Elucidating residue roles in engineered variants of AraC regulatory protein

Shuang-Yan Tang and Patrick C. Cirino*

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 10 September 2009; Revised 30 November 2009; Accepted 1 December 2009

DOI: 10.1002/pro.310

Published online 15 December 2009 proteinscience.org

Abstract: The AraC regulatory protein was previously engineered to control gene expression specifically in response to D-arabinose and not the native effector L-arabinose (Tang et al., J Am Chem Soc 2008;130:5267–5271). Mutations were targeted in the ligand-binding pocket and on the AraC N-terminal arm, which plays an important role in maintaining repressing or activating conformations in the absence or presence of effector, respectively. In this study, we analyze the contributions of individual mutations toward the overall mutant functions in an attempt to streamline future AraC design efforts. For a variety of point mutants, we quantify the induced expression response to D-arabinose (level of leaky expression, induction fold, half-maximal dose response, and effector specificity) and the binding affinity of the purified ligand-binding domain for D-arabinose. We find that mutations introduced in the N-terminal arm (design Position 8) strengthen the induction response, most likely by weakening interactions with the DNA-binding domain, but are not involved in ligand binding. Meanwhile, binding pocket mutations occurring further away from the arm (Positions 80 and 82) primarily contribute to maintaining repression in the absence of effector and do not show response to D-arabinose without the accompanying mutations. Combinations of mutations cooperatively couple molecular recognition to transcriptional activation, demonstrating the complexity of the AraC regulatory switch and the power of combinatorial protein design to alter effector specificity while maintaining regulatory function.

Keywords: AraC; effector specificity; ligand binding; molecular recognition; repression; regulatory switch

Introduction

The *Escherichia coli* AraC regulatory protein involved in L-arabinose metabolism controls gene expression at a number of promoters including P_{araBAD} (P_{BAD}).^{1–3} Schleif and coworkers have characterized AraC function and *ara* operon regulation and proposed the light switch mechanism depicted in Figure 1.⁴ In the absence of L-arabinose, the AraC homodimer functions as a repressor, while conforma-

tional change upon binding L-arabinose causes the dimer to act as a transcriptional activator.^{4–8} The 292-residue AraC protein is composed of two domains: the amino-terminal ligand-binding domain (LBD; residues 1–170) is responsible for dimerization and ligand (L-arabinose) binding, whereas the DNA-binding domain (DBD) is contained in the carboxy terminus.^{9,10} The three-dimensional structure of the AraC LBD in complex with L-arabinose has been reported.¹¹

In the repressing conformation, the two DBDs of an AraC homodimer contact the I1 and O2 DNA half-sites (separated by 210 bases), repressing transcription through formation of a DNA loop upstream of the P_{BAD} promoter. The stability of this conformation is largely attributed to interactions between the N-terminal arm of AraC (approximately composed of the first 18 residues) with the DBD.⁵ Upon binding

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: The National Science Foundation; Grant number: 0644678.

*Correspondence to: Patrick C. Cirino, Department of Chemical Engineering, The Pennsylvania State University, 226A Fenske Lab, University Park, PA 16802-4400.
E-mail: cirino@engr.psu.edu

Figure 1. Mechanism of dual regulation by AraC at the P_{BAD} promoter, adapted from Schleif.⁴ I1, I2, and O2 represent DNA binding half-sites. CRP is a coactivator (requiring cAMP) and RNA pol represents RNA polymerase. “E” represents the effector L-arabinose.

L-arabinose, the N-terminal arm folds over the sugar-binding pocket, thus relieving the repressing conformation and enabling the DBDs to instead bind the adjacent I1 and I2 half-sites, resulting in transcriptional activation at P_{BAD} .

Schleif and coworkers have studied the effects of various mutations on AraC functions.^{5,7,8,12,13} Residues 7–14 of the N-terminal arm of AraC are believed to make important interactions with the DBD in the absence of effector.^{5,13} The crystal structure of the LBD binding L-arabinose suggests that L-arabinose interacts directly with the arm via van der Waals interactions with residue F15 and a hydrogen bond between the main-chain carbonyl of residue P8. A network of water molecules mediates additional interactions between arabinose and the N-terminal arm.

Variants of AraC engineered to control gene expression specifically in response to small molecules of interest can find applications in molecular reporting, diagnostics, and metabolic engineering. Retention of both the repressing and activating functions of this protein is important for fully harnessing its stringent control properties. We recently described mutagenesis of the AraC L-arabinose-binding pocket to isolate mutants that are induced by D-arabinose instead of L-arabinose.¹⁴ Residue positions were selected for saturation mutagenesis based on their known involvement in AraC switching function and/or ligand binding, and mutants were selected based on their ability to regulate expression of green fluorescent protein (GFP) from promoter P_{BAD} . Two different libraries (Positions 8, 24, 80, and 82 were saturated in library A and Positions 8, 15, 24, and 80 were saturated in library B) yielded a variety of mutants having different regulatory prop-

erties (specifically, half-maximal dose response based on a whole-cell fluorescence assay, maximum induced fluorescence response, specificity toward D- versus L-arabinose, and level of leaky expression).¹⁴ We then sought to better understand residue roles in the selected mutants. Specifically, we were curious whether individual regulatory properties could be assigned to specific mutations in the isolated mutants, and whether particular mutations act cooperatively. We were also interested in knowing whether mutations in the N-terminal arm (design Position 8) contributed to ligand binding, induced switching, or both properties.

In this study, we quantify regulatory properties of mutants having combinations of mutations that were found in the originally isolated, D-arabinose-responsive AraC variants. We find that mutations in the binding pocket have a strong influence on the regulatory behavior of the N-terminal arm, and mutations introduced on the arm (Position 8) are not involved in ligand binding but do strengthen the induction response, perhaps by weakening interactions with the DBD. Meanwhile, binding pocket mutations occurring further away from the arm (Positions 80 and 82) primarily contribute to maintaining repression in the absence of effector. In general, the evolved regulatory properties arise from combinations of cooperatively acting mutations. Results from this study will help guide future efforts to alter AraC effector specificity while retaining desired regulatory features.

Results

Table I lists the mutation combinations studied. These mutations were identified from three mutants recovered from saturation mutagenesis library A (MutA1, MutA2, and MutA3) and one mutant from saturation library B (MutB1), all of which were selected based on their ability to be induced by D-arabinose and not L-arabinose.¹⁴ Using a P_{BAD} -GFP reporter system, whole-cell fluorescence was measured after induction with L- or D-arabinose and normalized by cell density. When possible, half-maximal D-arabinose dose response ($K_{d,app}$) was estimated and used to assess the relative induction sensitivity of each mutant to D-arabinose. In cases where the maximum fold increase in induced fluorescence over the uninduced state (induction fold) was less than ~3-fold, $K_{d,app}$ could not be reliably determined. Results from these fluorescence assays are summarized in Table I.

Table II summarizes the effect of mutations on specificity toward D-arabinose over L-arabinose, for those mutation combinations which still resulted in inducibility by D-arabinose (in Table I). The difference in induced expression of GFP by 100 mM D-arabinose in the absence versus presence of L-arabinose provides an indication of the extent to which

Table I. Whole-Cell Fluorescence Due to GFP Expression from Promoter P_{BAD} in Strains Expressing the Indicated AraC Variant and in the Presence of the Indicated Concentration of Effector

Variant	Mutations					Effectors			Induction fold	$K_{d, app}$ (D-arabinose)
	P8	F15	T24	H80	Y82	0 (leaky)	10 mM L-Arabinose	100 mM D-Arabinose		
Wild-type						170	58,000	210	—	—
MutA1	R		D	L	Q	2200	2500	37,000	16.8	18.9
P8R	R					2200	23,000	2600	1.2	—
MutA1-R8P			D	L	Q	2000	2000	15,000	7.5	9.8
MutA1-D24T	R			L	Q	3200	3000	2900	—	—
T24D			D			1200	3300	2600	2.2	—
MutA1-N	R		D			27,000	28,000	99,000	3.7	54.2
MutA1-C				L	Q	620	610	520	—	—
H80L				L		500	960	550	—	—
MutA1-Q82Y	R		D	L		25,000	25,000	77,000	3.1	34.2
MutA2	R		P	A	T	2200	2700	26,000	11.8	25.0
MutA2-R8P			P	A	T	4700	4500	23,000	4.9	19.9
MutA2-P24T	R			A	T	1500	1700	1600	—	—
T24P			P			3400	3200	11,000	3.2	24.2
MutA2-N	R		P			4900	5000	41,000	8.4	48.9
MutA2-C				A	T	760	690	690	—	—
H80A				A		230	230	250	—	—
Y82T					T	670	690	670	—	—
MutA2-T82Y	R		P	A		2700	2700	34,000	12.6	20.0
MutA3	D		R	Q	N	1600	1800	17,000	10.6	27.4
P8D	D					9500	27,000	14,000	1.5	—
MutA3-D8P			R	Q	N	350	210	240	—	—
MutA3-R24T	D			Q	N	2200	2300	6800	3.1	36.4
MutA3-N	D		R			21,000	21,000	26,000	1.2	—
MutA3-C				Q	N	710	700	680	—	—
T24R			R			310	370	280	—	—
MutB1	G	W	P	A		270	340	16,000	59.3	21.3
P8G	G					320	1200	760	2.4	—
MutB1-N	G	W	P			1200	1200	15,000	12.5	24.8
MutB1-G8P		W	P	A		1400	1400	1900	1.4	—

“N” and “C” represent variants carrying only the N-terminal (positions 8, 15, and 24) or C-terminal (positions 80 and 82) LBD mutations, respectively.

Values represent total cell suspension fluorescence normalized with respect to cell optical density (no correction for background fluorescence). Fluorescence of control cells not harboring GFP reporter plasmid is ~140. Induction fold was calculated as fluorescence in the presence of 100 mM D-arabinose divided by that in the absence of effector.

“—” indicates the value could not be reliably determined.

L-arabinose inhibits induction by D-arabinose (presumably by competing for binding). Although none of the mutants in Table II is induced by L-arabinose, inducibility by D-arabinose for each variant of the originally selected mutants (i.e., variants of MutA1, MutA2, MutA3, and MutB1) is significantly

Table II. Effect of L-Arabinose on Whole-Cell Fluorescence Resulting from GFP Expression (from P_{BAD}) in Strains Expressing the Indicated D-Arabinose-Responsive AraC Variant and in the Presence of 100 mM D-Arabinose (DA)

	Mutations					Without L-Arabinose			100 mM L-Arabinose		
	P8	F15	T24	H80	Y82	0 DA	100 DA	Induction fold	0 DA	100 DA	Induction fold
MutA1	R		D	L	Q	2200	37,000	17	2300	33,000	14
MutA1-R8P			D	L	Q	2000	15,000	7.5	2300	2800	1.2
MutA1-N	R		D			27,000	99,000	3.7	27,000	24,000	—
MutA1-Q82Y	R		D	L		25,000	77,000	3.1	22,000	21,000	—
MutA2	R		P	A	T	2200	26,000	12	1600	18,000	11
MutA2-R8P			P	A	T	4700	23,000	4.9	3200	6500	2.0
T24P			P			3400	11,000	3.2	2100	3800	1.8
MutA2-N	R		P			4900	41,000	8.4	2900	5600	1.9
MutA2-T82Y	R		P	A		2700	34,000	13	2400	7500	3.1
MutA3	D		R	Q	N	1600	17,000	11	1100	9100	8.3
MutA3-R24T	D			Q	N	2200	6800	3.1	2900	3600	1.2
MutB1	G	W	P	A		270	16,000	59	340	16,000	47
MutB1-N	G	W	P			1200	15,000	13	1300	1700	1.3

inhibited by 100 mM L-arabinose. These results suggest that all point mutations tested contribute to the specificity of each mutant toward D-arabinose.

In general, every mutation studied contributes significantly to at least one of the individual properties reported (in Tables I and II) for the corresponding original mutant. Furthermore, the effect of a mutation on a given property depends on the other mutations present, indicating cooperativity among mutations. As examples: (1) mutant T24D (in wild-type) and mutant MutA1-D24T show essentially no inducibility by L- or D-arabinose, whereas MutA1 is strongly induced by D-arabinose and not L-arabinose; and (2) leaky expression is low for mutants P8R, T24D, and MutA1, but it is very high for mutant MutA1-N. In contrast, there are no consistent additive properties attributed to any given mutation.

Role of Position 8 mutations

When constructing libraries for D-arabinose, we targeted N-terminal arm residue Position P8 for mutagenesis. To better understand the roles of the Position 8 mutations in the selected D-arabinose mutants (MutA1, MutA2, MutA3, and MutB1), we studied the P8 revertant of each D-arabinose mutant as well as each single point mutant (P8R, P8D, and P8G) and the role of Position 8 in variants having mutations only in the N-terminal residues (Positions 8, 15, and 24). In most mutants, replacing the Position 8 mutation with the wild-type residue (proline) significantly reduces the induction response in 100 mM D-arabinose (e.g., MutA1 compared with MutA1-R8P, MutA3 compared with MutA3-D8P, MutB1 compared with MutB1-G8P, MutA2-N compared with T24P, and MutA3-N compared with T24R). Meanwhile, MutA2-R8P shows only slightly lower GFP expression compared with MutA2, but it shows a significantly reduced fold increase in expression due to elevated leaky expression in the absence of inducer. MutB1-G8P also shows a significant increase in leaky expression (compared with MutB1) and negligible responsiveness to D-arabinose. In contrast, MutA3-D8P shows significantly lower leaky expression compared with MutA3, and no detectable response to D-arabinose.

Although the influence of P8 mutations on the phenotypes monitored is strongly dependent on the other mutations present, it is noteworthy that the value of $K_{d,app}$, when possible to estimate, did not increase upon restoring P8, and often significantly decreased (e.g., MutA1 vs. MutA1-R8P, MutA2 vs. MutA2-R8P, and MutA2-N vs. T24P). In contrast, replacing other point mutations with the corresponding wild-type amino acid generally resulted in elevated $K_{d,app}$ values (refer to Table I). Changes in GFP expression dose response do not necessarily reflect changes in affinity for D-arabinose, but this result suggests that Position 8 mutations contribute

to the strength of induced gene expression (as they were selected to do), but may not play an important role in binding D-arabinose. One explanation for this behavior is that the arm mutations cause weakened interactions between the arm and DBD, making the switch to the transcriptional activation conformation more favorable. Although the weakened interaction can result in elevated leaky expression in the absence of inducer (e.g., mutant MutA3), the mutants in this study were selected based on their ability to be induced by D-arabinose, and we attempted to eliminate mutants showing high levels of leaky expression. Thus, P8 mutants may promote weakening of arm–DBD interactions only after conformational changes resulting from effector binding. For example, the level of leaky expression for MutA1-R8P is similar to that of MutA1, indicating that weakened arm–DBD interactions due to the P8R mutation are significant only in the presence of effector. Single point mutants P8G, P8R, and P8D were all poorly induced by D-arabinose and lost selectivity toward D- over L-arabinose. In contrast, other single point mutations in the binding pocket but not within the arm eliminate the response to L-arabinose to a much greater extent than the arm mutations (T24D, H80L, T24P, H80A, Y82T, and T24R). These results support the more significant role of the arm in regulatory switching when compared with effector binding.

Inability of some P8 revertants to show significant induction (e.g., MutA3-D8P and MutB1-G8P) may be due to strengthened arm–domain interactions. However, in these cases, it is also possible that the Position 8 mutation plays a more important role in effector binding. To further study the role of P8 mutations in effector binding, we purified the LBDs of mutants MutA1, MutA1-R8P, MutA3, and MutA3-D8P (each carrying a C-terminal His₆-tag fusion) and estimated each protein's D-arabinose binding affinity (K_d). Two of the five tryptophan residues in the AraC LBD reside in the ligand-binding pocket, and measuring intrinsic tryptophan fluorescence quenching in the presence of different concentrations of ligand has been described as a means of monitoring binding.^{13,15} (Obtaining soluble, full-length AraC protein at concentrations necessary to monitor effector binding has proven difficult¹³). Note that by removing the DBD, arm–DBD interactions are eliminated. Mutation Y31V was also introduced in all mutants to improve the solubility of the LBD.¹⁶ This added point mutation caused a ~2-fold increase in the value of $K_{d,app}$ (measured from the whole-cell GFP expression response; results shown in Table S1 in Supporting Information), which is consistent with the effect of this mutation on wild-type AraC as reported by Weldon *et al.*¹⁶ As shown in Table III, the binding affinity of LBD mutant MutA1-R8P for D-arabinose is actually higher than

Table III. D-Arabinose Binding K_d for AraC LBD Mutants, Evaluated from Intrinsic Trp Fluorescence Quenching Assays

Mutants	Binding K_d (mM)
MutA1-Y31V	202 $\pm$ 14
MutA1-R8P-Y31V	85 $\pm$ 24
MutA3-Y31V	240 $\pm$ 7
MutA3-D8P-Y31V	231 $\pm$ 21

that of MutA1 (85 mM compared with 202 mM), whereas the affinity of MutA3-D8P is approximately the same as that of MutA3. These results support the conclusion from the GFP expression analyses that Position 8 mutations do not play an important role in effector binding. The similarity in D-arabinose affinity for both MutA3 and MutA3-D8P suggests that the poor GFP expression by MutA3-D8P is due to strengthened arm-DBD interactions rather than inability to bind D-arabinose. Note that the significantly larger values of K_d for these LBD proteins compared with the estimated half-maximal induction dose response concentrations are consistent with the results reported by Ross *et al.* for the case of wild-type AraC LBD binding L-arabinose ($K_d \sim 0.4$ mM).¹³ Similar results were obtained with wild-type AraC in our hands (L-arabinose $K_d \sim 0.5$ mM, compared with a half-maximal dose response ($K_{d,app}$) of ~ 10 μ M¹⁴). Meanwhile, the P8R mutant (in otherwise wild-type LBD) showed a significantly higher L-arabinose K_d of ~ 87 mM, consistent with the observed interaction between P8 and L-arabinose.¹¹ Data for wild-type and P8R LBDs binding L-arabinose are presented in Supporting Information. A possible explanation for the large difference between the half-maximal induction dose response and K_d measured from purified LBD is that the LBD structure or dimer is destabilized or otherwise altered in the absence of DBD and/or DNA binding. Alternately, the concentration of effector required to activate enough AraC to saturate expression from P_{BAD} may be far below the concentration required to saturate binding.

To verify that the observed changes in protein fluorescence in the presence of these relatively large concentrations of D-arabinose are not due to nonspecific interactions, we also measured the fluorescence quench in the same concentration of L-arabinose. MutA1 showed a negligible amount of fluorescence quench in 250 mM L-arabinose. Meanwhile, the P8 revertant showed a 48% quench in fluorescence in the presence of 160 mM L-arabinose relative to the fluorescence quench in 160 mM D-arabinose, which is consistent with the finding that P8R contributes to the specificity of this mutant for D-arabinose over L-arabinose, as demonstrated by the inhibition of GFP expression by 100 mM L-arabinose in the presence of 100 mM D-arabinose (Table II).

It is worth noting that the maximum changes in fluorescence for the LBD mutants binding D-arabinose (MutA1 and MutA1-R8P show a 35 and 25% fluorescence quench, respectively) are significantly larger than that observed for the wild-type AraC or mutant P8R LBDs binding L-arabinose (in which fluorescence is only quenched by 18 and 8%, respectively). This suggests that the D-arabinose mutant LBDs may undergo more significant structural changes upon ligand binding compared with wild-type AraC. Although this finding warrants further investigation into the effects of these mutations on LBD structure, it does not detract from the observation that the fluorescence quench for both P8 revertants does not show reduced sensitivity to D-arabinose, despite the fact that both P8 revertants show reduced strength of induction in the whole-cell assay. All results support the conclusion that the selected P8 mutations serve to enhance the induction response but not ligand binding.

Roles of N- versus C-terminal mutations

As shown in Table I, combinations of N-terminal LBD mutations (Positions 8 + 24 or 8 + 15 + 24) show responsiveness to the new effector (D-arabinose) but also contribute to elevated levels of leaky expression. Meanwhile, the C-terminal mutations (Positions 80 and 82) play a significant role in strengthening repression in the absence of effector but do not show response to D-arabinose without the N-terminal mutations. For example, compare MutA1-N to MutA1-C and MutA1-D24T (in which case the proposed weakened arm-DBD interactions due to P8R should promote induction of GFP expression if effector could bind). Similar behavior is observed for the other comparable mutants in Table I (note there are exceptions, for example, the effect of Position 80 + 82 mutations on leaky expression of mutant MutA1-D24T compared with P8R). Mutations at Positions 80 and 82, which lie far from the DBD and N-terminal arm, therefore make the AraC activation switch less favorable in the absence of effector. Possible mechanisms include indirect strengthening of arm-DBD interactions or hindrance of arm movement (in the absence of effector). The C-terminal mutations may also contribute to effector binding but only in coordination with the N-terminal mutations (i.e., cooperatively).

Discussion

In a previous study, two sets of four amino acid positions within the AraC ligand-binding pocket were selected for saturation mutagenesis, and mutants induced specifically by D-arabinose were isolated.¹⁴ In ongoing work in our laboratory, alternate sets of amino acid positions in or near the AraC-binding pocket have been randomized and resulted in a variety of mutants having novel effector specificity

(unpublished results). Directed evolution studies that aim to progressively improve a single function (e.g., stability or activity) often take advantage of additive properties contributed by iteratively inserted mutations.^{17,18} In contrast, simultaneous improvements in multiple properties (e.g., ligand binding and induction) or more drastic changes in a single property (e.g., very different substrate specificity) are more difficult to access via single mutations and may require the cooperative action of multiple point mutations, which in turn requires deeply insightful rational design via site-directed mutagenesis, or more targeted combinatorial library design followed by high-throughput screening for the properties of interest.^{18–20} We have shown that many combinations of mutations that cooperatively act to couple molecular recognition to transcriptional activation were accessed using multiple-site saturation mutagenesis, and that all point mutations examined contribute to a mutant's regulatory properties. Combinatorial methods are therefore well suited for engineering AraC while managing its complexity. However, our ability to precisely tune regulatory properties can be improved by better understanding functional roles of residues and identifying appropriate combinations of amino acid positions to target for mutagenesis.

Extensive biochemical characterization of AraC by Schleif and coworkers has demonstrated the important role of the N-terminal arm in mediating the repressing and activating properties of this regulatory protein.^{5,12,13,21} Our mutational analysis of D-arabinose-responsive AraC variants has shown that single point mutations lying on or near the arm result in elevated leaky expression, which is consistent with previous studies. Whereas P8 directly contacts L-arabinose in wild-type AraC, our study suggests that Position 8 does not play a role in effector binding for the D-arabinose mutants. Position 8 mutations do however enhance the strength of induced expression, perhaps by weakening arm–DBD interactions in the presence of effector. In general, changes to the binding pocket (via mutations and/or effector binding) can have a strong influence on the regulatory behavior of the N-terminal arm. A study by Rodgers *et al.* showed that in the absence of the LBD, the AraC arm does not have significant affinity for the DBD.²² Our study further demonstrates the network of interactions throughout the LBD that relay binding information to the arm and regulate the strength of arm–DBD interactions.

It should be stated that the possible effect of mutations on AraC structural stability have not been considered in this analysis. For example, it is possible that the observed decrease in leaky expression due to the N-terminal mutations is a result of altered solubility or structural stability, which in turn may reduce *in vivo* levels of functional protein

Figure 2. Model depicting AraC arm switching. In state “A” the N-terminal arm is contacting the DBD in the absence of effector and gene expression is turned off; in state “B” the N-terminal arm contacts the LBD in the absence of effector, resulting in leaky expression; in state “C”, AraC is binding effector (“E”) but the N-terminal arm remains in contact with the DBD and gene expression is repressed; state “D” represents induced AraC with the N-terminal arm contacting LBD in the presence of effector. ΔG_{AB} , ΔG_{AC} , and ΔG_{CD} represent free energy changes between states.

and reduce leaky expression. Conclusions drawn from this study therefore rely on the assumption that the overall mechanism of regulation at P_{BAD} by AraC is not changed relative to wild-type, that is, that the protein can repress in the absence of ligand (which it clearly does), that ligand binds in the LBD pocket, and that the N-terminal arm plays a role in mediating repression and activation. Further investigations into mutant structure, solubility, and stability are warranted, although such studies are complicated by the well-documented difficulties in obtaining pure and soluble full-length AraC protein.¹³

When designing AraC to respond to new effectors, interactions between the arm and DBD, LBD and effector should be taken into consideration. These factors are depicted in the four state model of the AraC switch, shown in Figure 2. In this model, the N-terminal arm is either contacting the DBD and therefore repressing (states A and C) or is in an activating conformation when not contacting the DBD, and instead contacting LBD and potentially effector (states B and D). Note that this model depicts an AraC monomer, while the protein actually regulates as a dimer at P_{BAD} . Both “OFF” states in this model correspond to an AraC dimer that maintains contacts with the O2 and I1 half-sites depicted in Figure 1 (resulting in DNA loop formation), whereas both “ON” states correspond to the dimer conformation that binds the I1 and I2 sites. Consider mutant MutA1 compared with MutA1-R8P in terms

of this model. Note that the level of leaky expression is essentially the same for both mutants. That is, the energetics of the A–B transition are not significantly changed because of the arm mutation. Meanwhile, the P8R mutation increases the K_d for D-arabinose, estimated from the fluorescence quench assay involving only the LBD (Table III). This reflects the energetics of effector binding through a B–D transition, as arm–DBD interactions are not a factor. These assumptions allow us to conclude that the net energy change between state “A” and state “D” ($-\Delta G_{AD}$) is less for MutA1 than for MutA1-R8P. A P8R mutation that weakens arm–DBD interactions upon effector binding corresponds to destabilizing state “C,” causing a larger value of ($-\Delta G_{CD}$). However, destabilizing “C” will also correspond to a lower value of ($-\Delta G_{AC}$). Thus, in spite of a smaller value of ($-\Delta G_{AD}$) for MutA1, the increase in ($-\Delta G_{CD}$) resulting from the “C”-destabilizing mutation (P8R) leads to a higher fraction of protein in the “ON” state for MutA1 compared with MutA1-R8P (in a given concentration of effector).

Note that ($-\Delta G_{CD}$) could also be increased by strengthening interactions between the arm and effector or LBD (stabilizing “D”). This would not cause “C” destabilization or a reduction in ($-\Delta G_{AC}$). This analysis helps to explain the relatively poor sensitivity of the D-arabinose mutants for D-arabinose compared with wild-type AraC for L-arabinose. The structure of AraC binding L-arabinose shows a direct hydrogen-bond interaction between L-arabinose and the main-chain carbonyl of P8 as well as other water-mediated interactions between effector and arm.¹¹ Lower sensitivity for the D-arabinose mutants is consistent with Position 8 mutations not interacting with effector but rather weakening arm–DBD interactions (it is still possible that other arm residues interact with D-arabinose).

Interactions between arm and effector are not compulsory for effector-induced arm switching. The importance of arm mutations or arm-effector interactions when engineering AraC for altered effector specificity depends on the desired level of sensitivity. Direct interactions between effector and arm will stabilize “D” and are more likely to yield higher affinity mutants. The challenge then is to identify arm mutations that facilitate arm–effector interactions, but do not at the same time greatly weaken arm–DBD interactions. This is especially challenging considering the much higher probability of mutations that weaken interactions rather than forming new ones. Furthermore, when screening an AraC mutant library, sufficiently low concentrations of inducer should be used so that mutants with strengthened arm–LBD interactions (provided such mutants are present!) can be differentiated from those with weakened arm–DBD interactions.

Materials and Methods

Site-directed mutagenesis

Site-directed mutagenesis was performed using a Quik-Change kit (Stratagene, La Jolla, CA). The mutations introduced were confirmed by sequencing.

GFP expression fluorescence assays

Strain HF19 harboring reporter plasmid (pPCC442)¹⁴ and wild-type or mutant AraC expression plasmid (pPCC423)¹⁴ were grown overnight in LB medium containing chloramphenicol and apramycin and 0.4 mM IPTG, then diluted to optical density (OD_{600}) = 0.01 in the same medium containing an appropriate concentration of inducer, and allowed to grow under inducing conditions for 10 h. A total of 100 μ L culture was centrifuged, and the cells were washed with 10 mM potassium phosphate buffer (pH 7.5) and resuspended in 200 μ L of the same buffer. The cell suspension OD_{600} was measured with a SPECTRAMax microplate spectrophotometer (Molecular Devices Corporation), and fluorescence emission was measured with a GENios FL fluorescence spectrometer (Tecan Austria GmbH) (360 nm excitation filter, 535/50 nm emission filter). The data were normalized with respect to OD_{600} . The background fluorescence due to buffer served as the blank in all measurements. All reported fluorescence and induction $K_{d,app}$ data represent the mean of at least three independent data points. The coefficient of variation (CV) for all the fluorescence and induction $K_{d,app}$ data reported in the tables was always less than 15% and less than 10% for most reported values.

Purification of AraC LBD

To enable direct cloning of *araC* LBD genes for construction of in-frame, C-terminal 6-His tag fusions, primers pET22b-for (5'-ctgaagctgtgtcaccaccaccaccac cactgagatccggct-3') and pET22b-rev (5'-cgcaagctgtgc gacggagctc-3') were used to amplify plasmid pET22b (Novagen). The PCR product was self-ligated, resulting in plasmid pPCC447 (the sequence was confirmed to not contain mutations introduced from PCR). Plasmid pPCC423 carrying the *araC* gene (wild-type or mutant) was digested with NdeI and HindIII, and the resulting *araC* N-terminal domain gene fragment (bases 1–543) was ligated into pPCC447 digested with the same enzymes. The resulting plasmid was pPCC448. Mutation Y31V was introduced into the AraC mutants using Quik-Change kit (Stratagene, La Jolla, CA) to improve solubility¹⁶ (not shown). The plasmid was then transformed into *E. coli* BL21 (DE3) (Novagen) for protein expression. Cells were grown at 37°C in LB medium containing 100 μ g/mL ampicillin to OD_{600} of 0.8–1.0 and then induced with 0.1 mM IPTG overnight. Cells were harvested and lysed by French

press in buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM mercaptoethanol with 0.5 mM PMSF, and 10 µg/mL each DNase I and RNase A added immediately before use. Supernatant from the lysed cells was loaded to a Ni-NTA column (Sigma-Aldrich). The column was washed with buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM mercaptoethanol, and 10 mM imidazole, and the bound protein was then eluted with buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM mercaptoethanol, and 1M imidazole.

Fluorescence measurements of arabinose binding

Purified AraC mutant LBDs were dialyzed extensively into buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, and 1 mM mercaptoethanol and further diluted in the same buffer to a concentration of 30 µg/mL. Fluorescence was measured in 10-mm quartz cuvettes continuously stirred with a magnetic stir bar. Samples were excited at 280 nm with slit widths adjusted for 1 nm spectral bandwidth, and emission was scanned between 300 and 450 nm with slit widths adjusted for 5 nm spectral bandwidth. L- or D-Arabinose stock solutions dissolved in water were added to the samples to achieve the desired concentrations. After 1 min of equilibration, fluorescence was measured on a Fluorolog 3–21 fluorescence spectrometer (Horiba Jobin Yvon, Edison NJ). Data were collected every second at 1 nm intervals (Fig. S1 in Supporting Information).

References

- Hendrickson W, Stoner C, Schleif R (1990) Characterization of the *Escherichia coli* araFGH and araJ promoters. *J Mol Biol* 215:497–510.
- Smith BR, Schleif R (1978) Nucleotide sequence of L-arabinose regulatory region of *Escherichia coli* K12. *J Biol Chem* 253:6931–6933.
- Wallace RG, Lee N, Fowler AV (1980) The araC gene of *Escherichia coli*: transcriptional and translational start-points and complete nucleotide sequence. *Gene* 12:179–190.
- Schleif R (2003) AraC protein: a love-hate relationship. *Bioessays* 25:274–282.
- Saviola B, Seabold R, Schleif RF (1998) Arm-domain interactions in AraC. *J Mol Biol* 278:539–548.
- Seabold RR, Schleif RF (1998) Apo-AraC actively seeks to loop. *J Mol Biol* 278:529–538.
- Reed WL, Schleif RF (1999) Hemiplegic mutations in AraC protein. *J Mol Biol* 294:417–425.
- Wu M, Schleif R (2001) Mapping arm-DNA-binding domain interactions in AraC. *J Mol Biol* 307:1001–1009.
- Bustos SA, Schleif RF (1993) Functional domains of the AraC protein. *Proc Natl Acad Sci USA* 90:5638–5642.
- Eustance RJ, Bustos SA, Schleif RF (1994) Reaching out. Locating and lengthening the interdomain linker in AraC protein. *J Mol Biol* 242:330–338.
- Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science* 276:421–425.
- Wu M, Schleif R (2001) Strengthened arm-dimerization domain interactions in AraC. *J Biol Chem* 276:2562–2564.
- Ross JJ, Gryczynski U, Schleif R (2003) Mutational analysis of residue roles in AraC function. *J Mol Biol* 328:85–93.
- Tang SY, Fazelinia H, Cirino PC (2008) AraC regulatory protein mutants with altered effector specificity. *J Am Chem Soc* 130:5267–5271.
- Lauble H, Georgalis Y, Heinemann U (1989) Studies on the domain structure of the *Salmonella typhimurium* AraC protein. *Eur J Biochem* 185:319–325.
- 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.
- Kelly RM, Dijkhuizen L, Leemhuis H (2009) Starch and alpha-glucan acting enzymes, modulating their properties by directed evolution. *J Biotechnol* 140:184–193.
- Lipovsek DMM, Lippow SM, Basu S, Baynes BM, Library construction for protein engineering. In: Park SJ, Cochran JR, Eds. (2010) Protein engineering and design. CRC Press, pp 83–108.
- Chen ZL, Katzenellenbogen BS, Katzenellenbogen JA, Zhao HM (2004) Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity. *J Biol Chem* 279:33855–33864.
- Chodorge M, Fourage L, Ravot G, Jermutus L, Minter R (2008) In vitro DNA recombination by L-Shuffling during ribosome display affinity maturation of an anti-Fas antibody increases the population of improved variants. *Protein Eng Des Sel* 21:343–351.
- Ghosh M, Schleif RF (2001) Biophysical evidence of arm-domain interactions in AraC. *Anal Biochem* 295:107–112.
- Rodgers ME, Holder ND, Dirla S, Schleif R (2009) Functional modes of the regulatory arm of AraC. *Proteins* 74:81–91.