

Analysis of amino acid substitutions in AraC variants that respond to triacetic acid lactone

Christopher S. Frei,¹ Zhiqing Wang,¹ Shuai Qian,¹ Samuel Deutsch,² Markus Sutter,² and Patrick C. Cirino^{1*}

¹Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas 77204

²Joint Genome Institute, 2800 Mitchell Drive Walnut Creek, California 94598

Received 16 October 2015; Accepted 4 January 2016

DOI: 10.1002/pro.2873

Published online 8 January 2016 proteinscience.org

Abstract: The *Escherichia coli* regulatory protein AraC regulates expression of *ara* genes in response to L-arabinose. In efforts to develop genetically encoded molecular reporters, we previously engineered an AraC variant that responds to the compound triacetic acid lactone (TAL). This variant (named “AraC-TAL1”) was isolated by screening a library of AraC variants, in which five amino acid positions in the ligand-binding pocket were simultaneously randomized. Screening was carried out through multiple rounds of alternating positive and negative fluorescence-activated cell sorting. Here we show that changing the screening protocol results in the identification of different TAL-responsive variants (nine new variants). Individual substituted residues within these variants were found to primarily act cooperatively toward the gene expression response. Finally, X-ray diffraction was used to solve the crystal structure of the apo AraC-TAL1 ligand-binding domain. The resolved crystal structure confirms that this variant takes on a structure nearly identical to the apo wild-type AraC ligand-binding domain (root-mean-square deviation 0.93 Å), suggesting that AraC-TAL1 behaves similar to wild-type with regard to ligand recognition and gene regulation. Our results provide amino acid sequence–function data sets for training and validating AraC modeling studies, and contribute to our understanding of how to design new biosensors based on AraC.

Keywords: molecular reporter; biosensor; directed evolution; regulatory protein; FACS; high-throughput screening; AraC; crystal structure; cooperative residues

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

Statement of significance: Ligand-induced regulatory proteins make good candidates as molecular biosensors for screening large libraries of genetic variants for biosynthesis of the corresponding ligand. Previous protein engineering efforts have shown that inducer specificity of AraC can be readily altered, allowing for design of customized biosensors based on this regulatory system. Structure and residue substitution analyses of AraC variants responding to the compound TAL will help guide further biosensor design.

Conflict of Interest: The authors of this manuscript declare no conflict of interest.

Grant sponsor: National Science Foundation (NSF); Grant number: CBET1135710.

*Correspondence to: Dr Patrick C. Cirino; Department of Chemical and Biomolecular Engineering, Biocatalysis Laboratory, S337 Engineering Building I, Houston, TX 77204-4004. E-mail: pccirino@central.uh.edu

Introduction

Transcriptional regulatory proteins induced by small molecules have emerged as useful molecular reporting tools in whole-cell screening.^{1–4} Here, the natural link between molecular recognition and gene expression is used to report the presence and production of a metabolite of interest. For cases where there is no known transcriptional regulatory protein that responds to a desired compound, an existing transcriptional regulatory protein may be engineered to exhibit altered specificity toward the compound of interest.⁵ In previous studies, we engineered the variants of the *Escherichia coli* regulatory protein AraC, natively induced by L-arabinose (L-ara), to instead specifically activate gene expression in response to D-arabinose,⁶ mevalonate,⁷ and triacetic acid lactone (TAL).⁸

TAL (4-hydroxy-6-methyl-2-pyrone) and other 2-pyrone lactones are derailment products of polyketide synthases (PKSs) and serve as precursors to many higher value products⁹; hence, a sensitive and specific TAL biosensor would be of value in optimizing polyketide producing strains. In a previous study, we isolated our TAL-responsive AraC variant by screening a combinatorial AraC library constructed by simultaneously randomizing five codons corresponding to five residues (P8, T24, H80, Y82, and H93) located within the AraC ligand-binding domain (a library of ~3.2 million variants). This AraC library was expressed in *E. coli* and TAL-induced expression of GFP from the P_{BAD} promoter was screened via multiple rounds of fluorescence-activated cell sorting (FACS), resulting in isolation of a single TAL-responsive variant, "AraC-TAL1." To our knowledge, no natural or other artificial transcriptional regulatory proteins responding to TAL or similar 2-pyrone lactones have been identified.

Selection of the five residue positions for mutagenesis was based on prior structural and mutational analyses. Crystal structures of the wild-type AraC (wt-AraC) ligand-binding domain in the absence of and in complex with L-ara were previously solved. The L-ara complexed structure revealed primary contacts between a single L-ara molecule in the ligand binding domain and residues P8, T24, R38, Y82, and H93, as well as several other residues indirectly interacting with L-ara through water-bridged hydrogen bonds.¹⁰ In addition, substantial conformational changes in the wt-AraC N-terminal arm (residues 1–18) upon ligand binding were observed.^{11,12} Substitutions at residue F15 dramatically affect the response to L-ara, resulting in constitutive and noninducible AraC variants.^{13,14} Residues P8 and L9 are believed to contribute the strongest individual interaction energy between the N-terminal arm and L-ara.¹³ Substitutions examined at residues 6–18 largely resulted in variants with loss of repressibility (i.e., constitutive), whereas substitutions at residues T24, R38, H80, and Y82 led to repressible but noninducible variants.¹⁴

With the goal of designing AraC-TAL variants that respond specifically to 2-pyrone lactones of interest (e.g., a compound reflecting altered starter- or extender-unit specificity of a PKS variant), here we aim to gain insights into molecular recognition by AraC-TAL1, and variants thereof. From additional screening of a library of AraC variants using alternate protocols, we describe the isolation and characterization of a variety of new AraC-TAL variants (each having four to five amino acid substitutions), from which patterns of amino acid substitutions were observed. Since single amino acid substitutions can dramatically alter the behavior of wt-AraC, we examined the individual and combined contributions of amino acid substitutions in AraC-TAL1 gene expression control

to determine if this variant would be subject to a similar level of rigidity. Finally, we solved the AraC-TAL1 ligand-binding domain structure by X-ray crystallography to gain further insights into the sequence-to-function relationships that may help guide further design and screening efforts to identify transcriptional regulatory proteins for new targets of interest.

Results

Isolation and analysis of new AraC-TAL clones

AraC-TAL1 was isolated after 11 rounds of FACS sorting, and during those sorts, cells were induced by TAL until late-stationary phase prior to sorting.⁸ Subsequent to that study, we optimized our AraC library screening protocol for isolating new variants responding to various small molecules (unpublished data). The new protocol includes enriching FACS endpoint populations using selections and screening in microtiter plate assays after fewer rounds of sorting, screening cells after shorter growth periods in the presence of the desired inducer ligand, and optimized cell recovery and media/growth conditions. For the case of TAL as the inducer, we discovered that different sorting strategies lead to the isolation of different TAL-responsive variants, which we discuss below. Here we describe nine new AraC-TAL variants isolated from different sorting strategies but the same AraC library as AraC-TAL1, containing NNS (N = A, T, G, C and S = G, C) sites at codon positions relative to residues P8, T24, H80, Y82, and H93 (SLib4).⁷ Library screening was based on green fluorescent protein (GFP) expression controlled by the AraC cognate promoter P_{BAD} (P_{BAD}-*gfpuv*). Our optimized screening protocols and FACS were used to screen the library as described below.

After five rounds of sorting, two distinct populations (endpoints EP1 and EP2) emerged from different sort paths, each showing enhanced expression of GFP in the presence of TAL (Fig. 1). From these endpoints, we discovered three unique TAL-responsive variants previously not isolated (AraC-TAL2, 3, and 4). Interestingly, the original AraC-TAL1 was not found in either endpoint population, despite showing a TAL response similar to those of the newly isolated clones. Only 4 out of 46 clones screened from EP1 and EP2 (23 colonies picked from each) showed a response to 5 mM TAL. Owing to this, we reasoned that these endpoint populations still retained high levels of sequence diversity and the populations required further enrichment to enhance the frequency of responding clones. To address this, we subjected each end-point population to an additional single round of selection. P_{BAD}-*bla* (β -lactamase) was integrated into the chromosome of HF19 and confers resistance to ampicillin upon AraC-mediated activation. Screening 23 clones from each of the resulting populations after selection led

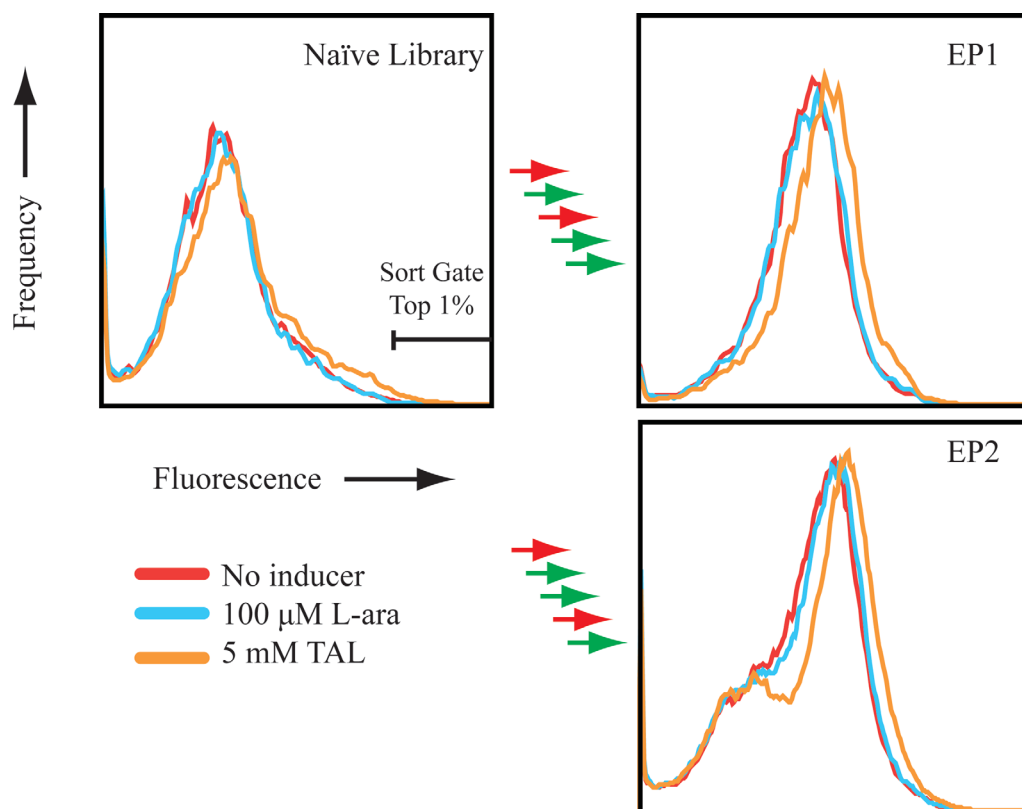


Figure 1. Histograms of flow cytometry data from endpoint populations after five rounds of FACS screening. The naïve library was sorted using two different sort schemes. First, the naïve library was screened with an initial round of negative sorting to enrich clones with functional repressibility. Four subsequent rounds of sorting (green arrows, positive sort; red arrows, negative sort) led to endpoint populations EP1 and EP2. Populations EP1 and EP2 were then subjected to further screening using selections and microtiter plate assays to isolate functional clones in the presence of 5 mM TAL.

to the discovery of six additional unique AraC-TAL variants (AraC-TAL5, 6, 7, 8, 9, and 10), and the isolation of the previous AraC-TAL1 and AraC-TAL4 variants. The amino acid substitutions of each AraC-TAL variant are reported in Table I. Note that only AraC-TAL5 was isolated from end-point population EP1, and neither AraC-TAL2 nor AraC-TAL3 was picked from either endpoint following selection. Fur-

ther optimization strategies of AraC library screening, including strategic placement of selection steps, media optimization, and gene copy number, are the topics of a forthcoming article.

We were curious as to why these new TAL-responsive clones (AraC-TAL2-10) were not isolated previously. Lower affinity for TAL and/or reduced activation may have led to these clones being

Table I. Residue Substitutions of AraC-TAL Variants.^a

Clone	Frequency ^a	Codon					Residue				
		8	24	80	82	93	8	24	80	82	93
WT-AraC		CCC	ACG	CAT	TAC	CAC	P	T	H	Y	H
AraC-TAL1	3	GTG	ATC	GGC	TTG	CGC	V	I	G	L	R
AraC-TAL2 ^b	-	GGG	CAC	CAC	AAG	CTG	G	H	H	K	L
AraC-TAL3 ^b	-	TCC	ATC	GGC	ATC	AGG	S	I	G	I	R
AraC-TAL4	5	AGC	CTG	GGC	CTC	CGC	S	L	G	L	R
AraC-TAL5	1	ATC	TTG	GGC	ATC	CGG	I	L	G	I	R
AraC-TAL6	3	GGG	TTG	CAC	AAG	GTC	G	L	H	K	V
AraC-TAL7	1	GTG	CTC	GGC	CTC	CGC	V	L	G	L	R
AraC-TAL8	1	GGG	CTG	CAC	AAG	TTC	G	L	H	K	F
AraC-TAL9	1	ACG	ATC	GGG	CTC	CGG	T	I	G	L	R
AraC-TAL10	1	GGC	CTG	GGC	ATC	CGC	G	L	G	I	R

^a Frequency with which the clone was isolated from 23 colonies screened from end-point population EP1 or EP2, following terminal selection step. Only AraC-TAL5 was isolated from population EP1.

^b Variant was isolated from previous screening, before incorporation of terminal selection step.

Table II. *AraC-TAL Variant Responses to Various Treatments*

Clone	Background (leakiness)	5 mM TAL	TAL fold response	$K_{d,app}$ ^a (mM)	Max _{1/2} ^b (mM)
wt-AraC	29 ± 6	22 ± 11	0.8	-	-
AraC-TAL1	110	2000	18.2	-	16.1
AraC-TAL2	130	1300	10.0	-	17.6
AraC-TAL3	140	1900	13.6	-	18.7
AraC-TAL4	130	1800	13.8	17.6	16.1
AraC-TAL5	140	2000 ± 500	14.3	16.5	14.9
AraC-TAL6	130	1100	8.5	-	20.1
AraC-TAL7	90	1200 ± 260	13.3	-	17.6
AraC-TAL8	150 ± 30	1500	10.0	-	16.6
AraC-TAL9	260	4100	15.8	12.9	9.1
AraC-TAL10	80	1900	24	8.9	10.0

The fluorescence per OD₅₉₅ is reported for each clone in the absence of any ligand (“Background”) or 5 mM TAL. The data were collected from three independent experiments and the averages are reported. The standard deviations were <20% of the average unless otherwise indicated. The fold-response of each clone in the presence of each ligand is reported as the fluorescence in the presence of the ligand divided by the background fluorescence.

^a $K_{d,app}$ was only calculated for clones in which the dose response reached saturation at TAL concentrations below the toxicity limit (25 mM).

^b Max_{1/2} is the TAL concentration relative to 50% of the maximum fold-response.

discarded under our previous stringent sort conditions. Therefore, we investigated the dose-dependent responses of all AraC-TAL variants (Supporting Information, Fig. S1). Above 25 mM TAL, cell growth is dramatically inhibited, preventing the measurement of a saturated response for most clones. However, each variant showed a dynamic range of response over 1–20 mM TAL and less than a two-fold response in the presence of up to 100 mM L-ara (data not shown). Fluorescence (GFP expression) in the presence of TAL (5 mM) and uninduced background fluorescence are reported in Table II. The uninduced background fluorescence (“leakiness”) of all variants is significantly greater than wt-AraC but similar to the previously isolated AraC-TAL1. AraC-TAL9 has a single amino acid substitution compared to AraC-TAL1 (V8T) and shows higher leakiness than all other variants. Induced fluorescence with AraC-TAL9 in the presence of 5 mM TAL was proportionally higher, leading to a fold-response similar to those of the other AraC-TAL variants. This result is consistent with earlier findings, suggesting that substitutions at P8 in the N-terminal arm more strongly influence repression than response to inducer.^{12,14,15} The individual dose responses to TAL, therefore, do not provide an obvious rationale for why the new variants were not previously isolated.

Despite their similar responses to TAL, two patterns of amino acid substitutions are present among the AraC-TAL variants. AraC-TAL variants 1, 3, 4, 5, 7, 9, and 10 show highly conserved amino acid sequences among the substituted residues: T24I or T24L, H80G, Y82I or Y82L, and H93R. Among these variants, AraC-TAL10 shows the highest fold-response which is associated to its low background fluorescence in the absence of TAL. A second pattern emerged from the AraC-TAL variants (AraC-TAL2, 6,

and 8) and shows weaker responses to TAL than those with the first pattern. All these variants (i) contain at least two positively charged amino acid substitutions (all others only contain one; Supporting Information, Table S2); (ii) have the least changes in substituted residue hydrophobicity (Supporting Information, Table S2); and (iii) were the only variants that do not include the substitution H93R. TAL is negatively charged at neutral pH (deprotonated at the 4-hydroxyl), and the positively charged substitution(s) may directly interact with the hydroxyl group. Meanwhile, substitutions with more hydrophobic amino acids should promote stronger interactions with the lactone ring and methyl group of TAL, as compared to the more polar pyranose ring of L-ara. A direct correlation was seen between the increase in amino acid substitution hydrophobicity and response to TAL (Supporting Information, Fig. S2), which could indicate a less specific response to molecules with hydrophobic functional groups.

In addition to TAL-responses, we looked into the specificity of each variant by measuring their response to structurally similar compounds phloroglucinol and 2,6-dimethyl-4-pyrone (Supporting Information, Fig. S3). No response to these compounds was detected (up to 25 mM for both). These results suggest selectivity toward TAL, though additional screening with TAL analogs will provide further insights into specificity and suggestions for decoy compounds in counter screens to evolve AraC variants that respond specifically to TAL or TAL analogs. Finally, it should be noted that the dose-dependent response of AraC-TAL1 to TAL was found to be unaffected by the presence of up to 10 mM L-ara (Supporting Information, Fig. S4). This suggests that L-ara does not competitively bind in the ligand-binding pocket.

Table III. Substitution Analysis of the Targeted AraC Ligand-Binding Pocket Residues

Clone	Residue					Fluorescence (rfu/OD)			Induction fold	
	8	24	80	82	93	Background (leaky)	500 μ M L-ara	7 mM TAL	500 μ M L-ara	7 mM TAL
wt-AraC	P	T	H	Y	H	72	20,500 $\pm$ 5200	77 $\pm$ 16	280	1.1
Mut10000	V	T	H	Y	H	300	13,700	330 $\pm$ 90	46	1.1
Mut01000	P	I	H	Y	H	100 $\pm$ 22	73 $\pm$ 16	100	0.7	1.0
Mut00100	P	T	G	Y	H	150	81	110	0.5	0.7
Mut00010	P	T	H	L	H	100	73	120 $\pm$ 27	0.7	1.2
Mut00001	P	T	H	Y	R	110	84	120 $\pm$ 28	0.8	1.1
Mut11000	V	I	H	Y	H	750	610 $\pm$ 170	740	0.8	1.0
Mut10100	V	T	G	Y	H	540 $\pm$ 140	530	600 $\pm$ 130	1.0	1.1
Mut10010	V	T	H	L	H	160	180	200	1.1	1.3
Mut10001	V	T	H	Y	R	450	450 $\pm$ 110	550	1.0	1.2
Mut01100	P	I	G	Y	H	90 $\pm$ 20	68	110 $\pm$ 30	0.8	1.2
Mut01010	P	I	H	L	H	100	76	120	0.8	1.2
Mut01001	P	I	H	Y	R	700	610	750	0.9	1.1
Mut00110	P	T	G	L	H	260	170	240 $\pm$ 60	0.7	0.9
Mut00101	P	T	G	Y	R	160	120	180 $\pm$ 50	0.8	1.1
Mut00011	P	T	H	L	R	120	91	130 $\pm$ 30	0.8	1.1
Mut11100	V	I	G	Y	H	1400	1300	1200	0.9	0.9
Mut11010	V	I	H	L	H	320	240	330	0.8	1.0
Mut11001	V	I	H	Y	R	280	240	290	0.9	1.0
Mut10110	V	T	G	L	H	1500	1400	1800 $\pm$ 450	0.9	1.2
Mut10101	V	T	G	Y	R	400	340	510	0.9	1.3
Mut10011	V	T	H	L	R	300 $\pm$ 80	240	360	0.8	1.2
Mut01110	P	I	G	L	H	190	130	180	0.7	0.9
Mut01101	P	I	G	Y	R	100 $\pm$ 20	68 $\pm$ 14	110	0.7	1.1
Mut01011	P	I	H	L	R	130	93	130 $\pm$ 30	0.7	1.0
Mut00111	P	T	G	L	R	170	150	230	0.9	1.4
Mut11110	V	I	G	L	H	680	670 $\pm$ 150	820	1.0	1.2
Mut11101	V	I	G	Y	R	140	96	160 $\pm$ 30	0.7	1.1
Mut11011	V	I	H	L	R	280	240	340	0.9	1.2
Mut10111	V	T	G	L	R	300	270 $\pm$ 90	1500	0.9	5.0
Mut01111	P	I	G	L	R	140	97	400	0.7	2.9
AraC-TAL1	V	I	G	L	R	380	420	5900	1.1	15.5

The 32 clones represent all combinations of residue substitutions between wt-AraC and AraC-TAL1. Variants are labeled according to the five residue positions carrying the wt-AraC ("0") or AraC-TAL1 ("1") amino acid. For example, "Mut10000" indicates that the wt-AraC residue at position P8 was changed to the AraC-TAL residue (P8V). The fluorescence, measured in relative fluorescence units per OD₅₉₅, is reported for each clone in the absence of ligand ("Background"), or in the presence of 500 μ M L-ara or 7 mM TAL. The data were collected from three independent experiments and the averages were reported. The standard deviations were <20% of the average unless otherwise indicated. The fold-increased fluorescence response of each clone in the presence of ligand is reported as the fluorescence in the presence of the ligand divided by the background fluorescence.

Amino acid substitutions in AraC-TAL variants reveal mostly cooperative interactions

Sequence analysis of AraC orthologues indicates that the amino acid substitution patterns of the ligand-binding domain in AraC-TAL variants are important to both ligand binding and the *on/off* switch.^{10,13,16} To better understand the roles of the five AraC-TAL1 amino acid substitutions and assess their potential cooperative effects, we investigated the TAL and L-ara responses of 32 AraC variants representing all combinations of wt-AraC or AraC-TAL1 residues, at the five target residues (Table III). Other than wt-AraC and AraC-TAL1, only three variants retain partial responses (>15% of wt-AraC or AraC-TAL to L-ara or TAL, respectively). Such intolerance to single substitutions in AraC-TAL1

points to cooperative interactions among these amino acids toward the gene expression response.

Weakened interactions between the AraC N-terminal arm and adjacent DNA-binding domain is expected to weaken gene repression at P_{BAD}.^{12,14,17} Variants with elevated leakiness resulting from the P8V substitution was therefore not unexpected (Table III). The induced expression response to L-ara is also much less affected by substitution P8V (Mut10000) compared to all other single substitutions (Mut01000, Mut00100, Mut00010, and Mut00001). A similar effect with a single substitution P8R was previously noted.¹⁵ The AraC-TAL1 variant with no substitution at P8 (Mut01111) also shows significantly reduced background fluorescence compared to AraC-TAL1, along with a dramatic,

Table IV. Gene Expression (Fluorescence) Response of AraC-TAL1 Variants with Single Alanine Substitutions

	Fluorescence			Induction fold	
	Background (leaky)	500 μ M L-ara	7 mM TAL	500 μ M L-ara	7 mM TAL
AraC-TAL1	400	330	4800	0.8	12.0
V8A	140	100	860	0.7	6.1
I24A	890	860	1700 $\pm$ 490	1.0	1.9
G80A	200	160	490	0.8	2.5
L82A	160	100	170	0.6	1.1
R93A	1800 $\pm$ 370	1400	1700	0.8	0.9

Fluorescence per OD₅₉₅ is reported for each clone in the absence of any ligand ("Background"), 500 μ M L-ara, or 7 mM TAL. The fold-induced fluorescence response of each clone in the presence of ligand is reported as the fluorescence in the presence of ligand divided by the background fluorescence. For cloning purposes, the ribosomal binding site (RBS) sequence upstream of these variants differs relative to that for data reported in the other tables, resulting in lower background and higher fold-induced fluorescence. The data were collected from three independent experiments and the averages are reported. Standard deviations were <20% of the average unless otherwise indicated.

though not complete, loss in induced response to TAL. Mut10111, AraC-TAL variant with a single wt-AraC substitution at residue 24, is the only other variant retaining a substantial response to TAL (>15% of AraC-TAL1 response).

In a similar analysis, we created and tested alanine-substitution variants of AraC-TAL1 to determine the contribution of each residue relative to a comparatively inert and small amino acid (rather than that of native wt-AraC amino acids). As shown in Table IV, substitution V8A retains a significant response to TAL, again supporting a stronger role of this residue in repression and arm switching, compared to ligand recognition. Alanine substitution at residue 24 also retains some response to TAL. Interestingly, variant G80A also shows response to TAL, while Mut11011 (AraC-TAL1 with Histidine at position 80) shows none. The larger histidine might crowd the binding pocket and exclude ligand binding. Finally, L82A and R93A show no response to TAL.

Structure of AraC-TAL1 ligand-binding domain reveals differences in the ligand-binding pocket

Structural determination of AraC-TAL1 ligand-binding domain in complex with TAL using X-ray crystallography was sought to illuminate details of the ligand-protein interactions. Conditions supporting crystal growth in the presence of TAL were not found. We were however able to solve the apo AraC-TAL1 ligand-binding domain structure at a resolution of 2.6 Å. Crystal parameters and refinement statistics for apo AraC-TAL1 are reported in Table V. There are three monomers in the asymmetric unit and the electron density is well defined and continuous for residues 17–168. The resulting crystal structure is shown in Figure 2(A) overlaid with apo wt-AraC. Unfortunately but not surprisingly, the structure of the N-terminal arm (residues 1–18) was not resolved. As noted earlier, this arm plays a crucial role in the transcriptional regulation of AraC.^{11,12,14,17,18}

The crystal structure of apo AraC-TAL1 is similar to those of apo and holo wt-AraC, with a root-mean-square deviation (RMSD) of 0.93 Å and 0.85 Å, respectively. The similarities to both the apo and holo structures of wt-AraC were not surprising due to the low RMSD (0.63 Å) between the two wild-type structures, differing significantly in only the N-terminal arm position. Still, the apo form of AraC-TAL1 is slightly better aligned with the holo form of wt-AraC. This could indicate that AraC-TAL1 is in a partially activated state and would help explain the elevated

Table V. Crystal Structure Data Collection and Structure Refinement Statistics

AraC-TAL1	
<i>Data collection</i>	
Wavelength (Å)	0.9765
Resolution range (Å)	39.22–2.61 (2.75–2.61)
Space group	P 3 ₁ 21
Unit cell (Å, °)	$a = 97.84$ $b = 97.84$ $c = 103.74$ $\alpha = \beta = 90$, $\gamma = 120$
No. of molecules in asymmetric unit	3
Total reflections	196,774 (28,508)
Unique reflections	17,941 (2575)
Multiplicity	11.0 (11.1)
Completeness (%)	100 (99.7)
$I/\sigma(I)$	18.6 (3.0)
R_{merge} (%) ^a	14.2 (89.0)
<i>Refinement</i>	
Reflections used in refinement	17,912
Reflections used for R -free	1785
R_{work} (%) ^b	19.8
R_{free} (%)	24.4
No. of nonhydrogen atoms	3910
RMSD bond lengths (Å)	0.004
RMSD bond angles (°)	0.64
Ramachandran favored (%)	96
Ramachandran allowed (%)	4.4
Ramachandran outliers (%)	0
Average β -factor (Å ²)	38.04

^a $R_{\text{merge}} = (\sum |I - \bar{I}| / \sum I) \times 100$.

^b R -factor = $(\sum |F_o - F_c| / \sum |F_o|) \times 100$, where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.

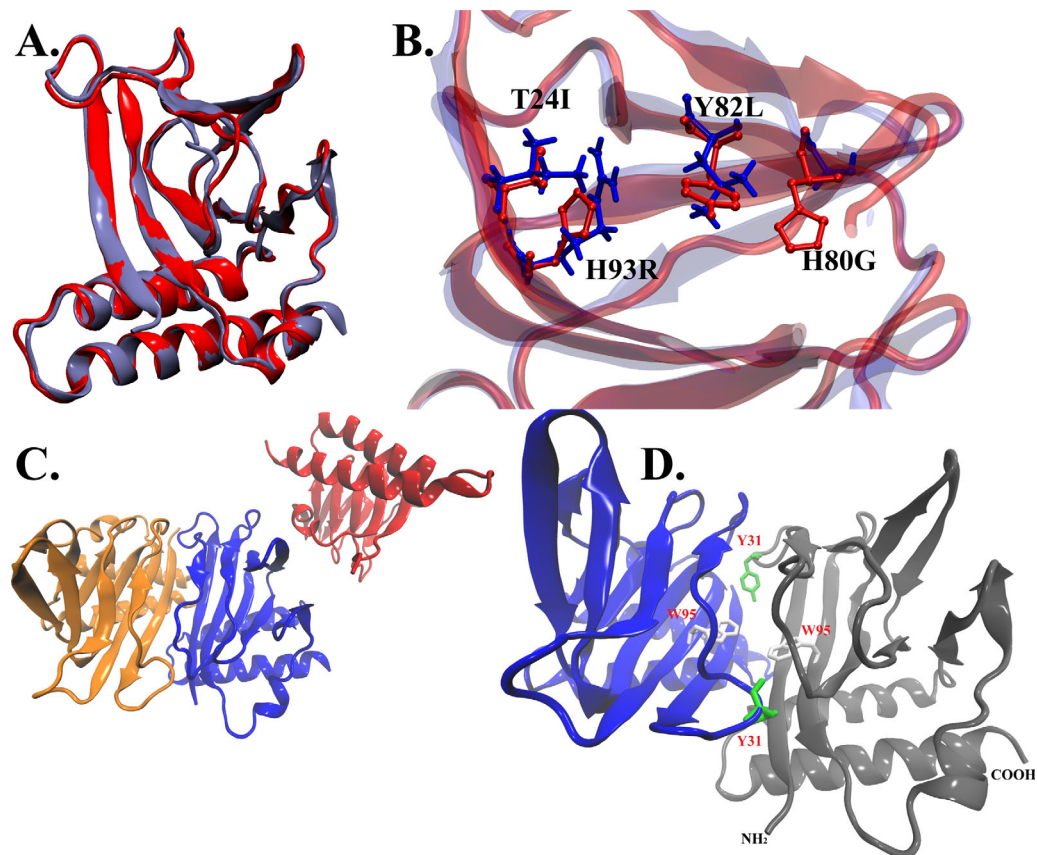


Figure 2. Comparison of AraC-TAL1 crystal structure with wt-AraC. (A) Overlay of the apo structures of wt-AraC (red) and AraC-TAL1 (blue) reveals strong similarities between the two structures. (B) The substituted residues of AraC-TAL1 (blue) are oriented similarly to the native residues of wt-AraC (red), extending into the ligand pocket. The amino acid substitution H80G shows a shift in the beta sheet $\beta 2$. (C) Each asymmetric unit of the AraC-TAL1 crystal structure contained three monomers. Two of the monomers (shown in orange and blue) appear to directly interact through the N-terminal β -barrels. (D) The β -kiss of the two monomers. Residues Y31 and W95 are highlighted, which interact in the apo form of wt-AraC, but do not form the same interaction in AraC-TAL1.

leakly expression allowed by all AraC-TAL variants. The resolved substituted residues—T24I, Y82L, and H93R (P8V is part of the N-terminal arm and was not in an ordered region of the structure)—protrude into the ligand pocket [Fig. 2(B)] and alter the binding pocket properties with minimal changes to the backbone positions. Substitution of H80G does, however, shift the position of beta sheet $\beta 2$ by 2.5 Å. Also, despite the substitution of more hydrophobic amino acids (T24I, H80G, and Y82L), the ratio of solvent-accessible surface area of hydrophobic and nonhydrophobic residues in the ligand binding domain remains relatively unchanged (1.284 and 1.281, respectively).

Two of the three monomer chains in the asymmetric unit of the ligand-binding domain formed a dimer through interactions between the N-terminal β -barrels [Fig. 2(C, D)]. wt-AraC also exhibits a dimer through a “ β -kiss” interface but only in the apo form, where a tyrosine (Y31) of the adjacent monomer fills the ligand-binding pocket and interacts with W95.^{10,19} However, the β -kiss of AraC-TAL1 is slightly rotated relative to the apo wt-AraC β -kiss, which prevents Y31 from filling the ligand pocket

and interacting with W95. Overall, the structure of apo-AraC-TAL1 suggests the regulatory mechanism is similar to that of wt-AraC but with a modified binding pocket that accepts different substrates. Resolving the N-terminal arm and structure in the presence of TAL remains an important factor in understanding the mechanism of AraC-TAL variants.

Discussion

AraC tightly regulates gene expression at promoter P_{BAD} , and the AraC- P_{BAD} regulatory system is an invaluable tool in applied molecular biotechnology and metabolic engineering. By altering AraC effector specificity, we developed this system to act as a reporter of TAL. Whereas attempts to isolate AraC variants with altered effector specificity from randomly mutated libraries were unsuccessful,⁶ simultaneously targeting multiple positions within the binding pocket yielded functional biosensor variants. Using FACS alone, we previously isolated a single TAL-responsive AraC variant. In this study, we describe a new set of TAL-responsive AraC variants that were isolated as a result of modified library

screening protocols that include the use of selection. We speculate that oversorting of the combinatorial libraries, a high degree of functional similarity between all variants, and/or inadequate endpoint population screening prevented prior identification of these new (and very similar) TAL variants.

It is worth noting that our work has sought TAL sensors that respond to “useful” concentrations of TAL, from a metabolic engineering standpoint. Screening was performed in 5 mM TAL, leading to AraC variants that are drastically less sensitive to inducer compared to wild-type AraC (which responds to micromolar L-ara). That our screening/selection strategy did not converge on a smaller subset of AraC variants is likely a direct result of this relatively low stringency (many similar solutions of comparable low sensitivity exist). On the other hand, for the case of seeking variants with greater sensitivity, increased stringency reduces the likelihood of isolating responsive variants. This is particularly challenging when the parent (wild-type AraC) shows no response even at high TAL concentrations. Moving forward, we aim to use both rational design and a directed evolution approach to engineer AraC-TAL variants that respond to TAL and TAL analogs with specificity and higher sensitivity. Such variants can serve as biosensors to report on altered substrate specificity of select polyketide synthase variants.

While it is not known to what extent the various substitutions affect each variant’s stability, overall fold, or solubility, our results collectively demonstrate nonadditive and cooperatively acting amino acid substitutions within a given variant. The crystal structure of an apo AraC-TAL ligand-binding domain variant is also shown to be nearly identical to that of wt-AraC. However, solving the crystal structure of AraC-TAL ligand-binding domain in complex with TAL remains a work in progress. Based on the structure of wt-AraC ligand-binding domain in complex with L-ara, we expect the N-terminal arm to be folded over the ligand-binding pocket in the AraC-TAL1 holo complex. This structure would help confirm the role of substitutions at position P8, which seem to be primarily involved with regulating repression. These results are consistent with previous studies indicating substitutions in the N-terminal arm weaken repression in the absence of a ligand.^{11,12,20} Understanding the orientation of TAL within the binding pocket will also help to understand molecular recognition and how to better design for selectivity. Notably, molecular docking studies with the current apo-structure of the AraC-TAL1 ligand-binding domain have been inconclusive, in that many potential TAL orientations show similar binding energies and are within the error of the energy calculations in the docking protocol (results not shown). This is not unexpected, given the relatively low sensitivity of all AraC-TAL variants to inducer TAL. However, the amino acid sequence–function data sets

provided here are useful for training and validating future AraC modeling and ligand docking studies, which in turn should help guide rational design approaches to fine-tune specificity.

Methods

General

Escherichia coli strains used in this study were MC1061 (F[−] Δ(ara-leu)7697 [araD139]B/r Δ(codB-lacI)3 galK16 galE15 λ-e14-mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2(r-m+)) for plasmid propagation, HF19, previously described by Tang *et al.*,⁶ for biosensor expression, and SQ12 for selections (described in Supporting Information). Parent plasmid vectors used in this study contained either a modified RSF1030 origin (pPCC442⁶) with high copy number (~200 copies per cell)²¹ or the pBR322 origin lacking the *rop* gene on the plasmid (pPCC423, pFG1, pFG29) which has medium copy number (30–60 copies per cell). Refer to Supporting Information for details on plasmid construction. Cultures were grown in either lysogeny broth (LB) or biosensor media (BM). BM was prepared by supplementing LB with 50 mM TES pH 7.2 and 1% glycerol. Unless otherwise stated, all cultures for FACS screening were grown at 37°C and 250 rpm in a New Brunswick Excella E25 incubator (Eppendorf; Hauppauge, NY). Flasks were all nonbaffled. Deepwell plates were purchased from Corning Life Sciences (Tewksbury, MA; Cat. No. 3960) and had square wells with conical bottoms for optimal aeration and pelleting. Cell culture optical densities (OD₅₉₅) were measured on a BMG Labtech (Ortenberg, Germany) NOVOstar with a 595 nm absorbance filter. All bulk fluorescence measurements were taken on a Molecular Devices (Sunnyvale, CA) Gemini EM Microplate Reader (Ex: 400 nm; Em: 510 nm).

TAL (4-hydroxy-6-methyl-2-pyrone) was purchased from Sigma-Aldrich (St. Louis, MO; Cat. No. H43415). The powder was directly dissolved in the appropriate medium to a final concentration of 50 mM and adjusted to pH 7.0 with 10M NaOH. The solution was sterile filtered with a 0.2 μm syringe filter and stored at 4°C for up to 1 month. L-arabinose was purchased from Sigma-Aldrich (Cat. No. A3256) and was prepared by dissolving the powder into biology grade water to a final concentration of 1 M. The stock was sterile filtered using a 0.2 μm syringe filter and stored at room temperature. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Research Products International (Mount Prospect, IL; Cat. No. I56000-25.0) dissolved in biology grade water to a final concentration of 100 mM (1000×) and sterile filtered as described above.

Fluorescence-activated cell sorting of naïve AraC library

Plasmid DNA of the naïve SLib4 library (10 ng) was transformed into 45 μ L of freshly prepared electroporation competent HF19 cells harboring pPCC442. Electroporation-competent cells were prepared according to Varadarajan *et al.*²² The 1 mL transformation outgrowths were diluted 10-fold in LB containing the appropriate antibiotic(s). The culture was then grown to an OD₅₉₅ of ~2 and diluted to an OD₅₉₅ of 0.2 in LB containing the appropriate antibiotic(s) to further dilute cells not harboring any plasmid. The subcultures were grown to an OD₅₉₅ of ~6 and diluted a final time to an OD₅₉₅ of 0.2 in BM containing the appropriate antibiotic(s) and 100 μ M IPTG. Each sample was subjected to three treatments: no ligand, 0.1 mM L-ara, and 5 mM TAL. After 6 h of growth, cells were harvested and washed once with PBS. The OD₅₉₅ and the bulk fluorescence were measured for each culture. The washed cells were diluted 1:100 in PBS and analyzed on a FACSJazz (BD Biosciences; Franklin Lakes, NJ). The appropriate cultures were then subjected to FACS. Clones were isolated by either collecting the top 1% of most fluorescent cells (positive sort) or collecting the bottom least fluorescent population relative to the bottom 99% least fluorescent cells of the wt-AraC population not induced (negative sort). For the initial two rounds of sorting, at least 10^8 cells were observed (5.5×10^7 and 1×10^6 cells were collected from the first round negative and positive sort, respectively). For all subsequent rounds of sorting, 5×10^5 cells were collected. After the sort finished, the samples were treated one of two ways: (1) samples from a negative sort were diluted 1:1 with 2 $\times$ YT medium (e.g., 20 mL of sorted sample mixed with 20 mL of 2 $\times$ YT), and half concentrations of antibiotics. Cultures were then incubated at 37°C and 250 rpm until they reached an OD₅₉₅ of ~2. They were diluted to OD₅₉₅ of 0.2 in LB containing the appropriate antibiotic(s). From here, samples were treated as described above for the subsequent round of sorting. (2) Samples from a positive sort were transferred to a centrifuge tube and treated according to Ramesh *et al.*²³ Briefly, the cells were pelleted by centrifugation at 17,900g for 10 min. The medium was discarded and the plasmid DNA was harvested using a modified protocol from a Zymo Research Zyppy Plasmid Miniprep Kit (Cat. No. D4036). The cell pellet (even if not visible) was suspended in 200 μ L of PBS. The lysis buffers were adjusted accordingly and a column from the Zymo Research DNA Clean and Concentrator Kit (Cat. No. D4005) was used to purify the plasmid DNA. Each sample was eluted from the column with 10 μ L of elution buffer. The isolated plasmid DNA was then transformed into electrocompetent HF19 cells harboring

pPCC422. The outgrowths were diluted and treated as described above for the subsequent round of sorting. After each sort, plasmid DNA was isolated from an aliquot of the subcultures prior to dilution in BM for future analysis and sequencing.

Selections and deep-well plate screening of endpoint populations

Selections of the endpoint populations were carried out as follows. Plasmid DNA (10 ng) from the endpoint population was transformed into electroporation competent SQ12 cells. The outgrowth from the transformations was diluted 10-fold and 100 μ L was plated on selection plates containing LB supplemented with 50 μ g/mL apramycin, 100 μ M IPTG, 5 mM TAL, and either 100 or 300 μ g/mL ampicillin. The plates were incubated at 37°C until colonies were easily visible on the plates (8–12 h). The colonies were scraped from the plates and the plasmid was isolated using a QIAGEN Plasmid DNA Miniprep kit. The isolated plasmid DNA was used to further screen the populations.

The plasmid DNA of the final populations was isolated and transformed into electroporation-competent HF19 cells harboring the pPCC442 reporter plasmid. Clones (23 total) were isolated from LB agar plates and streaked onto new LB agar plates. Quadruplicate 500 μ L starter cultures in 2 mL 96-well deep well (DW) plates were inoculated from each isolated clone. The starter culture was incubated for 6 h at 37°C 900 rpm in a Heidolph Titramaz 1000/Inkubator 1000. Quadruplicate 500 μ L subcultures in 96-well (DW) plate. Each sample was treated in 3 ways: (1) no ligand, (2) 100 μ M L-ara, and (3) 5 mM TAL. The cultures were inoculated to an OD₅₉₅ 0.2 from the respective starter culture. The subcultures were incubated for 6 h at 37°C 900 rpm in a Heidolph Titramaz 1000/Inkubator 1000. The cultures were washed once with 1 mL of phosphate-buffered saline and OD₅₉₅ and the bulk fluorescence were measured for each culture.

Deep-well plate dose responses

Isolated plasmid clones from the deep-well plate clone screening were digested and cloned into pFG29 as described above. The recloned mutants were transformed into electrocompetent HF19 cells. Clones were isolated from LB agar plates and quadruplicate 500 μ L starter cultures in 2 mL 96-well (DW) plates were inoculated from each isolated clone. The starter culture was incubated for 6 h at 37°C at 900 rpm in a Heidolph Titramaz 1000/Inkubator 1000. Quadruplicate 500 μ L subcultures in 96-well (DW) plate containing a range of concentrations of target ligand were inoculated by 50-fold dilutions of the respective starter culture. The subcultures were incubated for 6 h (OD₅₉₅ of ~10) at 37°C at 900 rpm. The cultures were washed once with 1 mL

of phosphate-buffered saline, and the OD₅₉₅ and the bulk fluorescence were measured for each culture.

Amino acid substitution analysis

All AraC-TAL amino acid substituted variants were cloned using Gibson Assembly.²⁴ The Gibson Assembly inserts were amplified from pFG29 and or pFG29-TAL, and the mutations were introduced by PCR primers. All PCR reactions were carried out using Phusion polymerase and the recommended PCR conditions for Phusion polymerase were followed. Annealing temperatures were obtained from NEB Tm Calculator (<http://tmcalculator.neb.com/>). The PCR fragments were purified using Zymo Research (Irvine, CA) Zymoclean Gel Recovery Kit (Cat. No. D4001). The vector was obtained by double digestion (BstAPI and AflII) of pFG29-TAL, and recovered with Zymoclean Gel Recovery Kit. Gibson Assembly was performed using Gibson Assembly Master Mix, and the recommended protocol from New England Biolabs NEBuilder (<http://nebuilder.neb.com/>) was followed. The cloning procedure is the same for alanine substitution assay, but a different template, pPCC1321, was used for PCR amplification. Plasmids pFG29-TAL and pPCC1321 are identical, but with the addition of an AvrII restriction site in the RBS region of AraC-TAL in pPCC1321.

HF19-competent cells were transformed individually with the 32 *araC/araC-TAL* mutant plasmids using electroporation and subsequently plated onto LB-agar plates supplemented with 50 µg/mL apramycin. Fresh colonies were inoculated into 500 µL LB cultures in 96-well (DW) plates supplemented with 50 µg/mL apramycin. The cultures were grown at 37°C and 900 rpm to OD₅₉₅ of 5–6 in a Heidolph Titramaz 1000/Inkubator 1000. The cultures were diluted to OD₅₉₅ of 0.2 in 200 µL BM supplemented with 50 µg/mL apramycin and 100 µM IPTG. The subcultures were incubated for 6 h at 37°C and 900 rpm in a Heidolph Titramaz 1000/Inkubator 1000. The cultures were washed once with 200 µL of phosphate-buffered saline before OD₅₉₅ and the bulk fluorescence were measured for each culture.

Protein crystallization and structure determination

Cloning: Primers pET45-araC-LBD-for and pET45-araC-LBD-rev were used to amplify the *araC-TAL* ligand-binding domain from pPCC1202. Amplified *araC-TAL* ligand-binding domain was cloned into SmaI single cut pET45b-SmaI vector by Gibson Assembly, and the resulting plasmid was named pPCC1212. Subsequently, PCR and Gibson assembly were used to remove a 6-bp DNA sequence between the enterokinase cutting site and the start codon of *araC-TAL* ligand-binding domain, resulting in plasmid pPCC1212D.

Protein Purification: BL21(DE3) cells harboring plasmid pPCC1212D were grown at 37°C overnight in LB medium with 100 µg/mL ampicillin. The overnight culture was used to inoculate five flasks of 500 mL LB supplemented with 100 µg/mL ampicillin to OD₅₉₅ of 0.05 at 37°C. At an OD₅₉₅ of ~1, cell cultures were chilled on ice for 15 min. IPTG (100 µM) was added to each subculture to induce protein expression at 15°C for 20 h. Cells were harvested and concentrated 100-fold in lysis buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM β-mercaptoethanol, and then 1 µg/mL DNase, 1 µg/mL RNase, and HaltTM protease and phosphatase inhibitor cocktail were added immediately before lysis. Lysis was performed by running the concentrated cells through three passes on a French Pressure Cell Press and subsequently centrifuged at 17,900g for 30 min. The presence of AraC-TAL1 ligand-binding-domain fragment was verified using polyacrylamide gel electrophoresis. It should be noted that the ligand-binding domain of all AraC-TAL variants show poor solubility and the presence of TAL during purification did not improve solubility (data not shown). The lysed cells supernatant was loaded onto a column containing HisPurTM Ni-NTA RseI (Thermo Scientific; Waltham, MA; Cat. No. 88221) for protein purification. The column was washed 10 times with buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM β-mercaptoethanol, and 100 mM imidazole. Target protein was eluted with buffer containing 15 mM Tris-HCl (pH 8.0), 0.1M NaCl, 5% glycerol, 1 mM β-mercaptoethanol, and 500 mM imidazole. A PD-10 desalting column (General Electric, Cat. No. 17-0851-01) was then used to remove β-mercaptoethanol and imidazole from the protein sample. The His-tag was cleaved using an Enterokinase Cleavage Capture Kit (EMD Millipore; Billerica, MA; Cat. No. 69067-3), and the digested protein sample was loaded onto an Amicon[®] Ultra-4 Centrifugal Filter Unit (EMD Millipore, Cat. No. UFC801008) for removing His-tag peptide and buffer exchange. The final protein sample was concentrated in buffer containing 10 mM Tris-HCl (pH 8.0) and 50 mM NaCl, which resulted in approximately 1 mg of purified AraC-TAL ligand-binding domain protein.

X-ray diffraction: AraC-TAL1 ligand-binding domain crystals were obtained from sitting drop experiments at room temperature by mixing 2 µL of protein solution (3 mg/mL) with 2 µL of reservoir solution containing 12–14% PEG-400, 200 mM CaCl₂ and 0.1M HEPES pH 7.0–8.0. Crystals were stabilized by adding 6 µL of reservoir solution and 6 µL of 50% glycerol solution to the drop, frozen in liquid nitrogen and measured at beam line 5.0.1 of the Advanced Light Source (ALS) at Lawrence Berkeley National Lab. Diffraction data were integrated with XDS and scaled with SCALA

(CCP4).^{25,26} The structure was solved by molecular replacement with AraC (PDB ID 2AAC). Automatic building of a model into the density using autosol was followed by manual rebuilding/refinement cycles using COOT and *phenix.refine*.²⁷ Statistics for diffraction data collection, structure determination and refinement are summarized in Table V. The resolved AraC-TAL1 ligand-binding domain crystal structure was analyzed using the software package Visual Molecular Dynamics (University of Illinois at Urbana-Champaign, IL).

Acknowledgments

We thank Dr Navin Varadarajan for his thoughtful insights in reviewing this manuscript. This work was also supported in part by the Laboratory of Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

References

- Schallmeyer M, Frunzke J, Eggeling L, Marienhagen J (2014) Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr Opin Biotech* 26:148–154.
- Dietrich JA, McKee AE, Keasling JD (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Ann Rev Biochem* 79:563–590.
- Eggeling L, Bott M, Marienhagen J (2015) Novel screening methods — biosensors. *Curr Opin Biotech* 35:30–36.
- Hansen LH, Sorensen SJ (2001) The use of whole-cell biosensors to detect and quantify compounds or conditions affecting biological systems. *Microbial Ecol* 42:483–494.
- Gredell JA, Frei CS, Cirino PC (2012) Protein and RNA engineering to customize microbial molecular reporting. *Biotech J* 7:477–499.
- Tang SY, Fazelinia H, Cirino PC (2008) AraC regulatory protein mutants with altered effector specificity. *J Am Chem Soc* 130:5267–5271.
- Tang S-Y, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Intl Ed* 50:1084–1086.
- Tang S-Y, Qian S, Akinterinwa O, Frei CS, Gredell JA, Cirino PC (2013) Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. *J Am Chem Soc* 135:10099–10103.
- Chia M, Schwartz TJ, Shanks BH, Dumesic JA (2012) Triacetic acid lactone as a potential biorenewable platform chemical. *Green Chem* 14:1850–1853.
- Soisson SM, MacDougallShackleton B, Schleif R, Wolberger C (1997) Structural basis for ligand-regulated oligomerization of AraC. *Science* 276:421–425.
- Reed WL, Schleif RF (1999) Hemiplegic mutations in AraC protein. *J Mol Biol* 294:417–425.
- Saviola B, Seabold R, Schleif RF (1998) Arm-domain interactions in AraC. *J Mol Biol* 278:539–548.
- Damjanovic A, Miller BT, Schleif R (2013) Understanding the basis of a class of paradoxical mutations in AraC through simulations. *Proteins* 81:490–498.
- Ross JJ, Gryczynski U, Schleif R (2003) Mutational analysis of residue roles in AraC function. *J Mol Biol* 328:85–93.
- Tang SY, Cirino PC (2010) Elucidating residue roles in engineered variants of AraC regulatory protein. *Protein Sci* 19:291–298.
- Schleif R (2010) AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol Rev* 34:779–796.
- Rodgers ME, Schleif R (2009) Solution structure of the DNA binding domain of AraC protein. *Proteins* 77:202–208.
- Seabold RR, Schleif RF (1998) Apo-AraC actively seeks to loop. *J Mol Biol* 278:529–538.
- 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.
- Wu M, Schleif R (2001) Strengthened arm-dimerization domain interactions in AraC. *J Biol Chem* 276:2562–2564.
- Phillips GJ, Park SK, Huber D (2000) High copy number plasmids compatible with commonly used cloning vectors. *Biotechniques* 28:400.
- Varadarajan N, Georgiou G, Cantor JR, Iverson BL (2009) Construction and flow cytometric screening of targeted enzyme libraries. *Nature Protoc* 4:893–901.
- Ramesh B, Frei CS, Cirino PC, Varadarajan N (2015) A method for direct plasmid recovery after FACS screening. *Biotechniques*, 59:157–161.
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345.
- Kabsch W (2010) XDS. *Acta Cryst D* 66:125–132.
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst D* 67:235–242.
- Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Cryst D* 68:352–367.