Regulation of the L-arabinose operon of *Escherichia coli*

Over forty years of research on the L-arabinose operon of *Escherichia coli* have provided insights into the mechanism of positive regulation of gene activity. This research also discovered DNA looping and the mechanism by which the regulatory protein changes its DNA-binding properties in response to the presence of arabinose. As is frequently seen in focused research on biological subjects, the initial studies were primarily genetic. Subsequently, the genetic approaches were augmented by physiological and then biochemical studies. Now biophysical studies are being conducted at the atomic level, but genetics still has a crucial role in the study of this system.

In the 1960s, from a series of classic genetic experiments Englesberg and co-workers[1–4] concluded that the arabinose operon in *Escherichia coli* might not be regulated by a negative control mechanism like those that had been found in the lambda phage and *lac* operon systems. Those initial genetic findings indicated the existence of positive regulation and stimulated more extensive genetic analyses of the arabinose gene system[5]. These were followed by biochemical investigations that confirmed the existence of a positive regulatory mechanism[6]. Since then, positive regulation has been found in many regulation systems, both prokaryotic and eukaryotic, and even in the *lac* and lambda phage systems. Continued investigation of the arabinose operon discovered DNA looping[7], a mechanism now known to be widely used in gene regulation. Most recently, the details by which the regulatory protein of the arabinose operon, AraC protein, responds to arabinose have been elucidated at the molecular, if not the atomic scale[8]. This mechanism, called the light-switch mechanism, involves ligand regulation of the position of an arm of the protein. Because the mechanism is relatively simple, it is appealing to consider that future attempts at engineering regulation by arbitrary ligands will be based on the light-switch mechanism in other proteins and enzymes. The arabinose system, however, is already of practical use in protein expression systems, because the *ara* promoter provides high levels of induced expression and low levels of uninduced expression.

The arabinose system: genes and behavior

The arabinose system enables *E. coli* and its relatives to take up the pentose L-arabinose from the growth medium using products of the unlinked *araE* and *araFGH* genes, and then convert intracellular arabinose in three steps catalyzed by the products of the *araBAD* genes to D-xylulose-5-phosphate (Fig. 1). This then enters the pentose phosphate shunt. AraC protein regulates expression of its own synthesis[9] and the other genes of the *ara* system. In the presence of arabinose, AraC stimulates initiation of mRNA synthesis from the promoters p_E, p_{FGH}, p_{BAD} (Ref. 10) and p_J, a promoter serving a gene of unknown function[11]. At p_{BAD}, the AraC protein not only acts positively to stimulate transcription in the presence of arabinose, but also acts negatively in the absence of arabinose to repress transcription initiation[3]; whereas at p_C, AraC acts negatively in the presence or absence of arabinose[9].

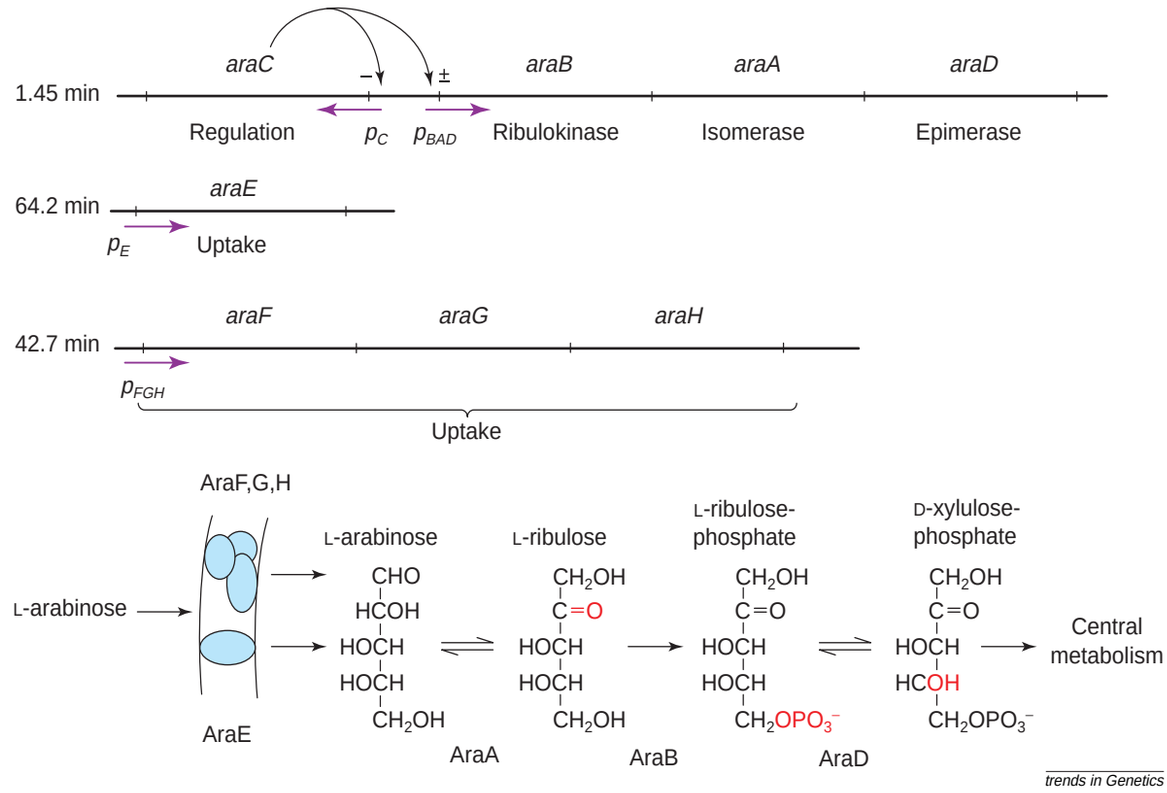
Careful experiments using arabinose to induce and rifamycin to block further mRNA synthesis have been carried out with special rifamycin-permeable cells. These show that initiation of *araBAD* mRNA synthesis begins within three seconds of the addition of arabinose[12]. Experiments on the unlinked *araE* and *araFGH* genes lack such time resolution, but show that approximately five minutes are required for full induction of these genes. Ultimately, the protein products of the *araBAD* genes are induced to approximately 300 times their uninduced level by AraC. Induction of the genes also shows 'catabolic sensitivity', a term resulting from the fact that inducibility is diminished in the presence of glucose and some other sugars. This sensitivity is mediated by the amount of cyclic AMP, which, in turn, modulates the activity of the cyclic AMP receptor protein, CRP, which is required in addition to AraC for full induction of the arabinose genes[6,13].

D-fucose (5-methyl-L-arabinose), a close structural analog of L-arabinose, blocks the growth of cells on arabinose. Fucose binds to AraC but does not normally activate transcription. The growth block induced by fucose is a geneticist's dream, as the isolation of fucose-resistant mutants is particularly simple. The fucose-resistant

Robert Schleif
bob@gene.bio.jhu.edu

Biology Dept, Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218, USA.

FIGURE 1. The L-arabinose operon



The genes required for the uptake and catabolism of L-arabinose in *Escherichia coli*, the operon structures, their approximate locations on the 100-minute circular genetic map, and the initial steps of arabinose catabolism showing the structures of the intermediates and the steps catalyzed by the products of the *ara* operon gene products. AraC acts positively and negatively at p_{BAD} (±), and negatively at p_C (−).

mutants lie in AraC, and some make the system constitutive; that is, the mutant AraC protein then activates transcription even in the absence of arabinose. Other mutants are activated not only by arabinose, but also by fucose[14]. Fucose-resistant mutants were critical in proving that the *ara* system is positively regulated.

How the *ara* system works

AraC protein functions as a homodimer. The monomer possesses two domains, a dimerization domain that also binds arabinose and a DNA-binding domain (Fig. 2a). These domains are loosely connected by a flexible linker that allows the dimeric protein (in the presence of arabinose) to bind to half-sites in their natural direct-repeat orientation separated either by four bases, the natural spacing found in the I_1 and I_2 or O_{1L} and O_{1R} elements, or by an additional 10 or 21 bases (one or two helical turns of the DNA) in artificial constructs[15]. The flexible linker also permits the protein to bind to inverted half-sites. In the absence of arabinose, an N-terminal arm of ~18 amino acids extends from the dimerization domain and binds to the side of the DNA-binding domain[8] away from the DNA, the 'back side'. The combination of the arm plus the linker holds each DNA-binding domain relatively rigidly to its dimerization domain. Consequently, as shown in Fig. 2b, the DNA-binding domains are then well orientated for binding to the I_1 and O_2 half-sites and forming a DNA loop, and are completely misorientated for binding to adjacent half-sites. Binding to adjacent half-sites (i.e. binding *cis*) would require substantially bending AraC or breaking at least one of the arm–DNA-binding domain interactions. Hence, it is energetically disfavored for AraC to bind to I_1–I_2 in the absence of arabinose, but energetically favored for AraC bind to nonadjacent half-sites and form a DNA loop.

Upon the binding of arabinose to the dimerization domains, it is energetically more favorable for the N-terminal arms of AraC to bind to the dimerization domains than to the DNA-binding domains (Fig. 2c). The arabinose-mediated release of the arms from their interactions with the DNA-binding domains relaxes the constraints holding them, and the domains are then more free to reorientate and assume any relative orientation. Most importantly, the now flexible AraC allows the overall energy state of the system of DNA–AraC to be lower if AraC binds to the adjacent I_1 and I_2 half-sites rather than looping between I_1 and O_2. As an aside, X-ray crystallography shows that arabinose induces only very small changes in the structure of the core of the dimerization domain[16], so it appears probable that the major determining factor of the arms' locations are the direct interactions between arabinose and the arms themselves.

The DNA loop that is formed in the absence of arabinose accomplishes several things for the arabinose system. First, its presence sterically blocks access of RNA polymerase to the p_{BAD} promoter[17], thus holding the basal level of p_{BAD} expression at a low level. The loop also blocks access of RNA polymerase to the p_C promoter, and might

also hinder the binding of the cyclic AMP receptor protein to its binding site alongside AraC. Looping to O_2 actively keeps AraC from occupying the I_2 half-site and inappropriately activating transcription from p_{BAD}. Finally, having AraC bound at the regulatory region in the absence of arabinose allows it to respond very quickly to the addition of arabinose as the potentially slow DNA-binding step has already taken place.

Neither the shift of the arms from the DNA-binding domains to the dimerization domains nor the shift of the DNA-binding domains from looping to binding *cis* requires directed movement of the arms or domains. Random diffusional motion should suffice. Because the arm is in equilibrium and continually dissociates from and reassociates to the two domains, it samples the arabinose occupancy of the dimerization domain. Overall, in the absence of arabinose, the arm spends most of its time bound to the DNA-binding domain, and in the presence of arabinose, most of the time bound to the dimerization domain. Diffusion also allows the DNA-binding domain that was bound at the O_2 half-site to shift position to the I_2 half-site and to induce p_{BAD} within a few seconds of the appearance of arabinose.

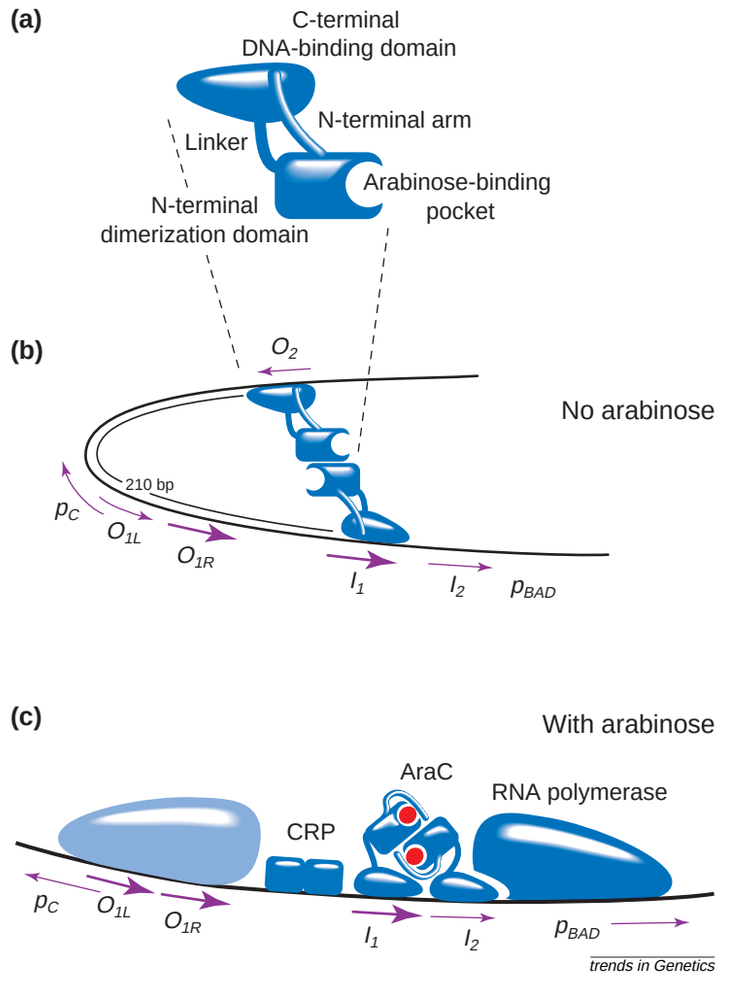
One phenomenon yet to be explained is the transient derepression of the p_C promoter following arabinose addition[18]. The activity of this promoter increases about tenfold shortly after the addition of arabinose, but returns to its pre-induction level about ten minutes later. This behavior is consistent with the hypothesis that opening the DNA loop upon the addition of arabinose allows RNA polymerase free access to the p_C promoter until AraC can bind at the O_1 site. AraC might then bind to O_1 slowly because its concentration free in the cytoplasm is low as most AraC could have bound nonspecifically to random DNA sequences as a result of the addition of arabinose. This hypothesis nicely explains a benefit of DNA looping – it permits rapid induction kinetics of the arabinose catabolic enzymes.

The AraC family

The first homologs of AraC found were the two regulatory proteins RhaR and RhaS of the rhamnose operon[19]. Study of this system began when it became clear that the insolubility and instability properties of AraC (and most of the AraC/XylS family members) hindered biochemical studies of the protein and its regulation mechanism. Since the discovery of these initial homologs of AraC, >100 additional bacterial regulatory proteins have been found that contain regions homologous to the DNA-binding domain of AraC (Ref. 20). The group is now called the AraC/XylS family. The homology between the DNA-binding domain of AraC and the other family members generally is a little less than 20%. Because, however, the homology extends over a major part of the proteins and they possess similar activities, they almost certainly possess similar tertiary structures. The similarity between the structures of the dimerization domains in the family members other than AraC is much less certain because their sequences show less similarity to one another. Members of the family possess two regions with homology to the helix-turn-helix DNA-binding motif that was first found in CRP and lambda phage repressor.

Two monomeric homologs of the AraC DNA-binding domain, MarA (Ref. 21) and Rob (Ref. 22), have been crystallized while bound to DNA and their structures determined by X-ray crystallography. As expected, their structures are nearly identical and contain two helix-turn-helix regions that contact DNA. This is compatible with the fact that a dimer of AraC contacts four adjacent major-groove regions on the DNA. The structure of the dimerization domain of AraC has also been determined, both in the presence and absence of arabinose[16], and in the presence of fucose[23]. The dimerization domain forms a pocket from β-sheets that binds arabinose, and dimerizes by an antiparallel coiled-coil. No other structure a AraC/XylS family member has been reported.

FIGURE 2. Regulation of the L-arabinose operon by arabinose

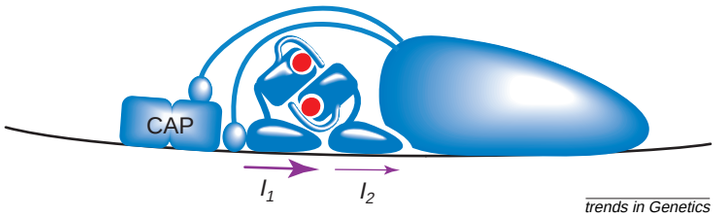


The domain structure of one subunit of the dimeric AraC protein (a) and the p_C and p_{BAD} regulatory regions in the absence (b) and presence (c) of arabinose. The regulatory elements O_2, I_1 and I_2 are 17-bp half-sites of similar sequence that each bind to one subunit of AraC, and O_1 is formed of two half-sites, O_{1L} and O_{1R}, that bind two subunits. In the absence of arabinose, RNA polymerase is hindered from binding to p_{BAD} and to p_C. The cyclic AMP receptor protein, CRP, is probably similarly hindered from binding to its DNA site. In the presence of arabinose, AraC binds primarily to the adjacent I_1 and I_2 half-sites instead of looping. Consequently, RNA polymerase has free access to p_{BAD} and CRP is free to bind as well. At p_C and O_1 RNA polymerase and AraC compete for binding.

Positive regulation: variations on the standard theme

Early on, genetic data indicated that the *ara* system was different from the other well-characterized regulation systems in that AraC actively turned on expression of *ara* genes when arabinose is present, rather than actively

FIGURE 3. RNA polymerase interactions of p_{BAD}



Probable contacts between the C-terminal domain of the α-subunit of RNA polymerase and the CAP protein and AraC protein and DNA. The α-subunit is depicted as contacting the DNA lying between the CRP- and AraC-binding sites because the α-subunit often does contact DNA, but no experimental data supporting such a DNA contact in the *ara* system have been reported. Additional contacts are made between RNA polymerase and the polymerase proximal subunit of AraC.

turning them off when arabinose is absent[5]. Because what appears to be a positive control system can merely be two negative control systems acting in series, and because negative regulation was known and accepted, it was of considerable importance to prove definitively that AraC was a positive regulator. A coupled transcription–translation system made the proof possible[6], showing that arabinose operon proteins could be synthesized under control of added AraC protein. Further, the addition of a mutant fucose-resistant AraC protein produced a system response matching the 'phenotype' of the added mutant protein.

With the understandings that came from dissection of the transcription initiation process on 'simple' promoters[24], the mechanisms by which positive regulators could work became more clear. Positive regulators could stimulate RNA polymerase binding or the formation of transcriptionally competent 'open' complexes of RNA polymerase and DNA. Indeed, extensive studies of the mechanisms by which CRP (Ref. 25) and lambda repressor functioned showed that positive regulators actually do increase binding of RNA polymerase to promoters and/or accelerate conversion of a closed RNA polymerase–DNA complex to an open complex capable of immediately initiating transcription. Studies on these systems also identified the important contacts between RNA polymerase and the activating proteins. The contact sites on RNA polymerase include the σ-factor, the N-terminal portion of the α-subunit, which forms part of the core RNA polymerase structure, and the C-terminal domain of the α-subunit, which is a small domain connected to the N-terminal domain by a linker of about ten amino acids.

To simplify measurement of the binding and open-complex-formation rates of RNA polymerase at p_{BAD}, the DNA-migration retardation assay, which had proven so useful in biochemical and biophysical studies of AraC[26], was adapted to detect open-complex formation[27]. With it, AraC was found to stimulate both the binding of RNA polymerase to DNA and the rate of open-complex formation. Various footprinting experiments carried out in conjunction with these studies also showed that RNA polymerase alters the DNA contacts made by the more upstream subunit of the homodimeric AraC, rather than those made by the subunit that binds near to the RNA polymerase –35 region. Altogether, the data suggest that RNA polymerase probably makes contact with both subunits of AraC, as well as with CRP. This is supported by the determination of residues in CRP that affect induction of p_{BAD} without affecting DNA binding by CRP (Fig. 3). These residues lie in a region called activation region three (AR3) of CRP (Ref. 28).

With 100 or more transcription regulators related to AraC, it would seem that the residues of AraC and the other family members that interact with RNA polymerase, and hence are essential for transcription activation, would be apparent upon sequence alignment. Unfortunately, this is not the case, although several highly conserved residues lying just beyond the second helix-turn-helix region are good candidates[20]. Not only does the alignment not pinpoint any candidate residues involved in the interaction, but extensive mutant hunts of AraC have also failed to reveal the residues involved in the contacts. It is possible then that any one of several residues suffice for activation of RNA polymerase, that the details of the interaction between different family members and RNA polymerase differ, or that the residues of AraC that contact RNA polymerase also participate in some other essential function of AraC. For example, a residue might be used by AraC to make a sequence-specific DNA contact and also contact RNA polymerase (although perhaps not at the same time).

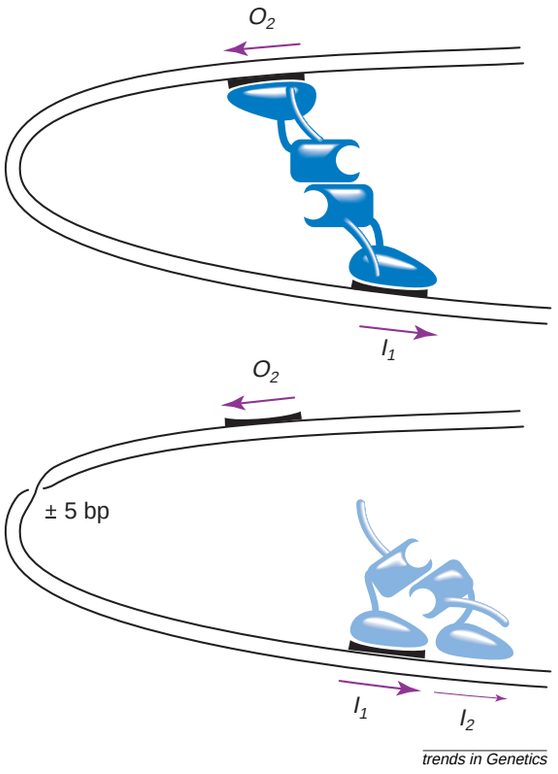
DNA looping: discovery, proof and biological use

DNA looping was discovered in classic textbook fashion. Englesberg had published data indicating that a particular deletion entering the p_{BAD} regulatory region from upstream caused the loss of repression by AraC, although the promoter could still be induced normally[3]. The possibility that a negative control element might lie upstream of all the elements required for positive control of p_{BAD} led to large-scale genetic mapping of the operon[29]. Unexpectedly, this work indicated that the negative control element, now known as O_2, might lie hundreds of base pairs upstream from all the sequences that were required for positive control[30]. This was then proven with the discovery and demonstration of DNA looping[7]. Looping of the DNA by AraC between a site within p_{BAD} and a site well upstream provided an explanation for how a negative regulatory element could act from far upstream of a promoter.

The critical test of DNA looping was the helical-twist experiment in which half a turn of the DNA helix was added to the sequence anywhere between upstream site required for full repression of p_{BAD} and the downstream site required for induction[7] (Fig. 4). Such a half-turn rotates one of the two sites to which AraC binds to the opposite side of the DNA, greatly hindering loop formation. Such an experiment can indicate the existence of DNA looping if the binding energies of AraC to the two half-sites are not so great that the excess binding energy is sufficient to twist the DNA and return the half-sites to the same face. The introduction of half-rotations diminished repression of p_{BAD}, whereas the introduction of integral numbers of rotations did not. This not only provided strong support for the looping idea, but also allowed determination of the helical twist of DNA *in vivo*[31]. Also consistent with the conclusion that relatively little energy is available for DNA looping in the *ara* system was the finding that the upper boundary for looping and binding *trans* by AraC in the absence of arabinose is only a little larger than the natural loop size of 210 base pairs (i.e. ~500 base pairs). The corresponding lower boundary is not known precisely, but is less than ~100 base pairs[31].

The hypothesis of DNA looping made the key prediction that the *I* site would be occupied by AraC even in the

FIGURE 4. The helical-twist experiment that demonstrated DNA looping



Introduction of half-integral turns between the O_2 and I_1 half-sites interfered with repression of p_{BAD} in the absence of arabinose giving rise to a five- to tenfold elevation of the basal level of transcription. Introduction of integral numbers of turns did not interfere with this repression.

absence of arabinose. Until this time it was thought that gene regulation was accomplished solely by regulating the binding or dissociation of a protein from DNA, and the prediction that AraC might remain bound to part of the *I* site in the absence and presence of arabinose was unanticipated. *In vivo* footprinting by dimethyl sulfate was developed to test this crucial idea[32]. Through good fortune, this technique works well because AraC bound to the I_1 and the more distal O_2 half-sites strongly enhances the reaction of dimethyl sulfate with a guanine in the sites. With *in vivo* footprinting, it was possible to show that the O_2 half-site and the I_1 half-sites are both occupied in the absence of arabinose and that, upon arabinose addition, occupancy of the O_2 half-site decreases. One of the most important supportive pieces of data from the footprinting is the fact that mutating the I_1 half-site so that AraC cannot activate transcription from p_{BAD} also abolishes binding of AraC to the O_2 half-site. That is, there is a strong cooperativity in the binding of AraC to I_1 and O_2, as is required by the DNA-looping hypothesis.

Until it was possible to replicate DNA looping *in vitro*, it was impossible to know the stoichiometry of AraC involved with looping. Careful experiments had shown that a dimer of AraC binds to linear DNA at the I_1–I_2 site[33]. Therefore, it had seemed plausible that DNA looping involved four subunits, two bound at *I* and two bound at O_2. *In vitro* DNA-looping experiments with supercoiled minicircles showed that, instead, a dimer of AraC loops the DNA[34]. Hence, one monomer contacts O_2, and one monomer contacts I_1. The *in vitro* DNA-looping experiments also showed that DNA looping in the *ara* system does not readily occur unless the DNA is supercoiled. Apparently, the energies available for DNA looping are insufficient to form the loop unless the DNA already is semi-looped because of the presence of supercoiling tension. The low energies involved, of course, were the key to the success of the helical-twist experiments. The low energies are also appropriate to a regulatory mechanism because a 100-fold shift in the equilibrium state of the *ara* regulatory system requires only a small energy input, approximately that resulting from a change of one hydrogen bond in water or a couple of typical van der Waals interactions.

After the discovery of DNA looping in the *ara* system, looping was found to occur in a number of other prokaryotic systems, including *lac*, *deo*, *gal* and *gln* (Ref. 35). DNA looping then became the accepted way to explain how eukaryotic enhancers could act from a distance. DNA looping accomplishes several things for cells. First, it allows multiple proteins to affect RNA polymerase and the initiation process, some from adjacent sites and some from distal sites, and second, the cooperativity inherent in the use of multiple DNA-binding sites increases the effective binding constants and allows regulatory proteins to function at very low concentrations[35].

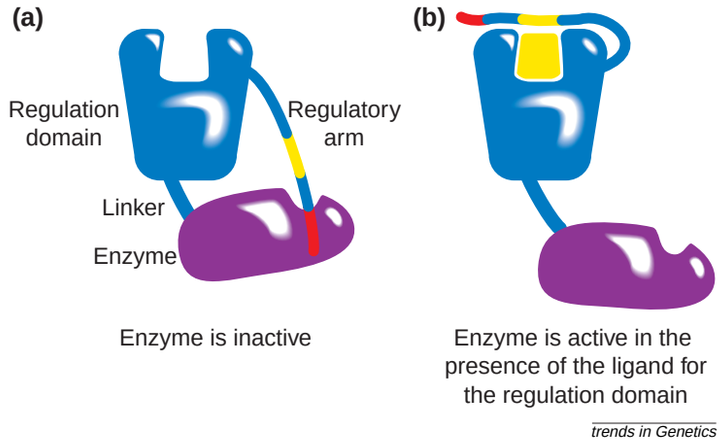
The light-switch mechanism

As described earlier, AraC responds to the presence of arabinose by shifting its N-terminal arm from the DNA-binding domains to the dimerization domains. This frees the DNA-binding domains and allows them to bind *cis* to I_1–I_2 where AraC then activates transcription from p_{BAD}. X-ray crystallography shows that in the presence of arabinose, most of the N-terminal 18 amino-acid arm of AraC is bound to the dimerization domain across the sugar bound in the central pocket of the domain (Fig. 2c)[16]. Crystallography also shows that in the absence of arabinose, electron density resulting from the arm is not resolved; that is, the arm does not occupy a single position. Although these data indicate that the presence of arabinose binds the arm to the dimerization domain, it shows nothing about the position of the arm in the absence of arabinose.

Several lines of genetic data show that the arm is bound to the DNA-binding domain in the absence of arabinose[8,36]. First, deleting the arm makes the protein act as though arabinose is present. This is as expected because without the arm, the DNA-binding domain is free, just as it would be in the presence of arabinose. Second, mutations that lie in the DNA-binding domain and make AraC poorly responsive to arabinose can be neutralized by second site mutations in the N-terminal arm.

Additional evidence for the mechanism is the existence of mutations that appear to strengthen or weaken arm interactions with the rest of the protein. A mutation that strengthens the arm–dimerization domain interaction sufficiently, allows the arm to bind to the dimerization domain and the mutant AraC then activates transcription in the absence of arabinose. Such a mutation has been found, and the basis of its stronger interaction can be understood structurally. The mutation introduces the negatively charged aspartic acid in the arm at a position that brings it very close to two positively charged residues in the dimerization domain (M. Wu, PhD thesis, Johns

FIGURE 5. Potential application of the light-switch mechanism to regulate an arbitrary enzyme



The terminal portion of the arm contains a peptide sequence that inhibits the enzyme, and the penultimate portion of the arm contains a peptide sequence that binds over the ligand when ligand is present. In the absence of ligand (a) the arm with inhibitor binds to and inactivates the enzyme, and in the presence of ligand (b), the arm binds over the ligand, thus freeing the enzyme of the inhibitory peptide sequence.

Hopkins University, 2000). Similarly, mutations that weaken the arm–DNA-binding domain interactions so that the arm does not bind at all to the DNA-binding domain should also cause the protein to be constitutively active. Mutations of this type were found by alanine and glutamic acid scanning of the surface of the DNA-binding domain (M. Wu, PhD thesis, Johns Hopkins University, 2000). The residues involved in the arm interaction lie in a line across the surface of the domain.

The mechanism used by AraC to respond to arabinose is called the light-switch mechanism by analogy to an electrical light switch. When the arm is in one position (bound to the DNA-binding domain), the system is off, and when the arm is in the other position (bound to the dimerization domain), the system is on. It is possible that the major arabinose-dependent variable determining whether or not the N-terminal arm of AraC binds to the dimerization domain is the presence or absence of the interactions between the arm and arabinose itself; that is, conformational changes in the dimerization domain of AraC surrounding the arabinose binding pocket are not of great importance. This would suggest that it might be relatively easy to append similar light-switch regulatory mechanisms onto other proteins. Arms can be engineered onto proteins, and in many cases it seems probable that peptide sequences can be found for portions of these arms using phage-display libraries such that an added arm will bind to the protein only in the presence or only in the absence of bound ligand. Additional amino acids in an arm that either activate or inhibit a protein linked to the domain would then allow allosteric regulation of the enzyme by the controlling ligand (Fig. 5). Although such systems might be constructed, they are unlikely to work without fine tuning. Obtaining the delicate balance of energies undoubtedly will require the use of powerful genetic selections.

Protein velcro: arm–domain interactions

The arm–domain interactions that are found in AraC protein provide a simple mechanism for generating interactions between domains and proteins[37]. Instead of requiring two preformed surfaces of complementary shape, an arm–domain interaction uses the structure of one of the interacting domains and a short peptide region attached to the other domain. This peptide region might have a defined shape only when it is bound to the first domain. If the arm were connected to a different protein, the two proteins could still be connected. Such an interaction is particularly convenient if any of a number of different proteins must all bind to one protein. Then it is simple to imagine that an arm that binds to the one protein can be added to each of the others. Nature indeed uses these ideas. PDZ domain proteins[38], G proteins[39,40], transcription factors[41–43], cellular sorting and transport proteins[44], and the DNA-replication machinery in prokaryotes[45] and eukaryotes all use arms for the association of one protein or a class of proteins with another protein[46]. Thus, arm–domain interactions are not confined to AraC protein and its regulation, but they are widely found in Nature. Because of their relative simplicity, we can imagine using them in the engineering of domain–domain interactions and possibly in the engineering of allosteric regulation into other proteins.

In summary, over the years, deeper and deeper study into the seemingly simple regulation system found in the arabinose operon of *Escherichia coli* has led to the discovery of DNA looping and the light-switch mechanism of allosteric regulation as well as a realization of the importance of arm–domain interactions in proteins.

References

1 Helling, R. and Weinberg, R. (1963) Complementation studies of arabinose genes in *Escherichia coli*. *Genetics* 48, 1397–1410
2 Sheppard, D. and Englesberg, E. (1967) Further evidence for positive control of the L-arabinose system by gene *araC*. *J. Mol. Biol.* 25, 443–454
3 Englesberg, E. *et al.* (1969) 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
4 Gielow, L. *et al.* (1971) Initiator constitutive mutants of the L-arabinose operon (OIBAD) of *Escherichia coli* B/r. *Genetics* 69, 289–302
5 Englesberg, E. *et al.* (1965) Positive control of enzyme synthesis by gene *C* in the L-arabinose system. *J. Bacteriol.* 90, 946–957
6 Greenblatt, J. and Schleif, R. (1971) Regulation of the arabinose operon *in vitro*. *Nat. New Biol.* 233, 166–170
7 Dunn, T. M. *et al.* (1984) An operator at –280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U. S. A.* 81, 5017–5020
8 Saviola *et al.* (1998) Arm–domain interactions in AraC. *J. Mol. Biol.* 278, 539–548
9 Casadaban, M. (1976) Regulation of the regulatory gene for the arabinose pathway, *araC*. *J. Mol. Biol.* 104, 557–566
10 Johnson, J. and Schleif, R. (1995) *In vivo* induction kinetics of the arabinose promoters in *Escherichia coli*. *J. Bacteriol.* 177, 3438–3442
11 Reeder, T. and 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
12 Hirsh, J. and Schleif, R. (1973) On the mechanism of action of L-arabinose *C* gene activator and lactose repressor. *J. Mol. Biol.* 80, 433–444
13 Haggerty, D. and Schleif, R. (1975) Kinetics of the onset of catabolite repression in *Escherichia coli* as determined by *lac* messenger ribonucleic acid initiations and intracellular cyclic adenosine 3′, 5′-monophosphate levels. *J. Bacteriol.* 123, 946–953
14 Beverin, S. *et al.* (1971) D-Fucose as a gratuitous inducer of the L-arabinose operon in strains of *Escherichia coli* B/r mutant in gene *araC*. *J. Bacteriol.* 107, 79–86
15 Carra, J. and Schleif, R. (1993) Variation of half-site organization and DNA looping by AraC protein. *EMBO J.* 12, 35–44
16 Soisson, S. *et al.* (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science* 276, 421–425
17 Hahn, S. *et al.* (1984) Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. *J. Mol. Biol.* 180, 61–72
18 Hahn, S. and Schleif, R. (1983) *In vivo* regulation of the *Escherichia coli araC* promoter. *J. Bacteriol.* 155, 593–600
19 Tobin, J. and 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
20 Gallegos, M. *et al.* (1997) AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* 61, 393–410
21 Rhee, S. *et al.* (1998) A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10413–10418

22 Kwon, H. *et al.* (2000) Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat. Struct. Biol.* 7, 424–430
23 Soisson, S. *et al.* (1997) The 1.6 Å crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *J. Mol. Biol.* 273, 226–237
24 Record, M. *et al.* (1996) In Escherichia coli *and* Salmonella *Cellular and Molecular Biology* (2nd edn) (Neidhard, F. *et al.*, eds), pp. 792–820, ASM Press
25 Busby, S. and Ebright, R. (1999) Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* 293, 199–213
26 Hendrickson, W. and Schleif, R. (1984) Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol.* 178, 611–628
27 Zhang, X, and Schleif, R. (1996) Transcription activation parameters at *ara* p_{BAD}. *J. Mol. Biol.* 258, 14–24
28 Zhang, Z. and Schleif, R. (1998). Catabolite gene activator protein mutations affecting activity of the *araBAD* promoter. *J. Bacteriol.* 180, 195–200
29 Schleif, R. (1972) Fine-structure deletion map of the *Escherichia coli* L-arabinose operon. *Proc. Nat. Acad. Sci. U. S. A.* 69, 3479–3484
30 Lis, J. and 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
31 Lee, D. and Schleif, R. (1989) *In vivo* DNA loops in *araCBAD*: size limit and helical repeat *Proc. Natl. Acad. Sci. U. S. A.* 86, 476–480
32 Martin, K. *et al.* (1986) The DNA loop model for *ara* repression: AraC protein occupies the proposed poop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Nat. Acad. Sci. U. S. A.* 83, 3654–3658
33 Hendrickson, W. and Schleif, R. (1985) A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. *Proc. Natl. Acad. Sci. U. S. A.* 82, 3129–3133
34 Lobel, R. and Schleif, R. (1990) DNA looping and unlooping by AraC protein. *Science* 250, 528–532
35 Schleif, R. (1992) DNA looping, *Annu. Rev. Biochem.* 61, 199–233
36 Reed, W. and Schleif, R. (1999) Hemiplegic mutations in AraC protein. *J. Mol. Biol.* 294, 417–425
37 Schleif, R. (1999) Arm–domain interactions in proteins: a review. *Proteins* 34, 1–3
38 Songyand, Z. *et al.* (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* 275, 73–77
39 Kostenis, E. *et al.* (1997) Genetic analysis of receptor-$G\alpha_q$ coupling selectivity. *J. Biol. Chem.* 272, 23675–23681
40 Sano, T. *et al.* (1997) A domain for G protein coupling in carboxyl-terminal tail of rat angiotensis II receptor type 1A. *J. Biol. Chem.* 272, 23631–23636
41 Kussie, P. *et al.* (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274, 948–953
42 Jabet, C. *et al.* (1999) NMR studies of the Pbx1 TALE homeodomain protein free in solution and sound to DNA: proposal for a mechanism of HoxB1–DNA complex assembly. *J. Mol. Biol.* 291, 521–530
43 Wong, K. and Geiduschek, E. P. (1998) Activator–sigma interaction: a hydrophobic segment mediates the interaction of a sigma family promoter recognition protein with a sliding clamp transcription activator. *J. Mol. Biol.* 284, 195–203
44 ter Haar, E. *et al.* (2000) Peptide-in-groove interactions link target proteins to the β-propeller of clathrin. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1096–1100
45 Shamoo, Y. and Steitz, T. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a ceptide from DNA polymerase and a polymerase editing complex. *Cell* 99, 155–166
46 Reynolds, N. *et al.* (2000) Essential interaction between the fission yeast DNA polymerase δ subunit Cdc27 and Pcn1 (PCNA) mediated through a C-terminal $p21^{Cip1}$-like PCNA binding motif. *EMBO J.* 19, 1108–1118

Towards an understanding of the genetics of human male infertility:

lessons from flies

It has been argued that about 4–5% of male adults suffer from infertility due to a genetic causation. From studies in the fruitfly *Drosophila*, there is evidence that up to 1500 recessive genes contribute to male fertility in that species. Here we suggest that the control of human male fertility is of at least comparable genetic complexity. However, because of small family size, conventional positional cloning methods for identifying human genes will have little impact on the dissection of male infertility. A critical selection of well-defined infertility phenotypes in model organisms, combined with identification of the genes involved and their orthologues in man, might reveal the genes that contribute to human male infertility.

Johannes H.P. Hackstein
hack@sci.kun.nl

Ron Hochstenbach*
p.f.r.hochstenbach@dmg.azu.nl

Peter L. Pearson*
p.l.pearson@med.uu.nl

Dept of Evolutionary Microbiology, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands.
*Dept of Medical Genetics, University Medical Center, room KC.04.084.2, PO Box 85090, NL-3508 AB Utrecht, The Netherlands

Approximately 4–5% of otherwise healthy men suffer from involuntary childlessness for which no clinical explanation can be given[1–3] (Box 1). There is a comparable spectrum and frequency of spermatogenenic defects in infertile males from the most divergent populations, arguing against the assumption that environmental influences play a major role (Table 1). Could the majority (if not all) of these various cases of unexplained (idiopathic) male infertility have a hereditary basis? If so, is the number of 'male fertility' genes and the frequency of mutant alleles sufficiently high to account for the extremely high incidence of idiopathic infertility? Based on the comparative analysis of male sterile mutations in the fruitfly *Drosophila melanogaster* and the mouse, we argue that the high incidence of human male infertility reflects a substantial genetic load in human populations due to autosomal recessive mutations.

0168-9525/00/$ – see front matter PII: S0168-9525(00)02140-5