

AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action

Robert Schleif

Biology Department, Johns Hopkins University, Baltimore, MD, USA

Correspondence: Robert Schleif, Biology Department, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA. Tel.: +1 410 516 5206; fax: +1 410 516 5213; e-mail: schleif@jhu.edu

Received 9 February 2010; revised 30 March 2010; accepted 5 April 2010.

DOI:10.1111/j.1574-6976.2010.00226.x

Editor: Juan L. Ramos

Keywords

allostery; transcription initiation; RNA polymerase; solubility; fluorescence; homology; fucose.

Abstract

This review covers the physiological aspects of regulation of the arabinose operon in *Escherichia coli* and the physical and regulatory properties of the operon's controlling gene, *araC*. It also describes the light switch mechanism as an explanation for many of the protein's properties. Although many thousands of homologs of AraC exist and regulate many diverse operons in response to many different inducers or physiological states, homologs that regulate arabinose-catabolizing genes in response to arabinose were identified. The sequence similarities among them are discussed in light of the known structure of the dimerization and DNA-binding domains of AraC.

The discovery and basic activity of the AraC protein

The *ara* gene system, AraC homologs, and AraC protein have been reviewed previously (Schleif, 1992, 1996, 2000, 2003; Gallegos *et al.*, 1997), but an updating and reevaluation are necessary in the light of recent advances. While this review mentions material covered in earlier reviews, it emphasizes material not reviewed previously, part of which has not been published previously.

The bacterium *Escherichia coli* can use L-arabinose, a five-carbon sugar and a constituent of some plant cell walls, as a source of carbon and energy. Initial work with the arabinose genes began as a laboratory exercise in a summer course at Cold Spring Harbor Laboratory. These studies led to the findings that four genes, *araA*, *araB*, *AraC*, and *araD*, code for proteins required for the uptake and conversion of L-arabinose to D-xylulose-5-phosphate. This final product then enters the pentose phosphate pathway (Fig. 1; Gross & Englesberg, 1959). Subsequent studies identified the enzymatic activities of AraA as L-arabinose isomerase that converts arabinose to L-ribulose, of AraB as a kinase that phosphorylates L-ribulose, and AraD as an epimerase that converts L-ribulose-phosphate to D-xylulose-phosphate

(Englesberg, 1961). These and additional studies also showed that while the levels of these proteins were considerably increased in wild-type cells grown in the presence of L-arabinose, the proteins were not similarly induced in AraC⁻ mutants (Sheppard & Englesberg, 1967). Because only a repressive type of gene regulation was well documented at the time, alternatives to AraC acting directly as an inducer or an activator of gene expression were favored until definitive *in vitro* proof was generated that the AraC protein was indeed a positive-acting gene regulator that turned on the synthesis of the other *ara* proteins in *E. coli* (Greenblatt & Schleif, 1971).

Further study of the *ara* system *E. coli* has revealed two transport systems (Kolodrubetz & Schleif, 1981a, b). In the lower affinity transport system, the transporter, the *araE* gene product (Lee *et al.*, 1981; Stoner & Schleif, 1983a), is bound to the inner membrane and utilizes the electrochemical potential to transport arabinose. The *araFGH* genes code for the arabinose-specific components of a high-affinity transport system, an ABC transporter (Hogg & Englesberg, 1969; Schleif, 1969; Brown & Hogg, 1972; Horazdovsky & Hogg, 1989). These are three proteins of the ATP-binding cassette transporter family. AraF is the periplasmic arabinose-binding protein, AraG is the

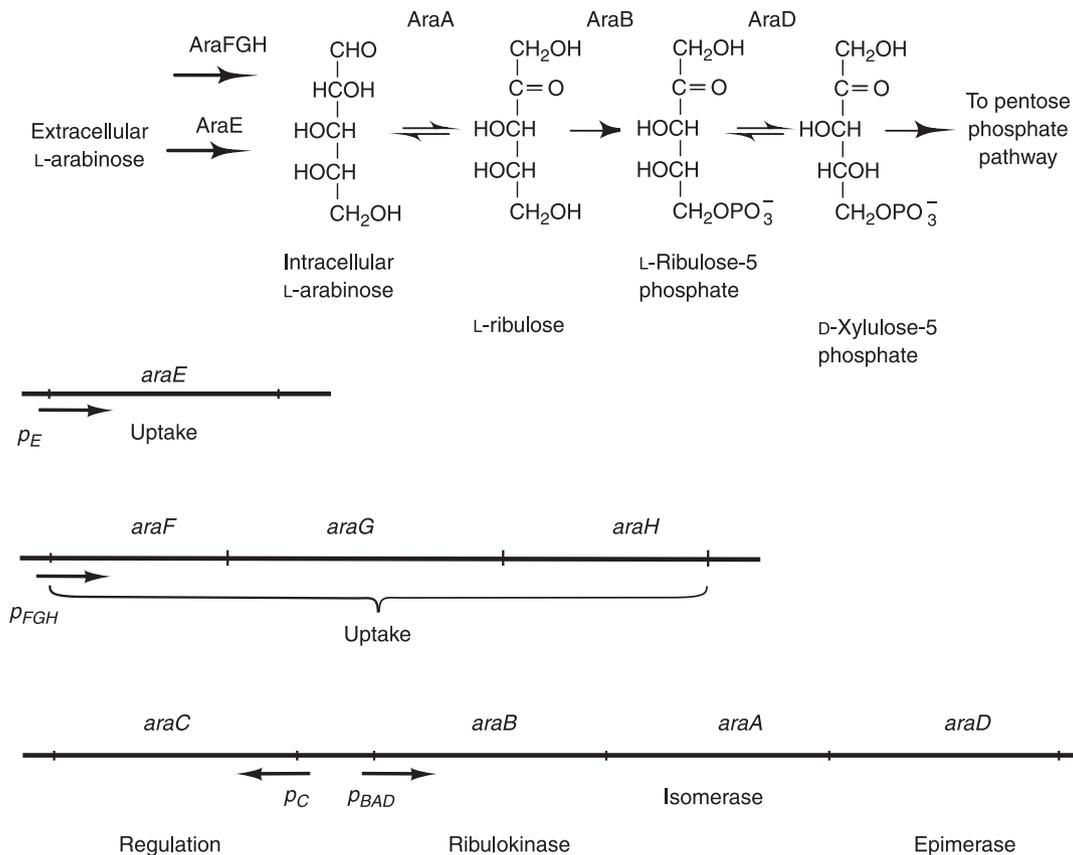


Fig. 1. The *Escherichia coli* genes coding for the proteins required for the uptake and catabolism of L-arabinose. Also shown is the metabolic pathway of this conversion.

ATP-binding component, and AraH is the membrane-bound component. The function of an additional protein induced by arabinose, the *araI* gene product (Reeder & Schleif, 1991), is as yet unknown (Carole *et al.*, 1999), but on the basis of its pattern of hydrophilic and hydrophobic residues, it has been conjectured to be a transporter either of arabinose-containing polymers or an exporter of arabinose.

Expression behavior of the *ara* promoters

Transcription of the *araBAD* genes under control of the p_{BAD} promoter is induced about 300-fold above a basal uninduced level within 3 s of the addition of arabinose to cells growing on glycerol in minimal salts medium (Schleif *et al.*, 1973; Hirsh & Schleif, 1973; Johnson & Schleif, 1995). The induction does not require the activity of the first enzyme of the arabinose catabolic pathway, implying that arabinose is the true inducer (Englesberg, 1961). This is not the case for the *lac* operon, where β -galactosidase is required to convert lactose to the true inducer, allolactose (Bourgeois & Jobe, 1972). Induction of the *ara* genes is negligible for cells growing on tryptone-yeast extract medium plus arabinose until an inhibitor, probably glucose, is consumed by the cells

(Lis & Schleif, 1973). A variety of *in vivo* and *in vitro* experiments further demonstrate the existence of catabolite repression and that the cyclic AMP receptor protein, CAP, plays an important role in the induction of the *ara* genes (Hahn *et al.*, 1984, 1986; Lobell & Schleif, 1991; Zhang & Schleif, 1998; Fig. 2).

Induction of the promoters for the active transport of arabinose, p_E and p_{FGH} , is only about 50-fold in response to arabinose (Johnson & Schleif, 1995). The promoter for the synthesis of AraC, p_C , also responds to the presence of arabinose. It is also stimulated by CAP, and is repressed by the AraC protein itself both by DNA looping in which a dimer of AraC binds to I_1 and to O_2 , and by repression in which a dimer of AraC binds to the O_1 pair of half-sites that partially overlap the p_C polymerase-binding region. Its regulatory behavior has been conveniently monitored by fusion of its promoter to the β -galactosidase gene (Casadaban, 1976). Upon arabinose addition, the promoter's activity increases up to 10-fold for about 10 min, and then reverts to almost its preinduction level (Ogden *et al.*, 1980; Stoner & Schleif, 1983b; Hahn & Schleif, 1983).

Because AraC is a regulatory protein, its *in vivo* level should be very low. Only a small excess of molecules of AraC

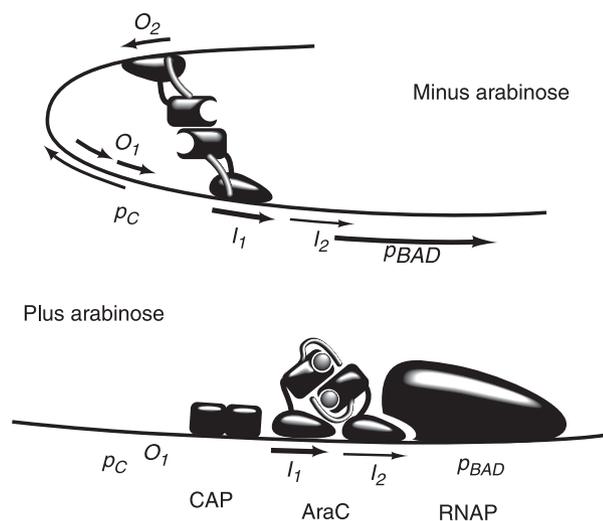


Fig. 2. The regulatory region of the *araCBAD* genes. I_1 and I_2 are termed half-sites as only a single subunit of AraC contacts each. They form what is sometimes referred to as the I site. O_1 consists of two half-sites and serves as an operator to the p_C promoter, whereas O_2 is a single half-site. The single CAP site serves both the p_C and the p_{BAD} promoters.

over the number of regulatory-binding sites on the chromosomal DNA for the protein would be needed. As expected, direct physical measurement of the normal level of AraC in a rapidly growing cell reveals a level of about 20 molecules per cell (Kolodrubetz & Schleif, 1981c).

Purification and solubility properties of AraC

Although clever genetic experiments and careful physiological measurements allow much to be deduced about the mechanism and function of a gene product, ultimately, definitive conclusions require making direct biochemical and biophysical measurements, and these require purified, active protein. The normal, low level of AraC in cells would require both a sensitive assay of the AraC protein to guide purification and roughly a 10^5 -fold enrichment from crude cell extracts to achieve 100% purity.

Originally, AraC could not be overproduced considerably (Steffen & Schleif, 1977), necessitating an assay of its activity in order to follow the protein during purification steps. The first reliable assay of AraC utilized its ability to stimulate *ara* mRNA synthesis and its translation in a coupled transcription–translation system. Ribulokinase that had been synthesized under the control of AraC was then measured (Greenblatt & Schleif, 1971). This assay allowed the purification of modest amounts of AraC and the determination of many of the protein's DNA-binding properties.

Clearly, a substantial overproduction of the AraC protein is necessary for full purification. Indeed, as the molecular biology tools for the overproduction of a protein became

available, they were applied to AraC. Unfortunately, although it became possible to so greatly overproduce most proteins by bacterial expression systems that they could be detected in crude extracts and followed during several simple purification steps by simple SDS gel electrophoresis, this was not possible at first for AraC. Substantial overproduction of the AraC protein generally yielded insoluble inclusion bodies. Even the refolding of inclusion bodies from urea or guanidine did not yield active AraC, although refolding of the AraC family member RhaS from urea in the presence of DNA does yield active protein (Egan & Schleif, 1994).

Many AraC family members possess similar solubility problems, and because they most obviously share homologous DNA-binding domains, these domains were expected to be insoluble. It was a surprise, however, to discover that the AraC DNA-binding domain itself is soluble, well behaved biochemically, and can be purified readily (Timmes *et al.*, 2004). Proteins often display a low solubility at pHs near their isoelectric points. Therefore, the fact that the isoelectric point of the dimerization domain of AraC is 6 and that of the AraC DNA-binding domain is 8 seems, to the detriment of biochemical studies, to minimize the solubility of AraC in the range of pHs in which the protein normally functions or in which most experiments would be conducted.

The original solution to the problem of purification of quantities of AraC sufficient for biochemical measurements was to overproduce the protein moderately (Steffen & Schleif, 1977; Schleif & Favreau, 1982). That is, to synthesize the protein in sufficient levels such that excessive purification was not required, but at levels sufficiently low that the protein did not precipitate while being synthesized or in the crude cell lysate. This approach allowed the purification of 25-mg quantities of AraC from 30 L of growth medium. Now, with the pET-based vector system for the overproduction of proteins, the understanding that the growth of cells at low temperatures retards protein aggregation, and that the solubility of partially purified or pure AraC is generally not a problem as long as a high salt concentration is maintained, it has become possible to obtain 100-mg quantities of pure AraC from 1 L of growth medium using simple purification procedures.

How AraC binds arabinose

It is not a trivial problem to appropriately regulate the expression of the proteins required for the uptake and catabolism of arabinose. It would seem sensible for cells not to induce significant expression of these proteins unless the extracellular arabinose concentration exceeds a value sufficient to support a reasonable rate of cell growth. If we assume that every molecule of arabinose reaching the cell

surface as a result of the combined effects of diffusion and cell motility is taken up and catabolized, then rough calculations indicate that it might be reasonable to induce when the extracellular arabinose concentration exceeds 10^{-7} M. The affinity of the AraC protein for arabinose, about 0.4 mM (Ross *et al.*, 2003), is much below this, indicating that at least one of the assumptions is far from correct.

On the other hand, however, the AraC protein is inside the cell, and the concentration of arabinose there depends not only on the extracellular arabinose concentration but also on the rates of arabinose uptake and catabolism. Furthermore, these two rates depend on the level of expression of the proteins that are controlled by AraC. Consequently, the system response will be nonlinear, that is, a slight induction can lead to a considerably more induction, and slightly less induction can lead to no induction at all. A side product of this nonlinear behavior is that the system will display a maintenance behavior, as has been observed (Siegele & Hu, 1997). That is, over a range of arabinose concentrations, an induced cell remains induced, while an uninduced cell possesses a low probability of inducing. Such a behavior can confound attempts to achieve uniform partial induction in all cells of a culture by inducing with intermediate concentrations of arabinose.

The crystal structure of the dimerization domain of AraC with arabinose bound reveals that the binding of arabinose generates an intricate network of hydrogen bonds between arabinose and residues lining the binding pocket (Soisson *et al.*, 1997). Analysis of the structure allowed Tang *et al.* (2008) to identify those residues most critical to the stereospecific binding of L-arabinose vs. D-arabinose. Upon randomization of these four residues and selection, it was possible to identify variants that bind D-arabinose in preference to L-arabinose.

Because the dimerization domain of AraC contains five tryptophan residues, one of which is located at the bottom of the arabinose-binding pocket, it is reasonable to expect that a change in fluorescence would be generated by the binding of arabinose. Indeed, upon the binding of arabinose, a reduction in the intrinsic fluorescence of AraC of up to ~2% can be observed. Greater reproducibility and accuracy, however, are obtained by measuring the 0.8-nm shift in the average emission wavelength (Weldon *et al.*, 2007). Surprisingly, the small wavelength shift shows considerably smaller fluctuations both in day-to-day measurements and in repeat measurements than the intensity shift. In part, this may be due to the fact that determination of the average emission wavelength combines 80–100 measurements.

Wild-type AraC protein does not bind just L-arabinose. D-fucose, a structural analog of L-arabinose (5-methyl L-arabinose), also binds to AraC (Ross *et al.*, 2003). Addi-

tionally, Lee *et al.* (2007) have found that the presence of IPTG interferes with arabinose induction and that mutants of AraC can be isolated in which arabinose remains an inducer, but the interfering ability of IPTG is much reduced. These results show that AraC also binds IPTG. That IPTG should bind to AraC and that it might interfere with arabinose induction is not altogether surprising. IPTG possesses a D-galactose moiety, and the ring structures of L-arabinose and D-galactose are identical. D-galactose itself appears not to be a strong inhibitor of arabinose induction as it binds very weakly, which is consistent with Lee and colleagues finding that in addition to residues in the arabinose-binding pocket, residues some distance from the arabinose-binding pocket appear to interact with part of the IPTG molecule.

DNA-binding properties of AraC

Once the AraC protein could be partially purified using the coupled transcription–translation assay, other assays could be tested and developed. One of the most convenient of these is the DNA migration retardation assay, also known as the gel shift assay or the EMSA assay. It was first reported for systems other than *ara* (Fried & Crothers, 1981; Garner & Revzin, 1981). Then, experiments with the AraC protein showed that the assay could be used for the detection of AraC in crude extracts, for assay of activity, the measurement of association and dissociation rates, and for the measurement of equilibrium-binding constants (Hendrickson & Schleif, 1984). It is also suitable for measuring the kinetics of binding of RNA polymerase to promoters and the formation of open complexes (Zhang *et al.* 1996).

Fluorescence anisotropy is the basis of another convenient assay of the binding of AraC to DNA (Timmes *et al.*, 2004). In principle, if a short piece of DNA containing a protein's binding site is fluorescently labeled, binding of the protein reduces the tumbling rate of the DNA. This can then be detected as an increase in the fluorescence anisotropy. In practice, the situation is slightly more complicated. The most convenient means of attaching a fluorophore to DNA utilizes six-carbon linkers. Rotation about these six bonds substantially decouples rotation of the fluorophore from that of the DNA. Such labeled DNA produces weak anisotropy signals on DNA binding. To see significant anisotropy signals, it appears that the fluorophore must be attached to the DNA in a position that allows it to make direct contact with the bound protein. It is this binding that reduces the fluorophore's tumbling and increases the anisotropy. Fortunately, the strength of the fluorophore–DNA binding usually does not considerably affect DNA-binding constants measured with the assay.

In the presence of arabinose, the association rate, k_1 , of the AraC protein to the I_1 – I_2 site is near the diffusion limit

of $2 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$, the dissociation rate, k_{-1} , is around 0.05 s^{-1} , and the equilibrium dissociation constant, K_d , is $2 \times 10^{-13} \text{ M}$ (Hendrickson & Schleif, 1984). Measurement of the latter value at physiological salt concentrations required special techniques to radiolabel the DNA to levels sufficient to allow its detection at the very low concentrations required in order that the concentration of free protein in solution be well approximated as the total protein added to the reaction mixture.

At equilibrium, the rate of association of protein and DNA, $k_1 \times [P] \times [\text{DNA}]$, equals the rate of dissociation, $k_{-1} \times [\text{PD}]$, where k_1 and k_{-1} are the association and dissociation rates. Because the equilibrium dissociation constant $K_d = [P] \times [D]/[\text{PD}]$, the dissociation rates and the equilibrium dissociation constant must obey the relationship $K_d = k_{-1}/k_1$. The three constants determined for AraC in the presence of arabinose agree reasonably closely with the relationship.

Binding of AraC to a single I_1 half-site is much weaker than to a pair of adjacent half-sites as is found at the *ara p_{BAD}* promoter. Consequently, the dissociation rate of AraC from a single half-site is much faster than from a pair of half-sites, and use of the gel-binding assay to detect binding or to measure binding constants then becomes problematic (Timmes *et al.*, 2004).

Protein-binding sites in the *ara* regulatory region

High-resolution deletion mapping, electron microscopy, DNase footprinting, hydroxyl radical footprinting, chemical protection, and premodification experiments have identified the binding sites of AraC, CAP, and RNA polymerase in the *araBAD* regulatory region (Fig. 2; Schleif, 1972; Lis & Schleif, 1975; Schleif & Lis, 1975; Hirsh & Schleif, 1976; Ogden *et al.*, 1980; Dunn *et al.*, 1984; Dunn & Schleif, 1984; Brunelle & Schleif, 1989; Carra & Schleif, 1993; Zhang *et al.*, 1996). Interestingly, the single CAP site serves both the divergently oriented p_C and p_{BAD} promoters (Dunn & Schleif, 1984).

It was a surprise initially to find that AraC contacted more than two adjacent major groove regions of DNA and that the binding site did not possess a strong inverted repeat symmetry like the other protein-binding sites known at the time (Hendrickson & Schleif, 1985). Although the consensus of the AraC-binding sites contains elements of inverted and direct repeat symmetry, the missing contact experiments mentioned above favor a direct repeat symmetry (Brunelle & Schleif, 1989). Consequently, the question of the symmetry of the half-sites for AraC binding was investigated more carefully with footprinting and with the synthesis of sites containing half-sites in various positions and orientations. These definitively showed that the natural

binding site of AraC at the p_{BAD} promoter is indeed a direct repeat (Carra & Schleif, 1993).

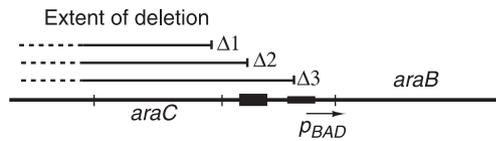
Does a direct repeat DNA-binding site imply that the binding protein possesses a similar direct repeat subunit structure? If each repeat unit of the DNA were contacted by a rigid, globular subunit, then the subunits would have to possess a direct repeat relationship with respect to each other. Such a linear structure would allow additional subunits to bind and the formation of polymers of indefinite length. Because AraC is a dimer in solution and was found to bind DNA only as independent dimers to triple and quadruple direct repeat DNA half-sites, the protein must be a dimer of a closed symmetry, that is, to possess a head-to-head symmetry (inverted repeat) (Carra & Schleif, 1993).

Thus, in order that the AraC protein be capable of binding to direct repeat half-sites, but not to form trimers, its DNA-binding domains cannot be integral parts of the dimerization domains. The two domains must be sufficiently independent that the dimerization domain can maintain a head-to-head symmetry while the DNA-binding domains bind to direct repeat half-sites. Such a modular domain structure is shown in Fig. 2.

The modular domain structure of AraC was proven by the construction of chimeras in which the dimerization domain was replaced with dimerizing coiled-coil domains or the DNA-binding domain was replaced with the DNA-binding domain of the LexA protein (Bustos & Schleif, 1993). The chimeras possessed the expected DNA binding and regulatory activities. Subsequently, it was shown that the inter-domain linker region could be lengthened or altered without altering the protein's regulatory properties (Eustance *et al.*, 1994).

Data that led to the discovery of DNA looping

Englesberg and colleagues found that two chromosomal deletions beginning upstream of the *araBAD* genes and ending within the *ara* operon possessed the peculiar property of being normally inducible, but having an elevated basal level (Fig. 3; Sheppard & Englesberg, 1967; Englesberg *et al.*, 1969a, b). Further work indicated that AraC possesses a repressing ability in addition to its inducing ability. In the two special deletions, however, the repressing ability was eliminated. This indicated that a site required for repression lies upstream from all the DNA sites required for induction at the p_{BAD} promoter. This implausible idea was tested using the power of lambda phage genetics to isolate and map hundreds of deletions ending within the *araBAD* operon (Schleif, 1972). The frequency of deletions ending between the site that is apparently required for repression and the sites required for induction indicated that the repression site is 200–500 base pairs (bp) upstream of all the sites required



Expression properties of deletions ending at different points in the *ara* regulatory region

Condition of expression	Deletion strain		
	Δ1	Δ2	Δ3
No AraC present	<1	<1	<1
AraC provided <i>in trans</i>	1	20	<1
AraC provided <i>in trans</i> + arabinose	200	200	<1

Fig. 3. The behavior of deletions that implied the existence of a site required for repression lying upstream from all the sites required for induction. The critical datum is shown in bold. The data show that deletion 1 behaves normally, as though it possesses all the essential regulatory sites. Deletion 2 can be interpreted as having lost a site that is required for repression so that any AraC in an inducing conformation is able to act on the remaining sites and activate transcription. When arabinose is provided, most of the AraC is imagined to be driven into the inducing state, and full induction is observed. Deletion 3 is presumed to have lost sites that are required for expression.

for induction (Schleif & Lis, 1975). Somewhat after the deletion mapping, it became possible to isolate plasmids carrying the arabinose operon and to conveniently generate and physically size deletions into the plasmid-borne *ara* genes (Dunn *et al.*, 1984; Dunn & Schleif, 1984). These showed that a site lying 210 bp upstream from the *ara* I_1 – I_2 site was required for repression. Footprinting experiments identified an AraC-binding site at this location.

Demonstrating DNA looping

At the time of the discovery of a site required for repression and lying hundreds of base pairs upstream from its target promoter, no convincing mechanism of action was known. Among the possibilities, however, was DNA looping in which the AraC protein bound at the upstream repression site interacted with an unknown component of the promoter or with the transcription initiation complex bound at the promoter. Helical twist experiments (Dunn *et al.*, 1984; Lee & Schleif, 1989; Fig. 4), provided strong evidence in favor of this idea. In these experiments, small insertions and deletions of known size were introduced between the upstream repression site and the sites required for induction. These insertions and deletions cyclically affected the ability of the upstream site to effect repression with a period of 10–11 bp. That is, insertions or deletions of 11, 22, and 33 bp could still repress normally, whereas insertions or deletions of 5, 15, 27 bp, etc. could not. Subsequent *in vivo* footprinting, mutational analysis, and *in vitro* DNA experiments provided

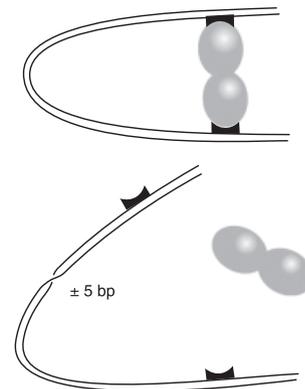


Fig. 4. Principle of the helical twist experiments. If two distally located sites are properly phased to allow DNA looping by the simultaneous binding of a protein to both, then the insertion or deletion of 5 p into a region between the sites rotates one site with respect to the other by half a helical turn of the DNA. This misorients the sites. Binding of a protein to both sites will now be energetically disfavored by the energy required to overtwist or undertwist the DNA between the binding sites by half a helical turn. In the case of DNA looping by AraC, but not so for some other looping systems, this energetic impediment was sufficient to drastically reduce DNA looping.

additional evidence for DNA looping. In addition, these showed that a dimer of AraC bound and formed the loop, with one subunit binding to the upstream half-site site named O_2 and the other subunit binding to the I_1 half-site (Martin *et al.*, 1986; Lobell & Schleif, 1990).

The value of looping

Why should nature use DNA looping as a mechanism in gene regulation? It seems to waste the DNA that constitutes the loop, and it seems more complex than systems in which the regulatory protein binds immediately alongside RNA polymerase or competes with polymerase binding. Several reasons can be advanced (Schleif, 1987, 1988).

The first is that the cooperativity generated by DNA looping allows a high occupancy of the binding sites, even though the affinity of the binding protein for each individual DNA site is quite low. The cooperativity arises because the binding of the protein to one of the two looping sites holds the protein in the vicinity of the other binding site, thus increasing the protein's effective concentration in the vicinity of the second site. It is the bridging that increases the binding. (It is an occasionally held misconception that multiple binding sites for a protein increase the solution concentration of the protein in their vicinity. This is not so, as the solution concentration, that is, the concentration of protein free in solution and not bound to a DNA-binding site, cannot be altered by the presence of nearby sites without violating the laws of thermodynamics.)

The cooperativity inherent in DNA looping means that at other sites, locations where looping is not possible, the binding of the protein will be very low. That is, DNA looping allows multiple DNA-binding sites to be used to increase the sequence specificity of DNA binding. This effect may be particularly important in eukaryotic cells, where thousands of different regulatory proteins must all be present in the nucleus. Hence, the concentration of each one is very low. If a protein is to bind at a gene's regulatory site, the protein's affinity for that site must be high. Unfortunately, a protein with such a high affinity for a site will bind so tightly that its dissociation may be so slow as to interfere with other biological processes such as DNA replication. Lowering the affinity of the protein for its site, but increasing its concentration will solve this problem, but would then require too high a protein concentration in the nucleus. DNA looping allows a relatively weak binding to individual sites, but when looping is possible, a high occupancy of the pair of looping sites.

In the case of the arabinose operon, a second function of DNA looping is to increase the dynamic regulatory range of *ara* gene expression (Saviola *et al.*, 1998). This effect is seen both experimentally and when modeling the *ara* system computationally. When the upstream repression site involved in looping is deleted, the basal level of expression of the arabinose operon is increased from 5- to 30-fold (Schleif & Lis, 1975; Dunn *et al.*, 1984). That is, DNA looping in the *ara* system interferes with the formation of the state that contributes to a large part of the operon's basal or uninduced expression level.

Computationally, the relative fraction of the *ara* operon copies that are free of protein, that are looped by AraC binding to the I_1 and O_2 half-sites, or that have AraC bound to the adjacent I_1 – I_2 half-sites in the inducing state can be found from the relative affinities of AraC for the various half-sites, the energetic costs of looping, and the intrinsic preference of AraC to loop (Seabold & Schleif, 1998). These relative affinities were precisely measured in a series of binding experiments in which two different DNA's were simultaneously present. The relative amount of binding to the two DNA's provided the relative affinities. The unknown energetic costs of looping and the intrinsic preferences of AraC were derived by fitting the expression levels of p_{BAD} to the equations yielding the occupancy of the various operon states when a variety of modifications had been made to the regulatory region, for example, with the O_2 half-site deleted. Calculations of the expression levels of various configurations of the regulatory region fit the experimental observations very well. The model clearly showed that a consequence of looping in the *ara* system is depletion of the state in which AraC is bound at I_1 – I_2 in the absence of arabinose. It is this state that provides the majority of the basal level of expression of the operon.

The half-sites I_1 , O_2 , and I_2 are involved in DNA looping and unlooping. The relative affinities of AraC for these three sites are roughly 1000, 100, and 10 (Seabold & Schleif, 1998). As would be expected, changing the sequence of the O_2 half-site to that of I_1 largely locks AraC in the looped state and prevents induction. Analogously, changing I_2 to bind AraC more tightly makes the operon constitutive and eliminates the requirement for arabinose for induction.

The activity of AraC and CAP in activating transcription

The transition of free RNA polymerase and the free promoter region to a promoter-bound RNA polymerase that is immediately capable of transcribing if provided with nucleotides often is well approximated as consisting of two steps. These are the reversible binding of RNA polymerase to the promoter in a closed complex and the transition of the closed complex to an open complex in which the promoter DNA is partially melted. The reversible binding is described with a dissociation constant, K_d , and because the formation of an open complex from the closed complex is well approximated as irreversible, it can be well characterized by a single rate constant, k_2 . Transcription activator proteins have been found that stimulate either or both of the steps. The DNA retardation gel migration assay is one of several convenient assays of open complexes, but it is well suited for the study of the *ara* p_{BAD} promoter. With such an assay, it was found that AraC stimulates both the binding of RNA polymerase and the transition from the closed to the open complex (Zhang *et al.*, 1996), a result subsequently found using a filter-binding assay with a hybrid *ara-lac* promoter (Lutz *et al.*, 2001).

In vivo experiments measuring the inducibility of the p_{BAD} promoter when the distal site to which AraC binds to form the repression loop, O_2 , was deleted showed that in the *ara* system, CAP both facilitates opening of the DNA loop and also either stimulates binding of RNA polymerase or the transition to the open complex (Lobell & Schleif, 1991).

Mutations in CAP that are defective in stimulating transcription from p_{BAD} were found to lie in two surface areas: AR1 and AR3 (Zhang & Schleif, 1998). Mutations in either region also affected the expression of *araFGH* and *rhaBAD*. Because a blank major groove of DNA lies between the CAP DNA-binding site and the AraC DNA-binding site, it seems unlikely that either AR1 or AR3 is involved in direct contacts with AraC. This is also consistent with the experimental finding of no detectable interaction between AraC and CAP upon their binding to *ara* region DNA. Instead of an AraC–CAP interaction, it seems more likely that one of the C-terminal domains of the α subunit of RNA polymerase contacts the AR1 region of the closer CAP subunit, and

that some other part of RNA polymerase contacts the AR3 region. Such an interaction seems plausible because the CAP–AraC–RNAP complex must substantially bend the DNA and it is very likely that in the complex CAP is positioned close to RNA polymerase despite the fact that the binding site for AraC lies between the binding sites for CAP and RNA polymerase.

The light switch mechanism of AraC explains looping and unlooping

Can we understand, at the atomic level, why AraC, in the absence of arabinose, prefers to loop DNA by binding to the well-separated I_1 and O_2 half-sites, and in the presence of arabinose, prefers to bind to the adjacent I_1 and I_2 half-sites? The light switch mechanism has been proposed to explain this and many other properties of AraC.

In the postulated mechanism (Saviola *et al.*, 1998), in the absence of arabinose, the two N-terminal arms of AraC interact with the DNA-binding domains to hold the DNA-binding domains in positions and orientations that favor DNA looping as shown in the top section of Fig. 2. In order for the AraC protein to bind to the adjacent I_1 and I_2 half-sites and stimulate transcription, at least one of the arm–DNA-binding domain interactions would have to be broken. The energetic cost of doing this is highly unfavorable. Consequently, the looped state is considerably more populated than the state of AraC binding to I_1 – I_2 .

In the presence of arabinose, the N-terminal regulatory arms are postulated to reposition, and as a result, to cease binding to the DNA-binding domains. Consequently, the DNA-binding domains become much more free to reorient and, as a result, can easily bind to the adjacent half-sites I_1 and I_2 . Because the sites are adjacent to one another, binding to them becomes the energetically favored state. Thus, the key controlling element in the behavior of AraC is the behavior of the N-terminal arm. The arm is therefore postulated to be free to interact with the DNA-binding domain in the absence of arabinose, and when arabinose binds in the pocket of the dimerization domain, the arm's interactions with the bound sugar hold the arm over the arabinose and prevent its interaction with the DNA-binding domain.

The results of nuclear magnetic resonance (NMR) experiments with the free N-terminal arm peptide and with arm peptide directly connected to the DNA-binding domain necessitate a refinement of the mechanism described above (Rodgers *et al.*, 2009). These show that the affinity of the arm for the DNA-binding domain is rather low, and that interaction of the arm with the DNA-binding domain occurs only when the arm has been structured by its interaction with the dimerization domain.

The extent of nature's use of light switch-type mechanisms involving N-terminal or C-terminal arms for controlled inter- or intra-domain interactions is not known, but arm–domain interactions are widespread in nature (Pabo *et al.*, 1982; Mak & Johnson, 1993; Lee *et al.*, 1996; Russo *et al.*, 1996; Berdis *et al.*, 1997; Luger *et al.*, 1997; Tai *et al.*, 1999; ter Haar *et al.*, 2000; de Beer *et al.*, 2000; Vetter & Wittinghofer, 2001).

Summary of the evidence for the light switch mechanism

Several lines of evidence led to the proposal of the light switch mechanism. First, in X-ray crystallographic studies of the dimerization domain, only the N-terminal arm was observed to significantly change structure upon arabinose binding. In the initial structure studies with the dimerization domain of a wild-type sequence in the presence of arabinose, residues of the arm were seen to be folded over the bound arabinose. The structure of the dimerization domain in the absence of arabinose was not initially helpful. In the absence of arabinose, the region formerly occupied by the arm was occupied by a subunit from another dimer of AraC (Soisson *et al.*, 1997). As this domain–domain interaction could have displaced the arm, the natural position and structure of the arm in the absence of arabinose could not be ascertained. In fact, in the minus arabinose structure, little electron density was observed for any of the arm residues, suggesting that in the crystallization, the arm had, indeed, been displaced and did not assume a unique structure.

In the crystal structure of the apo dimerization domain, not only could the arms not be seen, the arabinose-binding pockets were not empty. They were occupied by tyrosine residues from the opposing dimer. As it was possible that such an occupancy of the arabinose-binding pocket could have induced a structural change similar to that induced by arabinose, the similarity between the plus and the minus arabinose structures could have been misleading. Therefore, a mutant AraC was isolated in which the tyrosine was altered without changing the regulatory properties of AraC (Weldon *et al.*, 2007). This mutant dimerization domain was crystallized and its structure was determined. Fortunately, the mutant protein crystallized in a different lattice, and in this, the arabinose-binding pocket was occupied by a solvent. The structure of the core of the apo dimerization domain was seen to be virtually identical to that of the arabinose-bound dimerization domain. In this structure, the arm from residue 7 onwards could be discerned. Residues 7–18 of the arm possessed a completely different structure in the apo and holo structures (Fig. 5). The angle between the two monomers was also seen to have been shifted by 3°. The basis for this shift and the consequences are unknown at this point.

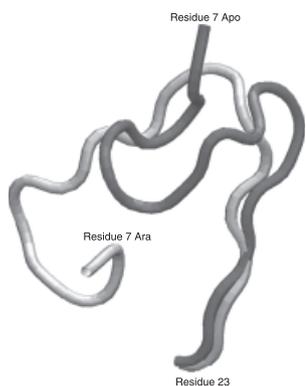


Fig. 5. The minus- and plus-arabinose backbone structures of residues 7–23 that constitute the N-terminal arm of AraC.

The drastic arabinose-controlled restructuring of the arm of AraC was utilized to construct a system in which the activity of β -galactosidase was modulated by the presence of arabinose (Fig. 6; Gryczynski & Schleif, 2004). Some deletions of the N-terminus of β -galactosidase leave the enzyme inactive. These α -deleted enzymes can be complemented by the addition of the α -peptide (Ullmann *et al.*, 1967; Zabin, 1982; Ullmann, 1992). Hence, placing a portion of the α -peptide sequence (carefully chosen to be compatible with the energetics of arm relocation in AraC) as an N-terminal extension on AraC yielded a system in which the availability of α -complementing activity could be modulated by the presence of arabinose. The construct showed, in fact, that the arm was more tightly bound to the dimerization domain in the absence of arabinose than in the presence of arabinose, a finding later shown more directly with NMR experiments (Rodgers *et al.*, 2009).

Another line of evidence that led to the formulation of the light switch mechanism is the behavior of mutations in the N-terminal arm. Through directed mutagenesis, the codons for each of the residues of the arm have been randomized (Ross *et al.*, 2003). At almost every position between residues 7 and 17, any change generates the constitutive phenotype of AraC. That is, the mutations leave AraC unable to loop and repress. Such a behavior is consistent with the arm's postulated activity in holding the DNA-binding domains such that looping is the energetically preferred state. It further suggests that most of the residues play important structural roles in the looping state.

The binding of D-fucose does not lead to induction. This behavior is easily reconciled with the light switch mechanism by postulating that fucose binding does not provide the interactions necessary for relocation of the N-terminal arm to the plus arabinose structure. Hence, the fucose-bound protein continues to repress.

Fucose provides for a simple genetic selection of repression-negative mutants in AraC (Englesberg *et al.*, 1965).

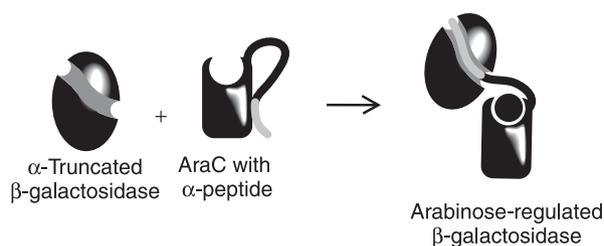


Fig. 6. Portable allostery. A construct in which the control of the N-terminal arm in AraC was used to add allosteric regulation by arabinose to β -galactosidase.

When both fucose and arabinose are present in the growth medium, the binding of fucose to AraC prevents induction, and hence, prevents growth. Mutants that can grow in the presence of arabinose and fucose are then easily isolated. Most of these are found to be constitutive, that is, they do not require either arabinose or fucose to be present for the arabinose operon to be highly induced. The fucose-resistant constitutive mutations impair the repression ability of AraC. Large numbers of constitutive mutants have been isolated and mapped in AraC making use of this property (Dirla *et al.*, 2009). For the most part, they lie in the N-terminal arm. This concentration stresses the importance of the arm to the repressing state of AraC.

Because the arm is postulated in the light switch mechanism to bind to the DNA-binding domain, it is surprising that the fucose selection does not yield constitutive mutations lying in the DNA-binding domain. When, however, more sensitive screening means than rapid growth in the presence of fucose are used, constitutive mutations are identified in the DNA-binding domain (Wu & Schleif, 2001a). These lie on the surface of the side opposite to that which binds DNA, but as yet, there is no direct physical proof that the N-terminal arm directly contacts the DNA-binding domain.

Mutations have also been found in the DNA-binding domain that appear to increase the strength of the interactions that hold AraC in the repressing state (Saviola *et al.*, 1998; Wu & Schleif, 2001b). These mutations drastically reduce inducibility *in vivo*. Mutations compensating for these uninducible mutations have been found in the arm, further strengthening the deduction that the arm and DNA-binding domain directly interact in the repressing state.

Several additional lines of evidence suggest that the DNA-binding domains of AraC are less constrained in the presence than in the absence of arabinose and that this positional and orientational freedom allows AraC to bind the adjacent half-site at P_{BAD} and induce. The first piece of evidence is that the protein formed by connecting two DNA-binding domains by a 20 amino acid flexible linker fully activates transcription (Harmer *et al.*, 2001). The second is obtained from examining the binding of AraC to two I_1

half-sites that are connected by flexible, single-stranded DNA segments from 6 to 24 nucleotides long. The binding affinity of AraC to the DNA with long linkers is unchanged by the presence of arabinose. In contrast, the binding of AraC to DNA containing the short linkers of six bases is tighter when arabinose is added (Harmer *et al.*, 2001; Rodgers & Schleif, 2008). This is interpreted as follows: the long linkers are sufficiently long as to allow both DNA half-sites to contact the DNA-binding domains of AraC no matter where these domains are positioned with respect to each other by their interactions with the dimerization domains of the protein. This would be true in the absence and presence of arabinose. With the short linkers, however, only in the presence of arabinose, when the DNA-binding domains would be free, would both DNA-binding domains be free to move close enough to one another that the two half-sites could be bound simultaneously.

Direct measurement of the postulated domain–domain interaction

The light switch mechanism predicts a direct interaction between the dimerization domains and the DNA-binding domains of AraC that is weakened or eliminated by the presence of arabinose. The experiments described so far in support of this point have all been indirect. A direct physical measurement of the expected domain–domain interaction would provide strong support for the mechanism.

The approximate strength of the expected interaction governs what kind of experiment may be performed to demonstrate its existence. In the absence of arabinose, the domains are to be mostly associated so that AraC will largely be in the looping state. In the presence of arabinose, the domains must be mostly unassociated so that they are free. This leaves them able to reorient to be able to bind to the direct repeat half-sites I_1 and I_2 . Because the domains are connected to one another by an eight amino acid inter-domain linker, perhaps 25 Å in length, the concentration of one domain in the presence of the other is very high. The concentration of one domain in the presence of the other is approximately the equivalent of a solution in which there is one molecule per cube of 50 Å on a side. This are 8×10^{21} molecules in the volume of a liter or a concentration of approximately 1 mM. At this concentration, in the presence of arabinose, the domains should not be associated. In the absence of arabinose, they should be associated. Thus, the strength of the domain–domain interaction for the two domains in solution should be on the order of 1 mM. This is a very weak interaction, one whose measurement with the two domains in solution would require that at least one of them be at about 1 mM in concentration. For the AraC dimerization domain, this would be around 24 mg mL⁻¹, a very high concentration.

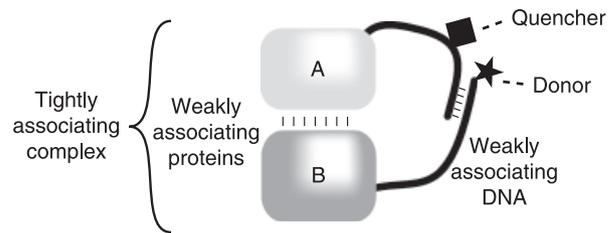


Fig. 7. The DNA assistance method for measuring very weak protein–protein interactions. Binding energy from the association of complementary regions of DNA is added to the binding energy of the proteins. The DNAs are also labeled with a fluorescent donor and a fluorescence quencher so that the association may be measured by fluorescence quenching.

Similar analysis shows that the strength of the interaction between transcription activators such as AraC and CAP and RNA polymerase is also very weak.

The DNA assistance method was devised to allow the detection and quantitation of very weak interactions between two proteins in solution (Frato & Schleif, 2009). In the method, single-stranded linker DNA sequences are conjugated to the protein (Fig. 7). At their ends, they contain short complementary regions. Hybridization of these regions provides additional binding energy for the protein–protein interaction. A fluorescent donor on one of the DNA molecules and a fluorescent quencher on the other DNA molecule allow a simple fluorescence assay of association. The strength of the protein–protein interaction is then calculated from the strength of the association of the protein–DNA complexes, a separate measurement of the association strength of the complementary DNA regions, taking into account the length of the flexible single-stranded linkers connecting the complementary regions to the proteins. The workability of the DNA assistance method was verified using model DNA–DNA interactions and interactions between two coiled-coil proteins.

When applied to AraC, the DNA assistance method detected an interaction between the domains of 0.37 mM in the absence of arabinose. This was weakened in the presence of arabinose to 0.67 mM. Although the magnitude of the interaction and the fact that it is weakened is as expected, deeper analysis shows that the binding energy difference is considerably smaller than what is required by the measured induction properties of the *ara* operon. This issue will be further discussed below.

At this point, it is not known to what extent modulation of the strength of domain–domain interactions is utilized by other AraC family members. Very preliminary experimental results with one AraC family member, XylS, have been interpreted to have detected a response to the XylS inducer in the interaction between dimerization domains and DNA-binding domains in solution (Dominguez-Cuevas *et al.*,

2008). Such a result might be consistent with a light switch-type mechanism operating in XylS in which the presence of a ligand frees the DNA-binding domain of XylS from the dimerization domain. To have detected an interaction at the micromolar concentrations of the two domains that were used in the experiments, however, the domain–domain interaction strength would have to be considerably tighter than 1 μ M. Then, upon binding of the ligand, the interaction would have to weaken to the millimolar range (if the length of the interdomain linker in XylS is similar to that found in AraC). Perhaps the simplest test of these striking findings would be measurements of the effects that the binding of the XylS ligand benzoate and DNA have on the binding of each other. According to the above estimations, benzoate should increase the affinity of XylS for DNA by more than a factor of 1000. It then follows from thermodynamic analysis that if binding of a ligand to XylS increases its DNA-binding affinity by more than a factor of 1000, then the binding of DNA must increase the affinity for the ligand by more than 1000, and this may be readily measured. It will be interesting to see whether this is ultimately found.

Structure of AraC

Efforts extending for over 20 years have failed to yield crystals of full-length AraC that diffract X-rays. Thus, the structure of the full protein is unknown. As mentioned earlier, the dimerization domain, however, has been crystallized in the absence and presence of arabinose (Soisson *et al.*, 1997; Weldon *et al.*, 2007). These structures show that arabinose binds in a β -barrel motif and that when arabinose is bound, the N-terminal arm of the protein folds over and covers the arabinose.

The DNA-binding domain of AraC can be purified, but it also resists crystallization. Its solution structure was therefore determined by NMR (Rodgers & Schleif, 2009). It was found to consist of the expected two copies of the DNA-binding helix–turn–helix motif. The fact that the domain was found to be well folded while not bound to DNA was unexpected. Its high sensitivity to proteases plus the low solubility of most AraC family members had led to the expectation that the common factor among the proteins, their low solubility, was attributable to their common component, their homologous DNA-binding domains. If these were all unfolded when not bound to DNA, the low solubility would have been readily explained.

Sequence analysis of the DNA-binding domain of AraC, and then of the evergrowing family of AraC DNA-binding domain homologs, suggested that the domain consisted of one helix–turn–helix DNA-binding motif and another DNA-contacting region with a rather low similarity to the canonical helix–turn–helix amino acid pattern. These pre-

dictions were tested with biochemical studies that identified specific amino acid–nucleotide interactions using the missing contact method (Brunelle & Schleif, 1989). The crystallization and structure determination of MarA and Rob while bound to DNA (Rhee *et al.*, 1998; Kwon *et al.*, 2000) provided definitive answers. These are two proteins with a sufficiently high homology to the DNA-binding domain of AraC to be confident that all three proteins share a similar tertiary structure. The MarA and Rob structures showed that both DNA-contacting regions in each of these proteins were helix–turn–helix motifs.

In the crystal structures of MarA and Rob, the two helix–turn–helix motif units of a single domain contact two adjacent major groove regions of the DNA, spanning a total distance of 17–19 bp. In the MarA structure, both recognition helices fit into the major groove, but in the Rob crystal structure, one of the recognition helices only lies across the groove. AraC also appears to fit its recognition helix into the major groove for Niland *et al.* (1996) have determined the effect on binding by AraC of singly altering each base of the *araI*₁ half-site. The resulting pattern of critical and noncritical bases is consistent with a structure of AraC–DNA in which the two recognition helices each extend into two adjacent major grooves, making base-specific contacts in each groove, and not making base-specific contacts in the minor groove lying between. It therefore seems likely that the RobA structure does not represent the true biological interaction between the protein and DNA.

Recently, the structure of the ToxT protein of *Vibrio cholerae*, an AraC ortholog with a moderate sequence similarity over both the dimerization and the DNA-binding domains, was published (Lowden *et al.*, 2010). Except for the absence of a region corresponding to the N-terminal arm of AraC, the tertiary structure of the two domains of this protein very closely parallels the structure of the two domains of AraC. The structure suggests that the DNA-binding properties of ToxT are controlled by immobilization–mobilization of the DNA-binding domain, somewhat like those proposed for AraC. In the case of ToxT, however, it is no DNA binding to DNA binding rather than DNA looping to binding *cis*.

AraC homologs and orthologs

Early on in the studies of AraC, it seemed likely that relatives of the AraC protein existed in bacteria. Therefore, genes of the rhamnose utilization pathway in *E. coli* were cloned, genetically and biochemically characterized, and indeed, found to possess not one, but two gene regulatory proteins related to AraC (Tobin & Schleif, 1987). Since that time, a number of additional homologs of AraC have been identified in *E. coli*, and now, in the fall of 2009, the NCBI database contains over 65 000 entries possessing biologically

significant similarity to AraC. Because most of these possess detectable sequence similarity only over the coding region for the DNA-binding domain of AraC, almost surely most of these regulate genes other than those coding for the uptake and catabolism of arabinose. A sizeable fraction of those whose function is known or can be inferred seem to be involved in the virulence or the control of the expression of extracellular proteins. Some of these may be regulated not by small molecules such as arabinose, but by other proteins that bind to the AraC ortholog (Plano, 2004).

Quite a number of the AraC orthologs possess detectable similarity to the *E. coli* AraC protein over both the dimerization and the DNA-binding domains. Some, those possessing only a handful of amino acid differences from the *E. coli* AraC protein, undoubtedly regulate *ara* operons. When the similarity reduces, however, it is not apparent which proteins regulate in response to arabinose and which are regulators that respond to different ligands and regulate genes other than those required for arabinose uptake and catabolism. In *E. coli*, the *araC* and the rhamnose regulatory genes *rhaR* and *rhaS* are immediately adjacent to the genes coding for the uptake and/or the catabolism of arabinose and rhamnose, respectively. Similar adjacency is seen in many other regulated bacterial gene systems. Thus, it seems highly likely that in a sequenced, but largely unstudied bacterium other than *E. coli*, a gene with sequence similarity to AraC that also lies adjacent to genes with high sequence similarity to the proteins of the *coli* arabinose operon that take up or catabolize arabinose is an authentic AraC ortholog that controls genes in response to arabinose.

Twenty orthologs of AraC whose coding gene lies adjacent to genes involved in the uptake or the metabolism of arabinose can readily be identified in the sequence databases as of fall 2009 (Table 1). Their sequence similarity to the *E. coli* AraC sequence ranges from 100% to about 50%. Notable is the tendency, with decreasing sequence similarity, to also find decreasing similarity in the arabinose-specific gene structure. Figure 8 shows the sequence alignments and the residues that are conserved among all the proteins.

Many of the conserved residues are as expected. For example, those that line the arabinose-binding pocket are highly conserved, as shown in Fig. 9, which depicts a single subunit of the dimerization domain in which arabinose in white in a Van der Waals representation is surrounded by the conserved residues that are shown in black.

The surface residues of the dimerization domain that are conserved might be expected to participate in conserved protein–protein or domain–domain interactions. Because the dimerization domain is not known to interact with other proteins, the conserved surface-exposed residues in this domain may be involved in interactions with the N-terminal arm or the DNA-binding domain. These residues are shown

Table 1. Homologs of AraC in the NCBI database as of fall 2009, their similarity to AraC and local gene order

Bacterial strain	NCBI similarity score to <i>coli</i> AraC	Number residues identical to <i>E. coli</i> AraC	<i>ara</i> gene order near AraC*
<i>Escherichia coli</i>	611	292/292	C*BAD
<i>Shigella boydii</i>	607	292	C*BAD
<i>Citrobacter</i> sp. 30_2	594	275	C*BAD
<i>Salmonella typhimurium</i> LT2	555	270	C*BAD
<i>Klebsiella pneumoniae</i> 342	538	266	C*BAD
<i>Cronobacter turicensis</i>	530	270	C*BAD
<i>Enterobacter sakazaki</i>	530	246	C*BAD
<i>Pectobacterium carotovorum</i>	321	202	A*B*FGHC
<i>Erwinia</i>	321	195	A*B*FGHC
<i>Dickeya dadantii</i> Ech703	318	198	A*B*FGHC
<i>Yersinia pseudotuberculosis</i> IP 32953	318	162	A*B*FGHC
<i>Serratia proteamaculans</i> 568	316	197	A*B*FGHC
<i>Aeromonas salmonicida</i>	313	197	A*D*B*FGHC
<i>Tolomonas auensis</i>	311	195	A*XB*FGHC
<i>Leptothrix cholodnii</i>	303	195	A*D*B*FGHC
<i>Vibrio furnissii</i>	258	177	H*G*F*BDAXC
<i>Actinobacillus succinogenes</i>	255	124	GHCBA
<i>Mannheimia succiniciproducens</i>	251	174	GHCBA?
<i>Reinekea</i> sp. MED297	250	175	Transport*A*C*BD
<i>Pantoea</i> sp. At-9b	229	167	A*B*Ftransport TransportC

*Genes are transcribed left-to-right unless indicated with an asterisk.

in Fig. 10. No obvious large patch of conserved residues is present, although the region including residues F15, R38, P39, and K43 worthy of further investigation.

The main dimerizing element of AraC appears to be a coiled-coil. Surprisingly, however, residues in this region are not highly conserved. It is an auxiliary dimerization interface region (Fig. 11), which is well conserved. Few regulatory mutations in the auxiliary dimerization interface have been isolated, and so it is unclear at this point what role the second dimerization region plays in the action of the AraC protein.

Figure 12 shows the conserved surface-exposed residues of the DNA-binding domain of AraC. As expected, many are part of the DNA contacting surface, which runs across the figure on the bottom-front of the domain. Any of the conserved residues that are not likely to be part of the DNA

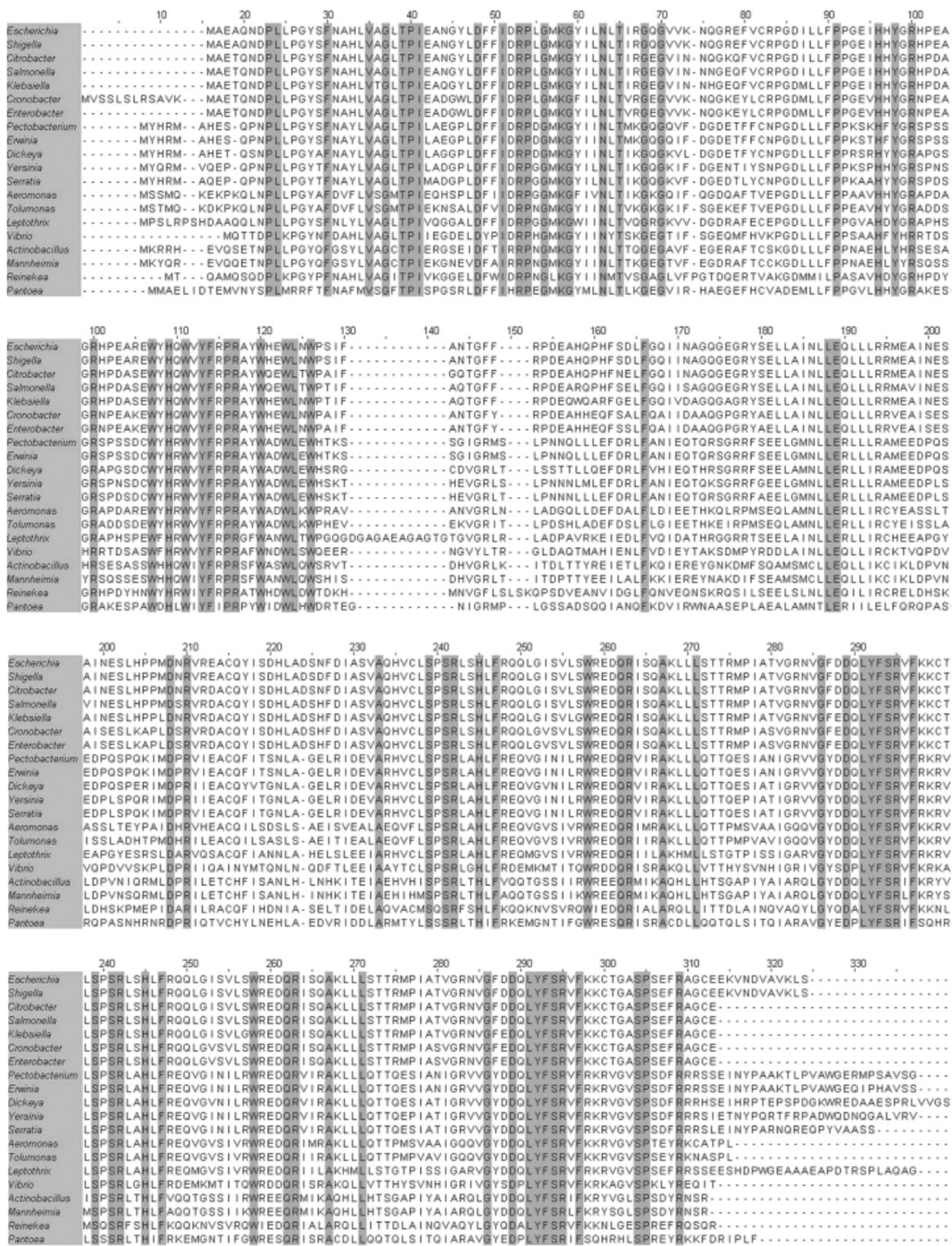


Fig. 8. Sequence alignments of homologs of AraC in which the fully conserved regions have a gray background.

contacting surface, S207, R178, Q230, and R277, are good candidates for contacting the dimerization domain or RNA polymerase. Without knowledge of the structure of DNA-

bound AraC, it is not possible to tell whether the highly conserved residue D257 contacts DNA or is likely to establish an important non-DNA interaction.

The regulatory protein of melibiose catabolism genes, MelR, possesses a clearly significant sequence similarity to AraC only over the DNA-binding domain. MelR orthologs in the NCBI database selected by the same criteria as used above for identifying orthologs of AraC also completely conserve the arginine corresponding to R277 of AraC.

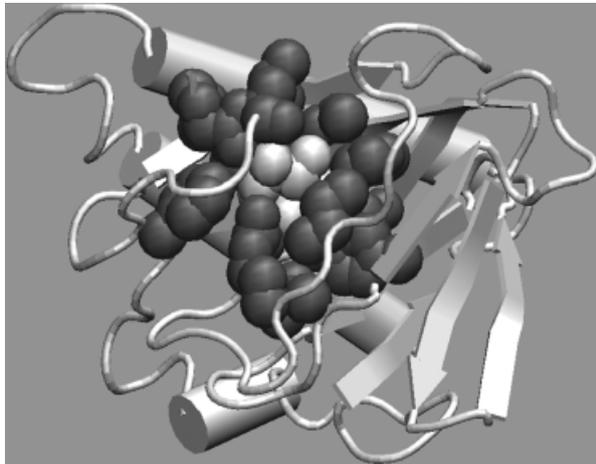


Fig. 9. All residues lining the arabinose-binding pocket in AraC are fully conserved in AraC homologs. Arabinose is shown in white, and the conserved surrounding residues, F15, T24, I36, R38, H80, Y82, W91, and H93 are in black. Only one subunit of the dimerization domain is shown.

Analogously, for the rhamnose regulators, RhaR and RhaS, the corresponding residue is highly, but not absolutely, conserved. Thus, residue 277 of AraC is identified as likely to be involved in important contacts to RNA polymerase.

RhaR and RhaS appear to be somewhat distant relatives of AraC. Their DNA-binding domains possess a clear similarity to AraC, but their N-terminal regions, which correspond to the dimerization and arabinose-binding domain in AraC, do not possess convincing sequence similarity to AraC. Because, however, the DNA-binding domain of AraC alone can activate transcription from *ara*

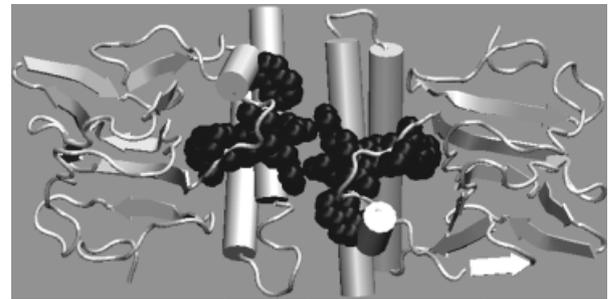


Fig. 11. The auxiliary dimerization interface of AraC. Conserved residues, F98, P100, R101, W104, W107, F134, L156, and E157 of this subdomain are shown in black.

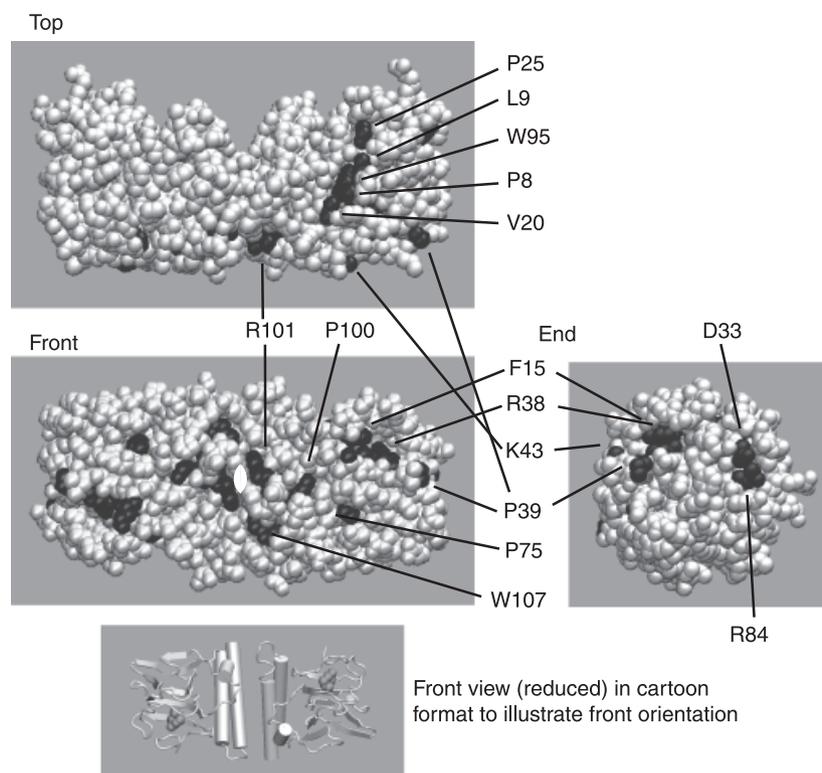


Fig. 10. Top, front, and end views of the dimerization domain in which the fully conserved surface residues are shown in black.

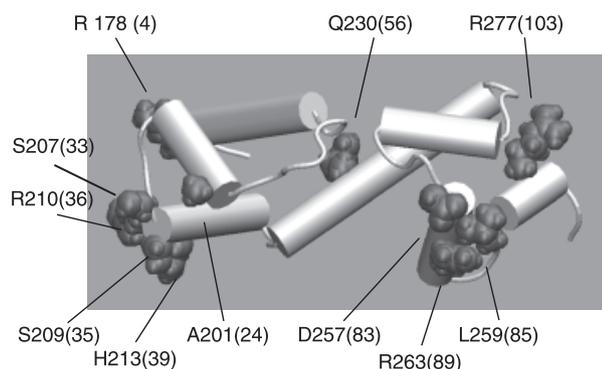


Fig. 12. The conserved surface-exposed residues of the DNA-binding domain of AraC. DNA binds across the bottom-front. The N-terminus is at the upper left. The supplemental numbering is that of the NMR structure of the domain.

p_{BAD} (Harmer *et al.*, 2001), just this portion of AraC interacts with RNA polymerase. Alanine scanning often yields small and unconvincing signals, but Egan's scans of RhaR and RhaS yielded fairly convincing data for contacts to RNA polymerase by RhaR residues D276, E284, and D285 and by RhaS residue 250 (Bhende & Egan, 2000; Wickstrum & Egan, 2004). In the sequence alignment to AraC, the corresponding residues in AraC are T248, D256, and D257 from RhaR and D257 from RhaS. In the twenty AraC orthologs, T248 is at a position of significant variability. D256 tends to be conserved, with aspartic acid appearing in 14 of the orthologs, glutamic acid appearing in four, and glutamine and serine appearing once. D257 of AraC is at a position that is fully conserved. These results strongly suggest that AraC residues D256 and D257 interact with RNA polymerase in activating the transcription of the *ara* genes. This prediction has not yet been tested.

Summary

The sites of AraC protein binding to DNA are by now well understood. Similarly, the looping and unlooping of AraC are well characterized. At a finer level, however, despite the fact that it is one of the more well-studied regulatory proteins, much remains to be learned about AraC. Although the light switch mechanism explains most of the major properties of AraC, there remain a number of 'lesser' activities of AraC that are not explained by the mechanism. AraC stimulates both the binding of RNA polymerase and the transition of RNA polymerase from a closed to an open complex, but precisely what residues participate in the interactions and what are the strengths of the interactions are not yet known. Finally, the atomic details that apparently lead to the relocation and restructuring of the N-terminal arm in the presence of arabinose are yet to be determined.

References

- Berdis A, Soumilion P & Benkovic S (1997) The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation. *P Natl Acad Sci USA* **93**: 12822–12827.
- Bhende PM & Egan SM (2000) Genetic evidence that transcription activation by RhaS involves specific amino acid contacts with sigma 70. *J Bacteriol* **182**: 4959–4969.
- Bourgeois S & Jobe A (1972) Lac Repressor-operator interaction. VI. Natural inducer of the *lac* operon. *J Mol Biol* **69**: 397–408.
- Brown CE & Hogg RW (1972) A second transport system of L-arabinose in *Escherichia coli* B/r controlled by the *araC* gene. *J Bacteriol* **111**: 606–613.
- Brunelle A & Schleif R (1989) Determining residue–base interactions between AraC protein and *araI* DNA. *J Mol Biol* **209**: 607–622.
- Bustos S & Schleif R (1993) Functional domains of AraC protein. *P Natl Acad Sci USA* **90**: 5638–5642.
- Carolé S, Pichoff S & Bouché JP (1999) *Escherichia coli* gene *ydeA* encodes a major facilitator pump which exports L-arabinose and isopropyl-β-D thiogalactopyranoside. *J Bacteriol* **181**: 5123–5125.
- Carra J & Schleif R (1993) Variation of half-site organization and DNA looping by AraC protein. *EMBO Journal* **12**: 35–44.
- Casadaban MJ (1976) Regulation of the regulatory gene for the arabinose pathway, *araC*. *J Mol Biol* **104**: 557–566.
- de Beer T, Hoofnagle A, Enmon J, Bowers R, Yamabhai M, Kay B & Overduin M (2000) Molecular mechanism of NPF recognition by EH domains. *Nat Struct Biol* **7**: 1018–1022.
- Dirla S, Chien HY & Schleif R (2009) Constitutive mutations in *E. coli* AraC protein. *J Bacteriol* **191**: 2668–2674.
- Dominguez-Cuevas P, Marin P, Busby S, Ramos J & Marques S (2008) Roles of effectors in XylS-dependent transcription activation: intramolecular domain derepression and DNA binding. *J Bacteriol* **190**: 3118–3238.
- Dunn T & Schleif R (1984) Deletion analysis of the *Escherichia coli* *pC* and *pBAD* promoters. *J Mol Biol* **180**: 201–204.
- Dunn T, Hahn S, Ogden S & Schleif R (1984) An Operator at – 280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *P Natl Acad Sci USA* **81**: 5017–5020.
- Egan S & Schleif R (1994) DNA-dependent renaturation of an insoluble DNA binding protein. *J Mol Biol* **243**: 821–829.
- Englesberg E (1961) Enzymatic characterization of 17 L-arabinose negative mutants of *Escherichia coli*. *J Bacteriol* **81**: 996–1006.
- Englesberg E, Sheppard D, Squires C & Meronk F Jr (1969a) 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* **43**: 281–298.
- Englesberg E, Squires C & Meronk F (1969b) The L-arabinose operon in *Escherichia coli* B/r: a genetic demonstration of two functional states of the product of a regulator gene. *P Natl Acad Sci USA* **62**: 1100–1107.

- Englesberg I, Irr J, Power J & Lee N (1965) Positive control of enzyme synthesis by gene C in the L-arabinose system. *J Bacteriol* **90**: 946–957.
- Eustance R, Bustos R & Schleif R (1994) Reaching out: locating and lengthening the interdomain linker in AraC protein. *J Mol Biol* **242**: 330–338.
- Frato KE & Schleif RF (2009) A DNA-assisted binding assay for weak protein–protein interactions. *J Mol Biol* **394**: 805–814.
- Fried M & Crothers D (1981) Equilibrium and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* **9**: 6505–6025.
- Gallegos M, Schleif R, Bairoch A, Hofmann K & Ramos J (1997) AraC/XylS family of transcriptional regulators. *Microbiol Mol Biol R* **61**: 393–410.
- Garner M & Revzin A (1981) 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* **9**: 3047–3060.
- Greenblatt J & Schleif R (1971) Regulation of the arabinose operon *in vitro*. *Nat New Biol* **233**: 166–170.
- Gross J & Englesberg E (1959) Determination of the order of mutational sites governing L-arabinose utilization in *Escherichia coli* B/r by transduction with phage Plbt. *Virology* **9**: 314–331.
- Gryczynski U & Schleif R (2004) A portable allosteric mechanism. *Proteins* **57**: 9–11.
- Hahn S & Schleif R (1983) *In vivo* regulation of the *Escherichia coli* AraC promoter. *J Bacteriol* **155**: 593–600.
- Hahn S, Dunn T & Schleif R (1984) Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. *J Mol Biol* **180**: 61–72.
- Hahn S, Hendrickson W & Schleif R (1986) Transcription of *Escherichia coli* ara *in vitro*, the cyclic amp receptor protein requirement for pBAD induction that depends on the presence and orientation of the araO₂ site. *J Mol Biol* **188**: 355–367.
- Harmer T, Wu M & Schleif R (2001) The role of rigidity in DNA looping–unlooping by AraC. *P Natl Acad Sci USA* **98**: 427–431.
- Hendrickson W & Schleif R (1984) Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J Mol Biol* **174**: 611–628.
- Hendrickson W & Schleif R (1985) A dimer of AraC protein contacts three adjacent major groove regions of the araI DNA site. *P Natl Acad Sci USA* **82**: 3129–3133.
- Hirsh J & Schleif R (1976) Electron microscopy of gene regulation: the L-arabinose operon. *P Natl Acad Sci USA* **73**: 1518–1522.
- Hirsh R & Schleif R (1973) *In vivo* experiments on the mechanism of action of L-arabinose C gene activator and lactose repressor. *J Mol Biol* **80**: 433–444.
- Hogg RF & Englesberg E (1969) L-arabinose binding protein from *Escheichia coli* B/r. *J Bacteriol* **100**: 423–432.
- Horzodovsky B & Hogg R (1989) Genetic reconstitution of the high-affinity L-arabinose transport system. *J Bacteriol* **171**: 3053–3059.
- Johnson C & Schleif R (1995) *In vivo* induction kinetics of the arabinose promoters in *Escherichia coli*. *J Bacteriol* **177**: 3438–3442.
- Kolodrubetz D & Schleif R (1981a) Regulation of the L-arabinose transport operons in *Escherichia coli*. *J Mol Biol* **151**: 215–227.
- Kolodrubetz D & Schleif R (1981b) L-arabinose transport systems in *Escherichia coli* K-12. *J Bacteriol* **148**: 472–479.
- Kolodrubetz D & Schleif R (1981c) Identification of AraC protein on two-dimensional gels, its *in vivo* instability and normal level. *J Mol Biol* **149**: 133–139.
- Kwon H, Bennik M, Demple B & Ellenberger T (2000) Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat Struct Biol* **7**: 424–430.
- Lee D & Schleif R (1989) *In vivo* DNA loops in AraCbad: size limit and helical repeat. *Proc Natl Acad Sci USA* **86**: 476–480.
- Lee JH, Al-Zarban S & Wilcox G (1981) Genetic characterization of the araE gene in *Salmonella typhimurium* LT2. *J Bacteriol* **146**: 293–304.
- Lee S, Owen K & Choi H (1996) Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. *Structure* **4**: 531–541.
- Lee SK, Chou HH, Ffleger BF, Newman JD, Yasuo Y & Keasling J (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl Environ Microb* **73**: 5711–5715.
- Lis JT & Schleif R (1973) Different cyclic AMP requirements for induction of the arabinose and lactose operons of *Escherichia coli*. *J Mol Biol* **79**: 149–162.
- Lis JT & Schleif R (1975) The regulatory region of the L-arabinose operon: its isolation on a 1000 base-pair fragment from DNA heteroduplexes. *J Mol Biol* **95**: 409–416.
- Lobell R & Schleif R (1990) DNA looping and unlooping by AraC protein. *Science* **250**: 528–532.
- Lobell RB & Schleif RF (1991) AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic amp receptor protein. *J Mol Biol* **218**: 45–54.
- Lowden M, Skorupski K, Pellegrine M, Chiorazzo M, Taylor R & Kull F (2010) Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. *P Natl Acad Sci USA* **107**: 2860–2865.
- Luger K, Mäder A, Richmond R, Sargent D & Richmond T (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.
- Lutz R, Lozinski T, Ellinger T & Bujard H (2001) Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res* **29**: 3873–3881.
- Mak A & Johnson A (1993) The carboxy-terminal tail of the homeo domain protein alpha2 is required for function with a second homeo domain protein. *Gene Dev* **7**: 1862–1870.
- Martin K, Huo L & Schleif R (1986) The DNA loop model for ara repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc Natl Acad Sci USA* **83**: 3654–3658.

- Niland P, Hühne R & Müller-Hill B (1996) How AraC interacts specifically with its target DNAs. *J Mol Biol* **264**: 667–674.
- Ogden S, Haggerty D, Stoner CM, Kolodrubetz D & Schleif R (1980) The *Escherichia coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *P Natl Acad Sci USA* **77**: 3346–3350.
- Pabo C, Krovatin W, Jeffrey A & Sauer R (1982) The N-terminal arms of lambda repressor wrap around the operator DNA. *Nature* **298**: 441–443.
- Plano G (2004) Modulation of AraC family member activity by protein ligands. *Mol Microbiol* **54**: 287–290.
- Reeder T & Schleif R (1991) Mapping, sequence, and apparent lack of function of *araJ*, a gene of the *Escherichia coli* arabinose operon. *J Bacteriol* **173**: 7765–7771.
- Rhee S, Martin R, Josner J & Davies D (1998) A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *P Natl Acad Sci USA* **95**: 10413–10418.
- Rodgers M & Schleif R (2008) DNA tape measurements of AraC. *Nucleic Acids Res* **36**: 404–410.
- Rodgers M & Schleif R (2009) Solution structure of the DNA binding domain of AraC protein. *Proteins* **77**: 202–208.
- Rodgers R, Holder N, Dirla S & Schleif R (2009) Functional modes of the regulatory arm of AraC. *Proteins* **74**: 81–91.
- Ross J, Gryczynski U & Schleif R (2003) Mutational analysis of residue roles in AraC function. *J Mol Biol* **328**: 85–93.
- Russo A, Jeffrey P, Patten A, Massagué J & Pavletich N (1996) Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**: 325–331.
- Saviola B, Seibold R & Schleif R (1998) Arm-domain interactions in AraC. *J Mol Biol* **278**: 539–548.
- Schleif R (1969) An L-arabinose binding protein and arabinose permeation in *E. coli*. *J Mol Biol* **46**: 185–196.
- Schleif R (1972) Fine-structure deletion map of the *Escherichia coli* L-arabinose operon. *P Natl Acad Sci USA* **69**: 3479–3484.
- Schleif R (1987) Why should DNA loop? *Nature* **327**: 369–370.
- Schleif R (1988) DNA looping. *Science* **240**: 127–128.
- Schleif R (1992) DNA looping. *Annu Rev Biochem* **61**: 199–223.
- Schleif R (1996) Two positively regulated systems, *ara* and *mal*. *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, 2nd edn (Neidhardt F, Curtiss R II, Ingraham J, Lin E, Magasanik B, Reznikoff W, Riley M, Schaechter M & Umberger, eds), pp. 1300–1309. American Society for Microbiology, Washington, DC.
- Schleif R (2000) Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* **16**: 559–565.
- Schleif R (2003) The AraC protein: a love–hate relationship. *BioEssays* **25**: 274–282.
- Schleif R & Favreau A (1982) Hyperproduction of AraC protein from *Escherichia coli*. *Biochemistry* **21**: 778–782.
- Schleif R & Lis JT (1975) The regulatory region of the L-arabinose operon: a physical, genetic and physiological study. *J Mol Biol* **95**: 417–431.
- Schleif R, Hess W, Finkelstein S & Ellis D (1973) Induction kinetics of the L-arabinose operon of *Escherichia coli*. *J Bacteriol* **115**: 9–14.
- Seibold R & Schleif R (1998) Apo–AraC actively seeks to loop. *J Mol Biol* **278**: 529–538.
- Siegele D & Hu JC (1997) Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *P Natl Acad Sci USA* **94**: 8168–8172.
- Sheppard E & Englesberg E (1967) Further evidence for positive control of the L-arabinose system by gene *araC*. *J Mol Biol* **24**: 443–454.
- Soisson S, MacDougall-Shackleton B, Schleif R & Wolberger C (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science* **276**: 421–425.
- Steffen D & Schleif R (1977) Overproducing AraC protein with lambda-arabinose transducing phage. *Mol Gen Genet* **157**: 333–339.
- Stoner C & Schleif RF (1983a) The *araE* low affinity L-arabinose transport promoter: cloning, sequence, transcription start site and DNA binding sites of regulatory proteins. *J Mol Biol* **171**: 369–381.
- Stoner C & Schleif RF (1983b) Transcription start site and induction kinetics of the AraC regulatory gene in *Escherichia coli* K-12. *J Mol Biol* **170**: 1049–1053.
- Tai A, Chuang J, Bode C, Wolfrum U & Sung C (1999) Rhodopsin's carboxy-terminal cytoplasmic tail acts as a membrane receptor for cytoplasmic dynein by binding to the dynein light chain Tctex-1. *Cell* **97**: 877–887.
- Tang S, Fazelinia H & Cirino P (2008) AraC regulatory protein mutants with altered effector specificity. *J Am Chem Soc* **130**: 5267–5271.
- Ter Haar E, Harrison S & Kirchausen T (2000) Peptide-in-groove interactions link target proteins to the β -propeller of clathrin. *P Natl Acad Sci USA* **97**: 1096–1100.
- Timmes A, Rodgers M & Schleif R (2004) Biochemical and physiological properties of the DNA binding domain of AraC protein. *J Mol Biol* **340**: 731–738.
- Tobin J & Schleif R (1987) Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. *J Mol Biol* **196**: 789–799.
- Ullmann A (1992) Complementation in β -galactosidase: from protein structure to genetic engineering. *BioEssays* **14**: 201–205.
- Ullmann A, Jacob F & Monod J (1967) Characterization by *in vitro* complementation of a peptide corresponding to an operator proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *J Mol Biol* **24**: 339–343.
- Vetter I & Wittinghofer A (2001) The guanine nucleotide-binding switch in three dimensions. *Science* **294**: 1299–1304.
- Weldon JE, Rodgers ME, Larkin C & Schleif RF (2007) Structure and properties of a truly apo form of AraC dimerization domain. *Proteins* **66**: 646–654.

- Wickstrum J & Egan S (2004) Transcription activators RhaS and RhaR. *J Bacteriol* **186**: 6277–6285.
- Wu M & Schleif R (2001a) Mapping arm–DNA-binding domain interactions in AraC. *J Mol Biol* **307**: 1001–1009.
- Wu M & Schleif R (2001b) Strengthened arm dimerization domain interactions in AraC. *J Biol Chem* **276**: 2562–2564.
- Zabin I (1982) β -Galactosidase α -complementation: a model for protein–protein interaction. *Mol Cell Biochem* **49**: 87–96.
- Zhang X & Schleif R (1998) Catabolite gene activator protein mutations affecting activity of the *araBAD* promoter. *J Bacteriol* **180**: 195–200.
- Zhang X, Reeder T & Schleif R (1996) Transcription activation parameters at *ara pBAD*. *J Mol Biol* **258**: 14–24.