AraC protein: a love–hate relationship

Robert Schleif

Summary

In the bacterium *Escherichia coli*, the AraC protein positively and negatively regulates expression of the proteins required for the uptake and catabolism of the sugar L-arabinose. This essay describes how work from my laboratory on this system spanning more than thirty years has aided our understanding of positive regulation, revealed DNA looping (a mechanism that explains many action-at-a-distance phenomena) and, more recently, has uncovered the mechanism by which arabinose shifts AraC from a state where it prefers to bind to two well-separated DNA half-sites and form a DNA loop to a state where it binds to two adjacent half-sites and activates transcription. This work required learning how to assay, purify, and work with a protein possessing highly uncooperative biochemical properties. Present work is focussed on understanding arabinose-responsive mechanism in atomic detail and is also directed towards understanding protein structure and function well enough to be able to engineer the allosteric mechanism seen in AraC onto other proteins. *BioEssays* 25:274–282, 2003. © 2003 Wiley Periodicals, Inc.

Initial perspective

@#$**^ – #*~** AraC protein! Eight person-years to develop an assay to detect the protein, ten-person years to develop a purification procedure, more than twenty person-years to crystallize, more than 125 person-years of studying its function. My favorite molecule indeed! Although much of the struggle took place before and during the development of genetic engineering techniques, a ''prehistoric'' time when detection and purification of a protein were so difficult as now to be legendary and beyond comprehension, work with AraC protein is still technically difficult. For example, despite the existence of today's marvelous protein hyperproduction strains and affinity-purification procedures, AraC remains fractious. Hypersynthesis yields insoluble and intractable precipitates, inhibiting satisfactory affinity purification schemes. Routine preparations today yield 1 mg of 90% pure AraC protein per liter of culture. Despite the difficulties, the work on AraC has been well worth the effort. Two basic gene regulatory principles have emerged from the work, DNA looping, and the light-switch mechanism for ligand response, and we expect that more remain to be discovered.

Biology Department, Johns Hopkins University, 3400 N. Charles St. Baltimore, MD 21218.
Funding agency: NIH; Grant number: GM18277.
DOI 10.1002/bies.10237
Published online in Wiley InterScience (www.interscience.wiley.com).

While I was a graduate student, Walter Gilbert and Benno Müller-Hill published their famous paper on the detection and partial purification of *lac* repressor[1] and Ellis Englesberg published his paper containing genetic evidence that convinced some scientists, but probably not even the majority, that AraC protein acted positively to induce expression of the arabinose utilization genes in *E. coli*.[2] Lacking both modesty and common sense, I decided that I wanted to learn how to apply Gilbert's biophysical and biochemical approaches and to elucidate the function and mechanism of AraC. I thought that solving the mechanism of AraC as a postdoctoral student would make it possible for me to find a decent job. I therefore asked to join Gilbert's group, and he accepted me. At the end of three years of very hard work, I had not even succeeded in detecting AraC, but I did find a job in the Biochemistry Department at Brandeis University. Now, nearly a full scientific career later, all of it directed to studies on AraC, we seem to have reached the level of understanding that I naively had hoped to achieve in a few years.

Genes that allow the bacterium *Escherichia coli* to grow on the sugar L-arabinose were first identified as a laboratory exercise at Cold Spring Harbor Laboratory, and then later mapped and studied (Fig. 1).[2] Mutations in the third of these genes to be characterized, AraC, were peculiar in that they eliminated expression of the AraB, AraA, and AraD gene products as well.[3–5] This essay primarily deals with events following the publication of Englesberg's evidence that AraC somehow turned on expression of the other arabinose operon gene products. Many scientists were more than reluctant to believe these data because they overturned the simple notion that all genes would be regulated by repressors, as had been demonstrated in the *lac* operon and in lambda phage. We now know that, in its overall regulatory properties, AraC protein does what many other gene regulatory proteins do. It turns on and off the synthesis of a few proteins depending on an intracellular condition. When arabinose is present, AraC activates transcription of the genes coding for the uptake and catabolism of arabinose. When arabinose is not present, AraC not only does not activate transcription from the *ara* promoters, but it also actively represses transcription from

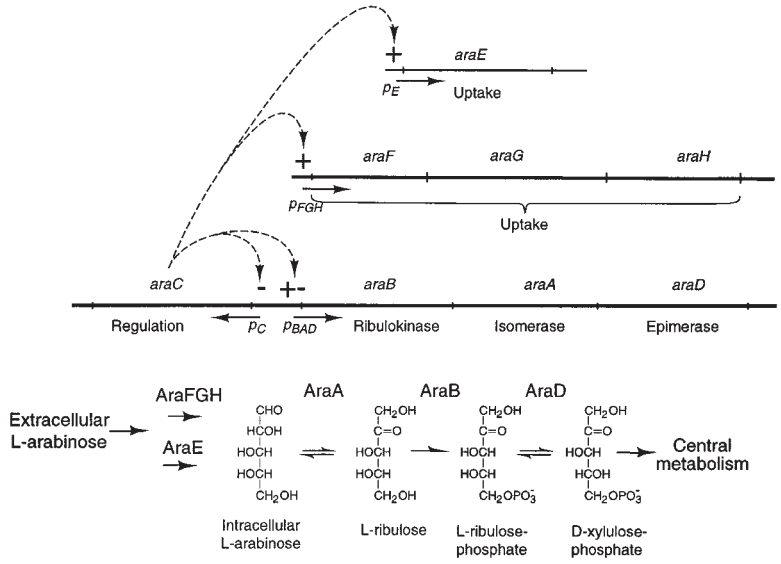


Figure 1. The three sets of arabinose-responsive genes of *Escherichia coli* and the metabolic pathway for the conversion of L-arabinose into D-xylulose-phosphate.

at least one of them, the intensively studied p_{BAD} promoter.[6,7] What, perhaps, is different from many systems is that the arabinose gene system is relatively simple, and that, despite the unpleasant biochemical properties of AraC, is somewhat amenable to genetic, biochemical, biophysical, and physiological studies. The apparent simplicity has made it possible over the years, to be learning enough about how it functions that graduate students and granting agencies have remained interested in the work. The liberty to apply diverse approaches to the same questions has multiplied the power of the individual approaches so that the inevitable consequences are that we now know much about how AraC protein regulates the arabinose operons.

The two basic gene regulatory principles discovered in the arabinose system have more than compensated for all the difficulty of studying the protein. First, the dimeric AraC protein molecule in the absence of arabinose contacts two half-sites on the DNA, I_1 and O_2, that are separated by 210 base pairs (Fig. 2). One monomer contacts one half-site, and the other monomer contacts the other half-site. This generates a DNA loop whose presence interferes with the access of RNA polymerase to two promoters in the looping region. When arabinose is present, instead of looping, AraC prefers to bind to the adjacent I_1 and I_2 half-sites on the DNA. In this state, AraC stimulates transcription of the p_{BAD} promoter via direct interactions with RNA polymerase.[8] Since its first discovery in the arabinose gene system, action at a distance by DNA looping has been observed in many gene regulatory systems.[9,10]

The second basic regulatory principle found in the *ara* system is the mechanism by which AraC responds to the presence of arabinose. The protein is a homodimer, and each of its monomers consists of a DNA-binding domain and of a dimerization domain that also contains an arabinose-binding pocket.[11–13] The critical factor is an N-terminal arm of about 18 amino acids that extends from the dimerization domain and, in the absence of arabinose, binds to the DNA-binding domain.[14,15] The combination of the covalent connection between the dimerization and DNA-binding domains and the binding of the N-terminal arms to the DNA-binding domains holds the DNA-binding domains in an orientation with respect to each other that favors their binding to the two well-separated I_1 and O_2 half-sites and forming a DNA loop (Fig. 2).[16] When arabinose is present, the N-terminal arms prefer, instead, to bind over the arabinose, and hence do not bind to the DNA-binding domains. This frees the DNA-binding domains and they can more easily bind to the adjacent I_1 and I_2 half-sites at the p_{BAD} promoter, a binding position from which AraC protein stimulates transcription from the promoter p_{BAD}. Thus AraC protein is on when its arms are in one position, and it is off when its arms are in the other position, hence the mechanism is named the light-switch mechanism. AraC can be thought of as a two-state molecular machine, with the state being determined by the presence or absence of arabinose.

The number of other proteins that are known to use arms or tails to make nonpermanent interactions with other domains or other proteins is growing.[9,10] The number is not large, however. Perhaps a sizeable number of systems utilize such interactions but have not yet been recognized to do so. The identification of these systems is particularly difficult as transient complexes are involved and because most of the protein structures that have been determined to date are of small, compact proteins or domains.

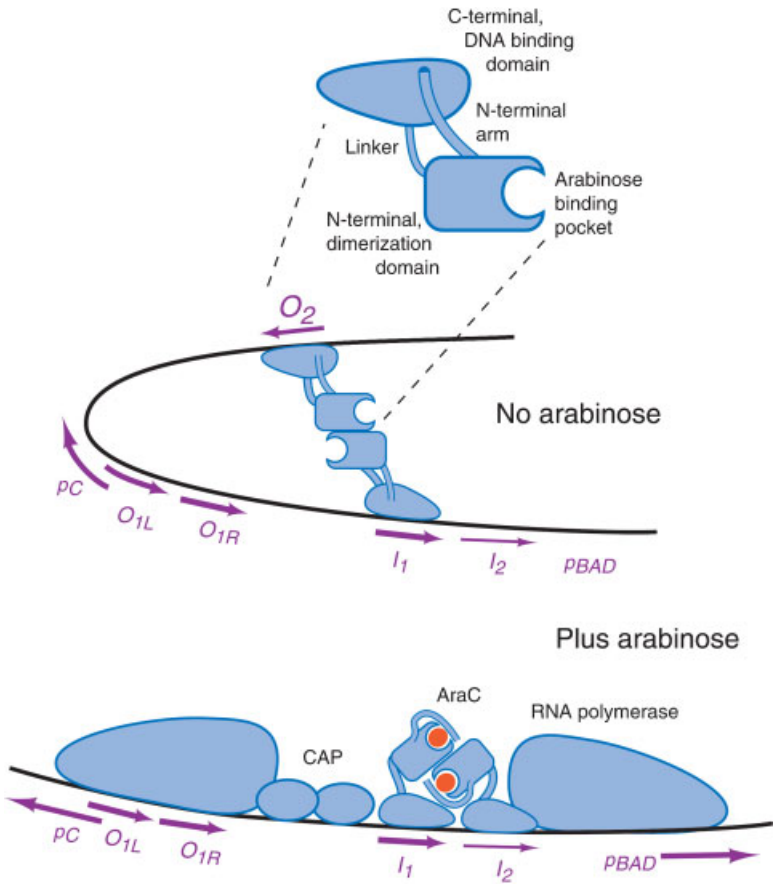


Figure 2. Binding of AraC in *trans* to the O_2 and I_1 half-sites to form a DNA loop in the absence of arabinose and its binding *cis* to the I_1 and I_2 half-sites in the presence of arabinose that leads to unlooping and induction of p_{BAD} and transient derepression of p_c and the light-switch mechanism.

Bad branch on a family tree

Before reviewing our current understanding of the mechanism by which AraC functions and the experiments that got us to this point, I should explain why AraC is such an ill-behaved protein. The root of the problem seems to be the protein's large DNA-contacting region. The reasoning is as follows: AraC protein is a dimer, and each monomer contacts two adjacent major groove regions of the DNA using two helix-turn-helix motifs.[17–22] These are facts. Now for the conjecture. The potential binding energy to DNA resulting from the large contact area of almost 40 base pairs could easily weld the protein onto the DNA and interfere with normal DNA metabolism like DNA replication. Even the *E. coli* RNA polymerase, which protects about 50 bases from nuclease digestion, makes intimate contact with less than 18 bases or base pairs.[23] One way to weaken the binding of a protein like AraC to DNA would be to use some of the DNA-binding energy to complete the folding of the protein. In other words, the protein likely does not complete folding in the absence of DNA. A partially folded DNA-binding domain would be very sensitive to proteases, and could have an excessive number of hydrophobic residues exposed. This would lead to aggregation as well as the irreversible binding of the protein to most column matrix materials or container walls—all properties of AraC.

The difficult biochemical properties of AraC protein early on led us to consider the possibility that other, better behaved, regulatory proteins might have structures similar to that of AraC and might function by the same mechanisms. The best candidate was the regulator of the rhamnose operon. Power had shown that the rhamnose genes likely were controlled in a positive fashion,[24] and since rhamnose is a sugar that is catabilized by a pathway strikingly similar to that utilized for arabinose, we thought it likely that AraC and the regulator of the rhamnose genes of *E. coli* would function by the same mechanism. Steve Hahn began the rhamnose work in my laboratory, and then James Tobin cloned and sequenced the rhamnose regulatory genes and developed a partial purification for small amounts of one of them.[25,26] Alas, RhaR was even more poorly behaved than AraC. The sequencing showed that, indeed, AraC, RhaR and a second regulator of the rhamnose genes, RhaS, shared sufficient sequence

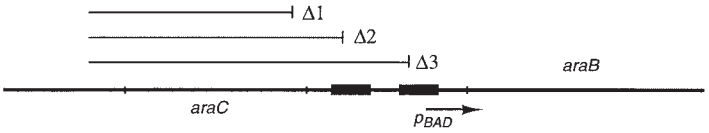
similarity over their C-terminal 100 amino acids that they very likely possessed similar tertiary structures. A wide variety of genetic and biochemical data associated this domain with DNA binding. Hence, the first three members of the family of DNA-binding proteins now known as the AraC/XylS family[27] had been identified. Since then the family has grown to over a hundred members. The other proteins that share sequence homology to the DNA-binding domain of AraC appear also to share the same intractable qualities. This is further evidence that the problematic behavior is a property of the DNA-binding domain, for the family members are highly divergent outside the DNA-binding domain. As a result of this poor behavior, very little biochemical information has been learned about any of the other proteins, that is, except for the two that are known to be monomers. Those two, RobA and MarA, are sufficiently well behaved that they have been purified and crystallized, and their structures determined while bound to DNA, exceptions proving the rule.[21,22] As these two proteins function as monomers, they do not possess excessive binding energy to DNA, and they can afford to be more completely folded in the absence of DNA, and hence are well behaved.

History and context of the discovery of DNA looping

One of the reasons for the existence of gene regulatory systems is to conserve resources by allowing the synthesis of selected gene products only when they are needed by the cell. Hence, the machineries of efficient regulatory systems themselves ought to consist of small numbers of molecules. While this is fine for the cell, it did not make life easy for the investigator in the era before genetic engineering. At that time, it was necessary to devise a sensitive assay for a specific regulatory protein based on a unique property of the protein. In the case of AraC, assays based on the binding of arabinose yielded proteins involved in active transport,[28,29] but not AraC. Consequently, the approach that seemed most likely to detect AraC protein seemed to be the ability of the protein to activate synthesis of the enzymes of the arabinose operon. Fortunately, Geofrey Zubay had just described the preparation of a cell extract that was capable of transcribing *lac*-containing DNA into RNA and translating the RNA into active β-galactosidase.[30] This system did yield the synthesis of arabinose enzymes upon addition of an AraC-containing extract to the extract capable of transcription–translation, although it was necessary, first, to increase the intracellular synthesis of AraC by increasing the number of AraC gene copies in cells used to make the AraC-containing extract, and, second, to eliminate ribulokinase from the extract used for transcription and translation.[29,31] The first requirement was met by infecting cells with a lambdoid phage carrying the arabinose operon, a phage that had first been made by the Susan Gottesman and John Beckwith.[32] This provided a cell extract enriched in AraC. This was added to the transcription–translation extract that was primed by the addition of more lambda-ara phage DNA. The second requirement was met by deleting the arabinose operon from the cells used to make both extracts. The synthesis of ribulokinase could then be detected by a sensitive radioisotope assay, which itself took more than 24 hours to perform. In the end, AraC could be detected, and this extremely cumbersome assay was used to guide the purification of AraC. Now, of course, most proteins of interest can be synthesized in such high levels that they may easily be detected and purified.

In parallel with the work on the detection of AraC protein and its purification for use in biochemical studies, we also pursued genetic studies. Part of the motivation for the genetic work was to assist the biochemical work, much like the isolation of the IPTG tight binding *lac* repressor mutants assisted identification of *lac* repressor,[1] and the isolation of overproducing mutants aided repressor purification.[33] Another motivation was to resolve a paradoxical finding that had been reported by Englesberg. He found a deletion entering the arabinose operon from the AraC end that removed the AraC gene and apparently ended in the regulatory region between AraC and AraB. The deletion left the p_{BAD} promoter normally inducible, but it eliminated repression of the promoter by AraC (Fig. 3).[6] How could AraC generate repression using a site upstream from all the DNA sites needed for the induction process? Possibly, the deletion ended just one or two nucleotides upstream of those required for induction, and the paradox was not real. Nonetheless, I thought that this finding was so unusual that I set about testing it with the isolation of new deletions. Hundreds were isolated, one per two-liter flask, and hundreds of point mutations were isolated, and a fine structure deletion map was constructed locating the deletion endpoints with respect to the point mutations. The locations of the deletions and point mutations appeared to be essentially random, and their high density allowed the size of the regulatory region between the AraC and AraB genes to be estimated at about 300 nucleotides.[34] More importantly, the number of deletions that reproduced the behavior of the Englesberg deletion suggested that the peculiar repression site lay several hundred nucleotides upstream from the sites that are required for induction.[35]

Progress in resolving the issue of upstream repression seemingly slowed while we developed methods to use the various *lambda-ara* transducing phages to isolate the *ara* regulatory region DNA and subportions of it.[36,37] This work partially succeeded, but was superseded by the development of genetic engineering techniques. With the new techniques we cloned and sequenced the regulatory region,[38] and, with a source of regulatory region DNA and pure AraC protein, we pursued biochemical studies. One of the first was the demonstration that the DNA migration retardation assay that had been described by Revzin and by Crothers could



Level of *ara* expression

	Δ1	Δ2	Δ3
No AraC present	<1	<1	<1
AraC provided *in trans*	1	**20**	<1
AraC provided *in trans* + arabinose	200	200	<1

Figure 3. The deletions that suggested the existence of repression of p_{BAD} that is dependent on an upstream DNA site. When a strain containing deletion two is provided with AraC in *trans*, the p_{BAD} promoter is normally inducible by arabinose but, surprisingly, in the absence of arabinose, the promoter possesses the elevated basal level of 20. The behaviors of the deletions one and two demonstrate the presence of an element lying between their endpoints whose absence leads to the elevated basal level.

be adapted to make biochemically meaningful measurements.[39–41] At this time, restriction enzymes were still not marketed, and it was the custom for the various molecular biology laboratories in the Boston area each to purify a different restriction enzyme and share it with the others. One day, an advertisement in the form of a Xerox sheet arrived from a new company, New England Biolabs, offering restriction enzymes for sale. This revolutionary commercialization quickly led to the preparation of a very similar, but fake advertisement that also included the sale of pure AraC. This was left on the desk of the graduate student working on the purification of AraC. The ultimate outcome of this was a devastating practical joke later played on me by my graduate students in collaboration with several faculty members.

The rapidly growing power of genetic engineering and the use of the DNA migration retardation assay along with pure AraC then allowed the careful study of the DNA-binding properties of AraC and the determination of the nucleotides that it contacted.[42,43] More importantly, it became possible to reinvestigate the question of repression from upstream sites.[44] More deletions were isolated, this time hundreds of times more easily than the first. Again a site was identified well upstream of the *ara* p_{BAD} promoter that was required for repression. Footprinting data verified the conclusion that AraC interacted with the upstream site.[45]

With a site identified well upstream of the promoter that was required for repression of the promoter, the question became one of mechanism. How could repression work from upstream. The most plausible mechanism seemed to me to be DNA looping, but my thoughts were rejected by everyone in my group. As a result, it was left to my technician and me to test the idea with the helical twist experiments (Fig. 4).[45] DNA looping was thus demonstrated, and at an opportune time, since enhancers in eukaryotic systems were being explored and were most puzzling in their action-at-a-distance behavior. We devised or refined a number of additional genetic, biochemical, and biophysical experiments to further test and demonstrate the existence of DNA looping.[46–48] Amongst these were the development and use of in vivo foot printing, and the measurement of the helical twist of DNA in vivo.

Finding the light-switch mechanism

Following the work on DNA looping, we discovered the extraordinary flexibility of AraC in being able to bind to direct repeat half-sites, inverted repeat half-sites, and half-sites separated by an additional helical turn of DNA.[49] This led to the expectation of separable dimerization and DNA-binding domains in AraC, a notion consistent with the then small family of proteins with sequence similarity to the DNA-binding domain of AraC. We developed both biochemical and genetic methods for the identification of domain boundaries and showed, indeed, that AraC possessed functionally independent domains.[50,51] For brevity, I shall omit descriptions of our work on the mechanism of transcription activation by AraC that was proceeding at the same time.

Up to this time very considerable effort had already been expended on attempts to crystallize AraC for structure determination by X-ray diffraction. At about this time, it became clear that understanding AraC function at any deeper level required knowledge of its tertiary structure. Hence, my technician and I began a crash and perhaps suicidal program for the crystallization of AraC. Although purification of sufficient protein required the growth of 30 liters of cells every third week for more than two years, the work of crystallization had been made simpler by the Hampton company that now sold multiple buffers for crystallization as well as siliconized cover slips. These simple conveniences substantially aided our crystallization efforts. In about half a year we had crystals, but were at first disappointed when we discovered that they were of

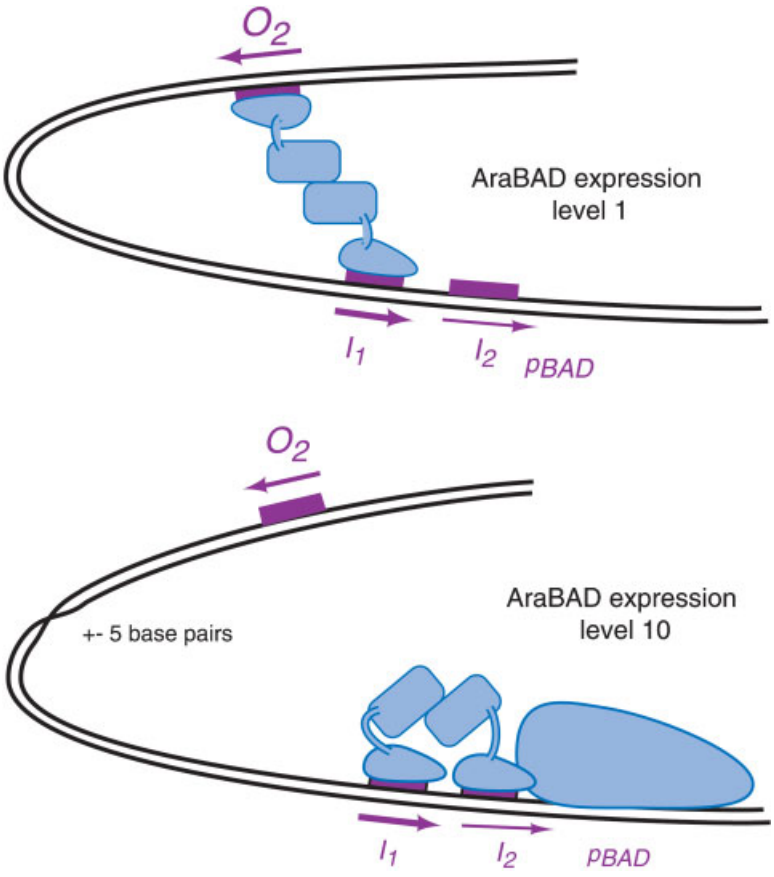


Figure 4. The helical twist experiments that demonstrated the existence of DNA looping. The insertion or deletion of five base pairs of DNA in a nonessential region between the promoter region and the O_2 half-site twist the two regions with respect to each other by half a helical turn, thereby placing the O_2 half-site on the back side of the DNA with respect to the I_1 half-site and the promoter.

AraC that had been proteolyzed during the crystallization incubation. Seeking to make lemonade from the lemons that we had been given, we protease digested pure AraC, further purified it, and attempted crystallization. The first crystal type that we could reproducibly grow was unsuitable for structure determination. The second grew much too slowly. We eventually learned how to coax the dimerization-arabinose-binding domain of AraC in the presence of arabinose to grow into a third form of crystal. While we looked lovingly upon our crystals as though they were our children, the crystallographers scorned them as too puny. Suitably motivated, we improved the growth conditions until the complaint was that the crystals were too large. Rather rapidly then, an X-ray diffraction group determined the structure of the AraC dimerization domain from the crystals that we provided.[13] Eventually, we also succeeded in growing crystals suitable for structure determination of AraC in the absence of arabinose and in the presence of an induction inhibitor, D-fucose.[52]

There are several initial responses to obtaining the structure of a protein one has studied for many years. Though not necessarily in this order, one looks to see (1) if the structure is consistent with all one knows about the protein, and (2) if there is anything unique or "novel" in the structure that would justify its publication in a leading journal. Soon one discovers that, although a structure contains an enormous amount of information, the information pertaining to mechanism of action is very well encrypted. Deciphering mechanism from structure is like trying to follow someone else's poorly documented computer program or to figure out the action of an electronic device from its wiring diagram. Eventually, we did come up with a mechanism for AraC that appeared to explain how the presence of arabinose could make the protein shift from DNA looping to binding to two adjacent half-sites. This mechanism depended heavily upon the structures of AraC in the presence and absence of arabinose and was consistent with the known biochemistry and genetics. We then tried different biochemical and genetics tests of the mechanism, but none worked! As we began to doubt the so-called mechanism, I spent many months considering alternatives. Typically this required assessing former experiments, deciding what each really told us and deciding which of our old experiments might have been misinterpreted or simply wrong. Very gradually, an idea developed that eventually became the light-switch mechanism. By this

time, we had a sizeable number of ''failed'' experiments that were attempts to prove the erroneous mechanism. Some of them were neutral with respect to the new idea, but several appeared to provide good support. In the next section I describe some of the experimental evidence in support of the light-switch mechanism.

Light-switch mechanism

As outlined earlier, in the absence of arabinose, the N-terminal arm extending from the N-terminal dimerization and arabinose-binding domain of AraC binds to the C-terminal DNA-binding domain. The combination of the arm's binding plus the covalent connection via the amino acids that link the N-terminal domain to the C-terminal domain holds the DNA-binding domains in the same relative orientation as the dimerization domains. This orientation is suitable for binding to the $I_1$and O_2 half-sites and forming a DNA loop. Binding to the adjacent I_1 and I_2 half-sites would require either bending the protein or breaking at least one of the arm–domain interactions. Apparently it is energetically more favorable for the protein to loop the DNA than to bind *cis* to I_1 and I_2. In the presence of arabinose, however, the interactions between the bound arabinose and residues of the N-terminal arm make it energetically more favorable for the arm to bind to the dimerization domain than to the DNA-binding domain. Hence the DNA-binding domains are no longer constrained in their orientations, and it is now energetically more favorable for them to bind to the I_1 and I_2 half-sites than to form a DNA loop.[14,16]

The mechanism as outlined above predicts that deletion of the N-terminal arms should leave the protein in an inducing state even in the absence of arabinose. It does.[14]

Mutations in the DNA-binding domain that weaken arm–DNA-binding domain interactions should make the protein unable to repress by DNA looping and hence put it in an inducing state even in the absence of arabinose. No such mutations were known when the mechanism was proposed. They were sought, and found, and the set of such mutations forms a stripe across the back side of the DNA-binding domain (whose structure can be reasonably accurately predicted on the basis of its homology to the MarA and RobA proteins) where the N-terminal arm presumably contacts the domain.[15,53] Additionally, mutations in the DNA-binding domain that leave the protein unresponsive to arabinose have been found.[14] These are predicted to strengthen arm interactions with the domain. The genetics data involving these and mutations in the arm that suppress the effects of such mutations are consistent with this interpretation. In the opposite way, mutations in the dimerization domain that appear to strengthen or weaken the interactions of the N-terminal arm with the domain have been found and are observed to have the predicted effects.[54]

According to the light-switch mechanism, it is the stiffness of the DNA and AraC that makes the protein favor looping in the absence of arabinose and favor binding *cis* in the presence of arabinose. If the stiffness of the DNA were eliminated by connecting the two half-sites with single-stranded DNA, such preferences should vanish. Such DNA can easily accommodate to wherever the DNA-binding domains prefer to be. Therefore, the dissociation rate of AraC from this split half-site DNA should be independent of the presence of arabinose whereas the dissocation rate of AraC from normal DNA should be much reduced by the presence of arabinose. Fig. 5 shows that arabinose affects only the dissociation rate from the normal DNA, the result expected if arabinose does not affect the intrinsic affinity of the DNA-binding domains for DNA.[55] This experimental technique permits easy determination of the mechanism by which a ligand modulates the affinity of a dimeric or oligomeric protein for DNA. If the ligand alters the intrinsic affinity of a DNA-binding domain for the DNA, the ligand will alter the affinity for DNA containing the two half-sites connected by single-stranded DNA. Alternatively, if the ligand modulates the affinity of the protein for DNA by changing the relative positioning of the DNA-binding sites, then the affinity of the protein for the special DNA should be independent of the presence of the ligand.

Another experiment that is consistent with the light-switch mechanism, but inconsistent with many alternatives is connecting two DNA-binding domains with a peptide linker, making it into a monomeric protein lacking the dimerization and arabinose-binding domain. This protein binds to adjacent half-sites and induces transcription, thus showing that the wild-type AraC protein is in the flexible or loose state in the presence of arabinose.[55]

The future

The natural course in the development of knowledge is the shift from observing and understanding phenomena in nature to using, designing and building things that work according to the same underlying principles. We are at a stage in our understanding of the arabinose operon regulatory mechanisms that it is appropriate to try to design other, but similar regulatory systems. For example, the light-switch mechanism is sufficiently simple that it is tempting to consider grafting it onto other proteins. The presence of a ligand could lead to the removal of an inhibitory peptide from an enzyme that is attached to the ligand-binding domain. More immediate objectives are to understand the details of the energetics of arm binding in AraC protein and in other proteins that use arm–domain interactions. We would like to predict and confirm the effects of alterations. Computations on energetics and dynamics of proteins coupled with the ability to construct and test the predictions undoubtedly will be crucial for future progress.

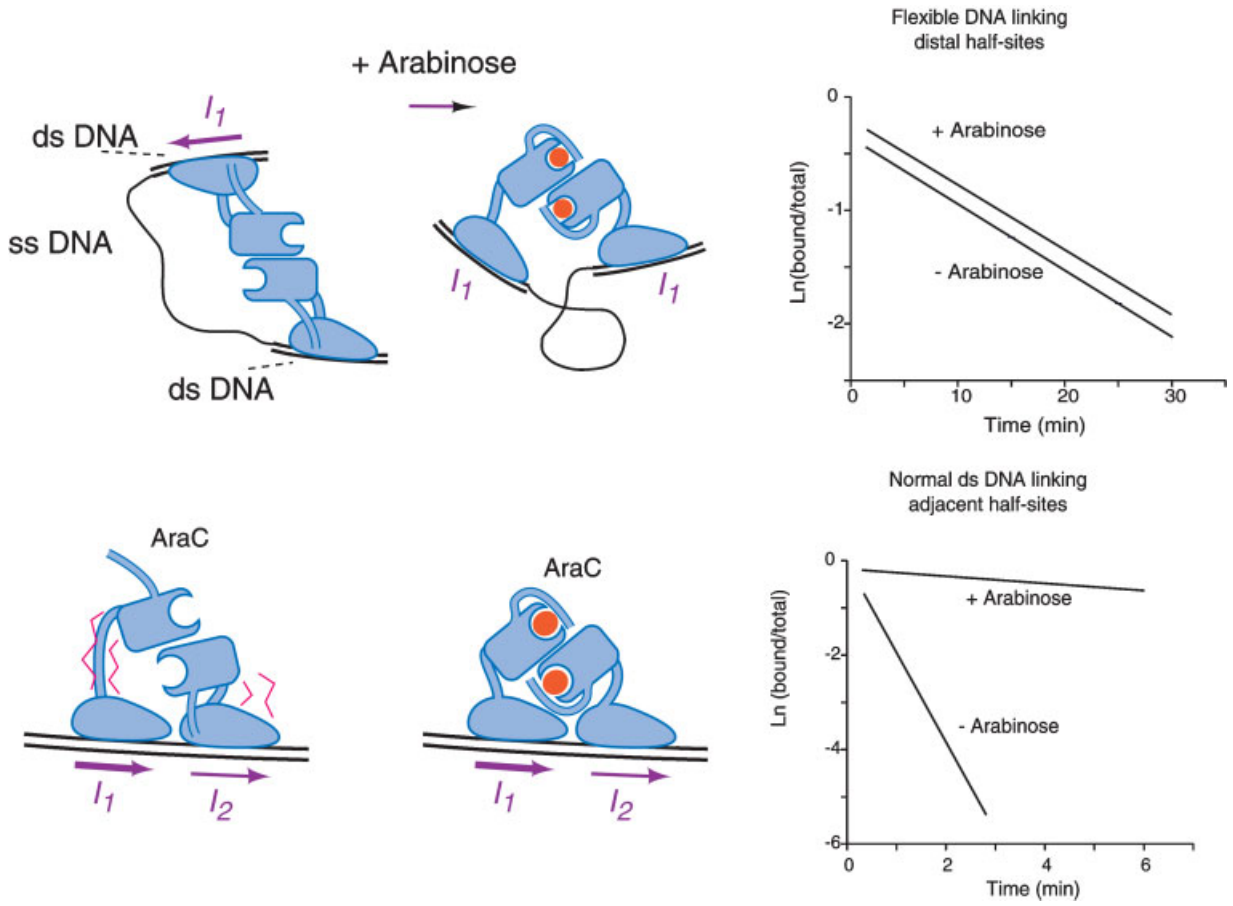


Figure 5. Rigidity experiments that showed that the effect of arabinose is on the relative positioning of the AraC DNA-binding domains rather than upon the intrinsic affinity of a domain for DNA. DNA whose two AraC-binding half-sites are connected by single-stranded DNA can easily be bound by AraC in the presence or absence of arabinose, and therefore displays identical dissociation rates from this DNA in the presence and absence of arabinose. The binding of AraC to normal double-stranded DNA in the absence of arabinose greatly strains AraC as indicated by the sawtooth symbols. Binding to the same DNA in the presence of arabinose does strain AraC. Hence, AraC dissociates more rapidly from this DNA in the absence of arabinose than in the presence.

Acknowledgments

About 40 graduate students, five postdoctoral associates and ten technical assistants have made contributions of inestimable value to the work described above. Only a few were mentioned by name, but all are remembered fondly. I thank Michael Rodgers for comments on this manuscript.

References

1. Gilbert W, Müller-Hill B. Isolation of the *lac* repressor. Proc Nat Acad Sci USA 1966;56:1891–1898.
2. Gross J, Englesberg E. Determination of the order of mutational sites governing L-arabinose utilization in *Escherichia coli* B/r by transduction with phage P1bt. Virology 1959;9:314–331.
3. Sheppard D, Englesberg E. Further evidence for positive control of the L-arabinose system by gene *araC*. J Mol Biol 1967;25:443–454.
4. Englesberg E, Irr J, Power J, Lee N. Positive control of enzyme synthesis by gene C in the L-arabinose system. J Bact 1965;90:946–957.
5. Gielow L, Largen M, Englesberg E. Initiator constitutive mutants of the L-arabinose operon (OIBAD) of *Escherichia coli* B/r. Genetics 1971;69: 289–302.
6. Englesberg E, Squires C, Meronk F. The L-arabinose operon in *Escherichia coli* B/r: a genetic demonstration of two functional states of the product of a regulator gene. Proc Nat Acad Sci USA 1969;62: 1100–1107.
7. Englesberg E, Sheppard D, Squires C, Meronk F Jr. An analysis of ''Revertants'' of a deletion mutant in the C gene of the L-arabinose gene complex in *Escherichia coli* B/r: isolation of initiator constitutive mutants (Ic). J Mol Biol 1969;43:281–298.
8. Zhang X, Reeder T, Schleif R. Transcription activation parameters at *ara* p_{BAD}. J Mol Biol 1996;258:14–24.
9. Schleif R. DNA Looping. Science 1988;240:127–128.
10. Schleif R. DNA Looping. Annu Rev Biochem 1992;61:199–223.
11. Steffen D, Schleif R. Overproducing *araC* protein with lambda-arabinose transducing phage. Molec Gen 1977;157:333–339.

12. Wilcox G, Meuris P. Stabilization and size of AraC protein. Mol Gen Genet 1976;145:97–100.
13. Soisson S, MacDougall-Shackleton B, Schleif R, Wolberger C. Structural basis for ligand-regulated oligomerization of AraC. Science 1997;276: 421–425.
14. Saviola B, Seabold R, Schleif R. Arm–domain interactions in AraC. J Mol Biol 1998;278:539–548.
15. Wu M, Schleif R. Mapping arm–DNA-binding domain interactions in AraC. J Mol Biol 2001;307:1001–1009.
16. Seabold R, Schleif R. Apo-AraC actively seeks to loop. J Mol Biol 1998; 278:529–538.
17. Ogden S, Haggerty D, Stoner CM, Kolodrubetz D, Schleif R. The *Escherichia coli* L-arabinose Operon: Binding sites of the regulatory proteins and a mechanism of positive and negative regulation. Proc Natl Acad Sci USA 1980;77:3346–3350.
18. Hendrickson W, Schleif R. A dimer of AraC protein contacts three adjacent major groove regions of the araI DNA site. Proc Natl Acad Sci USA 1985;82:3129–3133.
19. Brunelle A, Schleif R. Determining residue–base interactions between AraC protein and *araI* DNA. J Mol Biol 1989;209:607–622.
20. Carra J, Schleif R. Variation of half-site organization and DNA looping by AraC protein. EMBO J 1993;12:35–44.
21. Rhee S, Martin RG, Rosner JL, Davies DR. A novel DNA-binding motif in MarA: The first structure for an AraC family transcriptional activator. Proc Natl Acad Sci USA 1998;95:10413–10418.
22. Kwon HJ, Bennik MH, Demple B, Ellenberger T. Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. Nat Struct Biol 2000;7:424–430.
23. Youderian P, Bouvier S, Susskind M. Sequence determinants of promoter specificity. Cell 1982;30:843–853.
24. Power J. The L-Rhamnose genetic system in *Escherichia coli* K-12. Genetics 1967;55:557–568.
25. Tobin JF, Schleif RF. Positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J Mol Biol 1987;196:789–799.
26. Tobin JF, Schleif RF. Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of *Escherichia coli*. J Mol Biol 1990;211:75–89.
27. Gallegos M, Schleif R, Bairoch A, Hofmann K, Ramos J. AraC/XylS Family of transcriptional regulators. Microbiology and Molecular Biology Reviews 1997;61:393–410.
28. Hogg RW, Englesberg E. L-arabinose binding protein from *Escherichia coli* B/r. Bacteriol 1969;100:423–432.
29. Schleif R. An L-arabinose binding protein and arabinose permeation in *E. coli*. J Mol Biol 1969;46:185–196.
30. Zubay G, Chambers D, Cheong I. Cell-free studies on the regulation of the *lac* operon. In: Beckwith J, Zipser D, editors. The Lactose Operon. New York: Cold Spring Harbor Press; 1970. p 375–391.
31. Greenblatt J, Schleif R. Regulation of the arabinose operon in vitro. Nature New Biol 1971;233:166–170.
32. Gottesman S, Beckwith JR. Directed transposition of the arabinose operon: a technique for the isolation of specialized transducing bacteriophages for any *Escherichia coli* gene. J Mol Biol 1969;44:117–127.
33. Müller-Hill B, Crapo L, Gilbert W. Mutants that make more *lac* repressor. Proc Nat Acad Sci USA 1968;59:1259–1264.
34. Schleif R. Fine-structure deletion map of the *Escherichia coli* L-arabinose operon. Proc Nat Acad Sci USA 1972;69:3479–3484.
35. Schleif R, Lis J. The regulatory region of the L-arabinose operon: a physical, genetic and physiological study. J Mol Biol 1975;95:417–431.
36. Lis J, Schleif R. The isolation and characterization of plaque-forming arabinose transducing bacteriophage. J Mol Biol 1975;95:395–407.
37. Lis J, Schleif R. The regulatory region of the L-arabinose operon: its isolation on a 1000 base-pair fragment from DNA heteroduplexes. J Mol Biol 1975;95:409–416.
38. Smith BR, Schleif R. Nucleotide sequence of the L-arabinose regulatory region of *Escherichia coli* K-12. J Biol Chem 1978;253:6931–6933.
39. Garner M, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA Regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 1981;9:3047–3060.
40. Fried M, Crothers DM. Equilibria and kinetics of *lac* repressor–operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 1981;9:6505–6525.
41. Hendrickson W, Schleif R. Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. J Mol Biol 1984;178:611–628.
42. Ogden S, Haggerty D, Stoner CM, Kolodrubetz D, Schleif R. The *Escherichia coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. Proc Natl Acad Sci USA 1980;77:3346–3350.
43. Schleif R, Favreau A. Hyperproduction of *araC* protein from *Escherichia coli*. Biochemistry 1982;21:778–782.
44. Dunn T, Schleif R. Deletion analysis of the *Escherichia coli* P_C and P_{BAD} promoters. J Mol Biol 1984;180:201–204.
45. Dunn T, Hahn S, Ogden S, Schleif R. An operator at -280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. Proc Natl Acad Sci USA 1984;81:5017–5020.
46. Martin K, Huo L, Schleif R. The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites in vivo and repression-negative mutations lie in these same sites. Proc Nat Acad Sci USA 1986;83:3654–3658.
47. Lee D, Schleif R. In vivo DNA loops in *araCBAD*: size limit and helical repeat. Proc Nat Acad Sci USA 1989;86:476–480.
48. Lobell R, Schleif R. DNA looping and unlooping by AraC protein. Science 1990;250:528–532.
49. Carra J, Schleif R. Variation of half-site organization and DNA looping by AraC protein. EMBO J 1993;12:35–44.
50. Bustos S, Schleif R. Functional domains of AraC protein. Proc Natl Acad Sci USA 1993;90:5638–5642.
51. Eustance R, Bustos S, Schleif R. Reaching out: locating and lengthening the interdomain linker in AraC protein. J Mol Biol 1994;242:330–338.
52. Soisson S, MacDougall-Shackleton B, Schleif R, Wolberger C. The 1.6 Å crystal structure of the AraC sugar-inding and dimerization domain complexed with D-fucose. J Mol Biol 1997;273:226–237.
53. Reed W, Schleif R. Hemiplegic mutations in AraC protein. J Mol Biol 1999;294:417–425.
54. Wu M, Schleif R. Strengthened arm–dimerization domain interactions in AraC. J Biol Chem 2001;276:2562–2564.
55. 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.