

Original Article

Engineering sensitivity and specificity of AraC-based biosensors responsive to triacetic acid lactone and orsellinic acid

Zhiqing Wang^{1,†}, Aarti Doshi^{2,†}, Ratul Chowdhury³, Yixi Wang¹, Costas D. Maranas³, and Patrick C. Cirino^{1,2,*}

¹Department of Chemical and Biomolecular Engineering, University of Houston, 4726 Calhoun Rd, Houston, TX 77204-4004 Houston, TX, USA, ²Department of Biology and Biochemistry, University of Houston, 3507 Cullen Blvd, Houston, TX 77204-5008 Houston, TX, USA, and ³Department of Chemical and Biomedical Engineering, Penn State University, University Park, PA 16802-4400 PA, USA

*To whom correspondence should be addressed. E-mail: pccirino@uh.edu

[†]These authors contributed equally.

Received 11 August 2020; Revised 1 October 2020; Editorial Decision 9 October 2020; Accepted 9 October 2020

Abstract

We previously described the design of triacetic acid lactone (TAL) biosensor ‘AraC-TAL1’, based on the AraC regulatory protein. Although useful as a tool to screen for enhanced TAL biosynthesis, this variant shows elevated background (leaky) expression, poor sensitivity and relaxed inducer specificity, including responsiveness to orsellinic acid (OA). More sensitive biosensors specific to either TAL or OA can aid in the study and engineering of polyketide synthases that produce these and similar compounds. In this work, we employed a TetA-based dual-selection to isolate new TAL-responsive AraC variants showing reduced background expression and improved TAL sensitivity. To improve TAL specificity, OA was included as a ‘decoy’ ligand during negative selection, resulting in the isolation of a TAL biosensor that is inhibited by OA. Finally, to engineer OA-specific AraC variants, the iterative protein redesign and optimization computational framework was employed, followed by 2 rounds of directed evolution, resulting in a biosensor with 24-fold improved OA/TAL specificity, relative to AraC-TAL1.

Key words: synthetic biology, protein engineering, polyketides, high throughput screening, molecular reporters, IPRO

Introduction

Regulatory protein-based biosensors that respond to small molecule ligands have been used to engineer and optimize enzymes and biosynthetic pathways (Rogers *et al.*, 2016; Dietrich *et al.*, 2010; Zhang and Keasling, 2011; Wang and Cirino, 2016). Two major applications for biosensors in metabolic engineering are: (i) to facilitate high-throughput screening/selection of large enzyme/pathway libraries and (ii) dynamic regulation to balance cell growth and biochemical production. In both scenarios, the sensor to be employed must be equipped with the relevant ligand sensitivity and specificity. However, engineered regulatory proteins often display lowered sensitivity toward the non-native ligands and relaxed specificity as compared with their wild-type counterparts (Collins *et al.*, 2006; Reed *et al.*,

2012; Tang *et al.*, 2013; Chen *et al.*, 2015; Lönneborg *et al.*, 2012; Taylor *et al.*, 2016). Furthermore, many engineered regulatory proteins display elevated background (leaky) expression levels (Lönneborg *et al.*, 2012; Chen *et al.*, 2015; de los Santos *et al.*, 2016); for some applications, high background expression can be detrimental (Loew *et al.*, 2010; Lachmann *et al.*, 2015).

Regulatory protein AraC was previously engineered to respond to non-native ligands, and those variants facilitated high-throughput screening of enzymes and pathways (Tang *et al.*, 2013; Tang and Cirino, 2011; Qian *et al.*, 2019). One AraC variant, AraC-TAL1, responds to triacetic acid lactone (TAL). AraC-TAL1 was used in blue/white screening of libraries of the type III polyketide synthase (PKS), 2-pyrone synthase (Tang *et al.*, 2013), as well as screening of

host genome libraries (gene deletions and overexpression) (Li et al., 2018), for enhanced TAL production by engineered *Escherichia coli*. This TAL sensor displays elevated background expression as compared with wild-type AraC, as well as relaxed substrate specificity and relatively low sensitivity to TAL (half-maximum dose response occurs at 4 mM TAL concentration). More interesting applications of a TAL biosensor, or a biosensor that responds to similar, minimal polyketides such as orsellinic acid (OA), lie in the identification of novel PKS variants whose poor functional expression and/or low catalytic activity demand significantly greater sensitivity and/or specificity than AraC-TAL1 (Yeom et al., 2018). A similar constraint was recently described by Thompson et al. (2020), for the case of engineering biosynthesis of the nylon precursor caprolactam. A highly sensitive caprolactam biosensor identified from *Pseudomonas putida* ($K_d \approx 5 \mu\text{M}$) proved useful for detecting minute caprolactam levels, enabling rapid and accurate screening of novel production pathways (Thompson et al., 2020).

In the present study, we first sought variants of AraC-TAL1 showing reduced background expression and higher TAL sensitivity, using a TetA-based dual selection to rapidly enrich a library of ligand binding domain (LBD) variants. Seven variants of AraC-TAL1 having improved sensitivity and/or lowered background expression (comparable to that of wild-type AraC) were isolated using this selection platform. All seven AraC-TAL1 variants show various degrees of relaxed inducer specificity, including response to OA. A subsequent round of selection in which OA was included as a decoy ligand resulted in isolation of variant AraC-TAL⁺/OA[−] (carrying nine total amino acid substitutions relative to wild-type AraC), in which OA competitively inhibits the TAL-induced response.

OA is a common tetraketide product of type I PKS systems (Sanchez et al., 2010), and an OA-producing type III PKS from *Rhododendron dauricum* has been described (Taura et al., 2016). Importantly, TAL is a common ‘derailment’ triketide byproduct during OA biosynthesis (Taura et al., 2016). An OA sensor useful for engineering OA biosynthesis, or for studying/engineering chain elongation vs termination during tetraketide biosynthesis, must therefore show specificity toward OA and not TAL. We accordingly sought variants with higher specificity toward OA, using computational protein design followed by directed evolution. Variant AraC-OA8, with 12 total amino acid substitutions relative to wild-type AraC, shows a 24-fold increase in the OA specificity as compared with AraC-TAL1 and retains low background expression, comparable to that of wild-type AraC. Collectively, the new AraC variants described have potential utility as biosensors for engineering PKS specificity and improving OA biosynthesis, as well as transcriptional regulators in related engineered biosynthesis pathways.

Materials and Methods

Plasmid construction

Plasmid pPCC1322 (map and sequence in Supplementary Data) was constructed from pFG29-TAL (Frei et al., 2016) as follows: the f1 origin, together with ~1 kb non-coding sequence flanking the f1 origin, was removed from pFG29-TAL, and the ribosome binding site (RBS) region of AraC-TAL1 was modified to achieve a lower translation initiation rate for AraC-TAL1 or its variants. Library enrichment with *tetA*-dual selection was carried out with plasmids pPCC1340 and pPCC1342. Plasmid pPCC1340 was constructed by swapping *gfp* of pPCC1322 with *tetA*. *TetA* was amplified from PC05 chromosome (Khankal et al., 2009). Plasmid pPCC1342 was constructed by replacing the RBS of *tetA* with a stronger RBS

(designed with RBS Calculator). For comparison, the TetA translation initiation rates, calculated using the RBS Calculator (Espah Borujeni et al., 2014; Espah Borujeni and Salis, 2016, 2017; Salis et al., 2009) (<https://salislab.net/software/>), are 20607 au with plasmid pPCC1340 and 571456 au with pPCC1342.

T4 ligase and all restriction enzymes were purchased from New England Biolabs. T4 ligase was used for ligation reactions. NEBuilder® HiFi DNA Assembly Master Mix was used for Gibson Assembly (Gibson et al., 2009). High-fidelity PCR in this work was performed using Phusion® High-Fidelity DNA Polymerase or Q5® High-Fidelity DNA Polymerase. PCR conditions followed NEB Tm Calculator (<https://tmcaltcalculator.neb.com/>) and vendor's instructions for the polymerases. Zymoclean™ Gel DNA Recovery Kit was used for gel purification of DNA fragments.

Library construction

Random mutagenesis libraries were generated by amplifying the LBD of the respective parent AraC variant using the Agilent GeneMorph® II Random Mutagenesis Kit. For TetA selections, the library insert was ligated into pPCC1342, in place of the gene encoding AraC-TAL1. For libraries with AraC-OA6 and AraC-OA7 as parents (green fluorescent protein (GFP) fluorescence-based screening in microtiter plates), the library inserts were ligated into pPCC1322 vector. Library error rates were determined by sequencing 10 random library clones: error rates were ~4.5 mutations/kb for TetA selection libraries and ~4.2 mutations/kb for green fluorescent protein (GFP) screening libraries.

Combinatorial assembly of mutations identified in variants AraC-TAL12 to AraC-TAL17 was performed by assembly of purified PCR fragments amplified from each mutant (including parent AraC-TAL1). Primers were designed such that each fragment covers 1 mutation. For 2 mutations that are in proximity, fragments were amplified by PCR such that the fragments include all 4 possible combinations of the 2 mutations. The fragments were then assembled with Gibson Assembly into pPCC1322 vector for fluorescence-based microtiter plate screening. Sequencing of randomly picked clones confirmed correct library assembly. A nearly identical assembly approach was used to construct the library representing all 32 combinations of the 5 single- or double-amino acid substitutions returned by iterative protein redesign and optimization (IPRO) (those in AraC-OA1 through AraC-OA5).

IPRO to isolate OA-specific AraC-based biosensors

Using the modeled structure of AraC-TAL14 as the starting point for computational design, OA-specific AraC variants were obtained using an IPRO approach (Pantazes et al., 2015; Chowdhury et al., 2020), with unrestricted choice of substituted amino acid type for positions Pro8, Pro11, Thr24, Pro25, Gly30, Leu72, His80, Tyr82, Arg89, His93, Gly135 and Asn139. These residue positions include those that have previously been targeted for saturation mutagenesis (Tang et al., 2008), in addition to residue positions that appeared in new TAL-specific AraC variants described in this work. The protein redesigns were driven by the primary objective of enhancing OA interaction (sequence redesign step) and secondary objective of eliminating TAL interaction (only side chain repacking without introducing further amino acid changes). Note that ligand binding in AraC does not necessarily induce the requisite conformational changes to activate gene expression. Hence, ligand-interacting AraC variants with no induced GFP expression are expected to be encountered during experimental characterization of the IPRO-returned

designs. After sampling several design trajectories that progressively incorporate amino acid substitutions that improve binding to OA without destabilizing the variant structure (by >25% of the AraC-TAL14), the top five designs with the highest OA/TAL interaction energy ratios and with single or double amino acid substitutions were chosen for experimental characterization.

Fluorescence assays for measuring GFP expression

GFP expression under the control of AraC-based biosensors was measured using a microtiter plate-based fluorescence assay. For characterizing individual variants and constructing dose response curves, single colonies of HF19 (Tang *et al.*, 2008) transformants was inoculated into 500 μ l LB supplemented with 50 μ g/ml apramycin, in 96-deep-well-plates. After 6–12 hours of growth at 37°C 900 RPM on a shaking platform, the cultures were then diluted into fresh LB containing 50 μ g/ml apramycin and 100 μ M IPTG (isopropyl β -D-1-thiogalactopyranoside), with or without inducer ligand(s) of interest at final concentrations as indicated in the presented results. The subcultures were grown for 4–6 hours at 37°C 900 RPM in 96-well deep well plates. The cells were next pelleted and washed with PBS buffer before measurements. Fluorescence was measured on SpectraMax[®] Gemini[™] EM Microplate Spectrofluorometer from Molecular Devices[®]. Optical Density at 595 nm (OD₅₉₅) was measured on BMG Labtech NOVOstar Microplate reader. Fluorescence intensity was normalized by OD₅₉₅. Fold-induced GFP expression was calculated by dividing the normalized fluorescence intensity in presence of the inducer by the normalized fluorescence intensity in absence of the inducer. Note that cell toxicity, as noted by significantly impaired growth and reduced GFP expression, occurs at TAL concentrations above $\sim$ 4 mM and OA concentrations above $\sim$ 2 mM. Induced GFP expression was not studied at such toxic inducer concentrations.

For microtiter plate-based fluorescence screening of AraC libraries, single colonies of fresh HF19 transformants (450–500 colonies from each library) were inoculated into 500 μ l LB supplemented with 50 μ g/ml apramycin, in 96-deep-well-plates. HF19 cells transformed with pPCC1322 containing wild-type AraC or AraC-TAL14 in place of AraC-TAL1 were used as controls on each screening plate. The library cultures were then handled as described above for the case of individual clones. For the small combinatorial library comprising IPRO-predicted substitutions, 3 sets of subcultures were prepared for screening: with no inducer, 0.5 mM OA, or 3 mM TAL. For the random mutagenesis library, subcultures were prepared with no inducer, 1 mM OA or 3 mM TAL. Top clones from the combinatorial library were defined as those which had the highest OA specificity ratio (ratio of fold-induced GFP with OA to the fold-induced GFP with TAL), whereas the best clones from the random mutagenesis library were those with the highest fold-induced GFP with OA. All data points reported represent the average values of at least two independent replicates.

TetA dual-selection conditions

Enriching AraC-TAL1 variants with reduced background expression and higher sensitivity

HF19 competent cells were transformed with the library plasmid pool and the transformants were inoculated into 500 ml LB supplemented with 50 μ g/ml apramycin, 100 μ M IPTG and 1 mM NiCl₂. The negative selection was performed with vector pPCC1342 (strong *tetA* RBS), starting with $\sim$ 10⁸ transformants and harvested when OD₅₉₅ was above 1 (total $\sim$ 2.5 $\times$ 10¹¹ CFU). After each round of

negative selection, the library was harvested and the plasmids were extracted. The library insert was then re-cloned into the pPCC1340 backbone (weak *tetA* RBS) for positive selection on LB-Agar plates in the presence of 50 μ g/ml apramycin, 100 μ M IPTG, 1% glycerol, 50 mM TES (pH 7), 8 μ g/mg Tc and 0.5 mM TAL. Library transformants were plated directly on positive selection plates and cells were harvested by scraping after $\sim$ 13 hours of incubation at 37°C for plasmid extraction. The harvested library then was re-cloned back to pPCC1342 vector for the next round of negative selection. After 3 rounds of dual-selection, the enriched library was cloned to pPCC1322 backbone for an end-point screening with fluorescence assay in 96-well deep-well plates. A total of 94 colonies of HF19 transformants with the enriched library were assayed in the fluorescence assay for their leakiness and responses to TAL.

Enriching AraC-TAL1 variants that do not respond to OA

Here, both negative and positive selections were carried out using pPCC1340 (weak *tetA* RBS). Negative selection was carried out in LB supplemented with 50 μ g/ml apramycin, 100 μ M IPTG, 0.5 mM NiCl₂ and 2 mM OA, starting with $\sim$ 10⁸ transformants and harvested when OD₅₉₅ was above 1 (total $\sim$ 2.5 $\times$ 10¹¹ CFU). The plasmids were then extracted for re-transformation for positive selection (re-transformation was performed to eliminate any potential mutations appearing in the host genome). Positive selection was carried out on LB-Agar plates as described above. After 3 rounds of dual-selection, the enriched library was cloned to pPCC1322 backbone for an end-point screening with fluorescence assay in 96-well deep-well plates. A total of 94 colonies of HF19 transformants with the enriched library were assayed in the fluorescence assay for their responses to 1 mM TAL and 1 mM OA.

Results

TAL sensors with improved sensitivity and reduced background expression

The *tetA*-encoded class C tetracycline resistance protein (TetA) is a tetracycline/H⁺ antiporter that confers simultaneous tetracycline (Tc) resistance and nickel (Ni²⁺) sensitivity to *E. coli* (Stavropoulos and Strathdee, 2000). These combined features have enabled the development of dual selection systems in *E. coli* based only on *tetA* expression, for engineering riboswitches (Nomura and Yokobayashi, 2007; Muranaka *et al.*, 2009) and genome editing (Ryu *et al.*, 2017). Attracted to its simplicity for library screening, we optimized the *tetA* dual selection system to isolate AraC-based biosensors showing improved TAL sensitivity and reduced background expression as compared with AraC-TAL1 (Fig. 1A). The gene encoding the AraC-TAL1 LBD was amplified using error-prone PCR, and the random mutation library was expressed in *E. coli* strain HF19. The expression of *tetA* was placed under the control of promoter P_{BAD}, regulated by an AraC-TAL1 variant expressed from the same plasmid. AraC-TAL1 variants allowing leaky expression of *tetA* in absence of TAL (and in presence of Ni²⁺) are eliminated by virtue of their Ni²⁺ sensitivity during negative selection. Meanwhile inclusion of Tc (8 μ g/mg) allows for positive selection of AraC-TAL1 variants induced by a relatively low concentration of TAL (0.5 mM). Iterative rounds of negative selection enriched AraC-TAL1 variants having low background in absence of TAL, with positive rounds added after each negative selection step, to enrich only variants responsive to TAL.

Following three rounds of dual selection, we identified six AraC-TAL1 variants (named AraC-TAL12 to AraC-TAL17) having higher

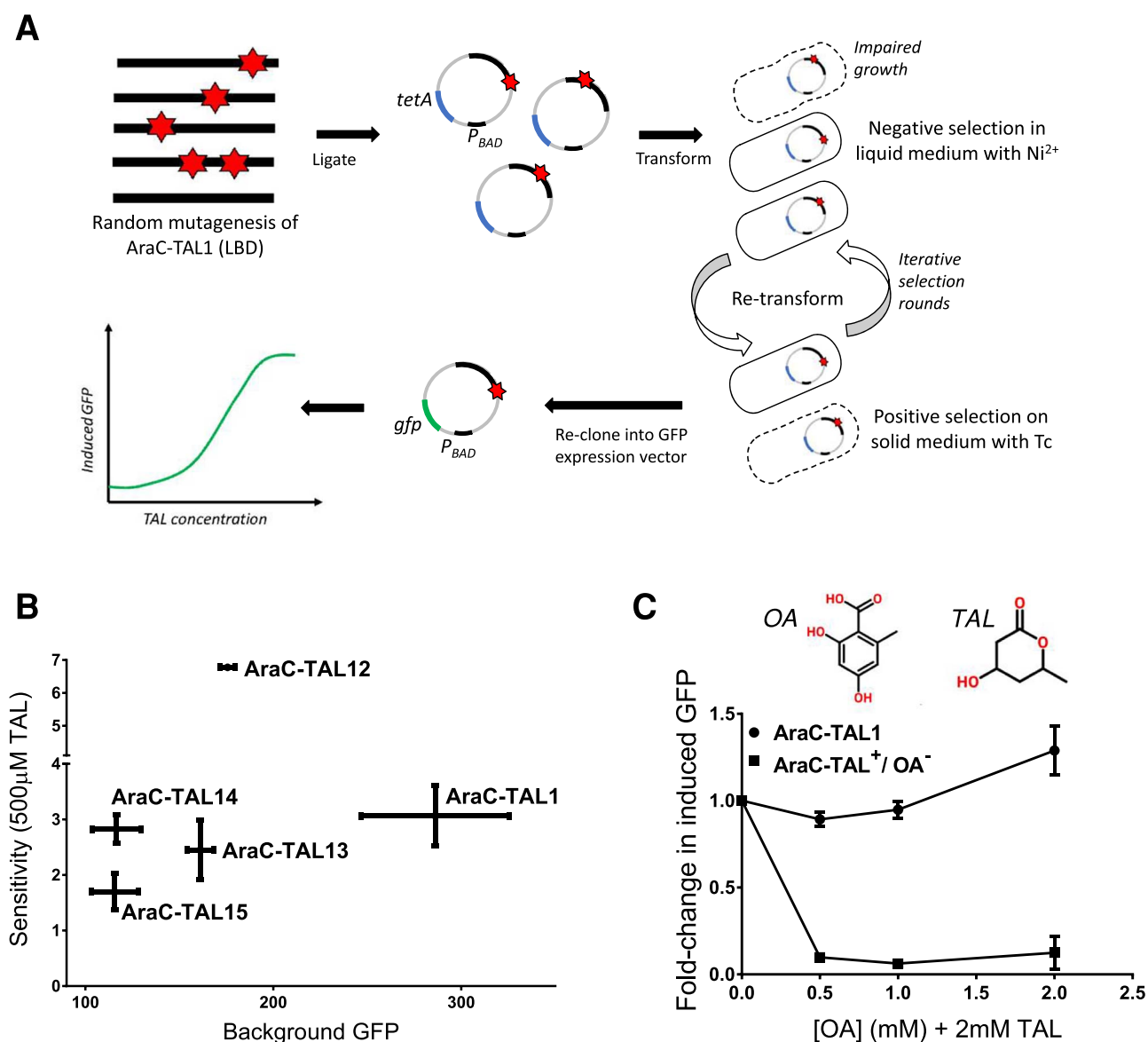


Fig. 1 AraC-based TAL sensors with improved sensitivity and reduced background expression. **A**) Iterative rounds of negative selection (in presence of Ni^{2+} and absence of TAL) in liquid medium and positive selection (in presence of Tc and TAL) on solid medium to isolate AraC-TAL1 variants with higher sensitivity and lower background expression as compared with AraC-TAL1. The selected mutants were cloned into a P_{BAD} -*gfp* expression vector for end-point screening and fluorescence-based dose response characterizations. **B**) TAL sensors with improved sensitivity and/or reduced background expression, relative to AraC-TAL1. Sensitivity is defined as the induced GFP expression response (as measured by fluorescence intensity) when cells were grown in media supplemented with 500 μ M TAL. Background expression for wild-type AraC is 89 ± 3 . **C**) OA inhibits TAL-induced GFP expression in variant AraC-TAL⁺/OA⁻. All fluorescence measurements were normalized by the measured cell density (RFU/OD595). Background GFP represents the fluorescence measurements (RFU/OD595) in the absence of inducer. Induced GFP expression indicates the normalized fluorescence value in presence of the inducer divided by the normalized fluorescence value in absence of the inducer. Data points are the average of 2 values, and error bars represent the range.

sensitivity and lower background expression than the AraC-TAL1 parent. Sequencing of the selected mutants identified single amino acid substitutions in the AraC-TAL1 LBD (Table I). The six AraC-TAL1 variants were further characterized using a P_{BAD} -*gfp* expression vector, for TAL-dependent fluorescence measurements (results summarized in Fig. 1B; dose response curves shown in Fig. S1). AraC-TAL12 shows > 2-fold greater induced GFP expression with 0.5 mM TAL as compared with the AraC-TAL1 parent (6.8-fold vs 3.1-fold). Meanwhile AraC-TAL14 and AraC-TAL15 show the tightest repression in the absence of TAL, with background GFP expression

levels comparable to that of wild-type AraC (Fig. 1B). Although sensitivity as it relates to a dose response is most commonly defined as the concentration at half of the saturation signal, the toxicity of these inducers prevents the identification of a true maximum response. Instead, we define TAL sensitivity as the ratio of the GFP expression level at 0.5 mM TAL to the background GFP expression level without inducer.

A small library was next constructed to test whether any combinations of the amino acid substitutions in variants AraC-TAL12 to AraC-TAL17 would further improve sensitivity or reduce

Table I. Amino acid substitutions in AraC-based biosensors described in this work

AraC variant	Amino acid substitutions
AraC-TAL1	(wild-type AraC), P8V, T24I, H80G, Y82L, H93R
AraC-TAL12	(AraC-TAL1), P11L
AraC-TAL13	(AraC-TAL1), E165G
AraC-TAL14	(AraC-TAL1), L72V
AraC-TAL15	(AraC-TAL1), G135W
AraC-TAL16	(AraC-TAL1), L160F
AraC-TAL17	(AraC-TAL1), Q54R
AraC-TAL ⁺ /OA ⁻	(AraC-TAL1), P25Q, G30S, R89C, N139S
AraC-OA1	(AraC-TAL14), L82M
AraC-OA2	(AraC-TAL14), V72D
AraC-OA3	(AraC-TAL14), I24D, P25G
AraC-OA4	(All TAL14), G135L, N139G
AraC-OA5	(AraC-TAL14), V8D
AraC-OA6	(AraC-TAL14), P25G
AraC-OA7	(AraC-TAL14), P25G, I36F, P128S, A140V
AraC-OA8	(AraC-OA7), H81Q, Q142L

background expression. Details of library construction and screening are provided in Materials and Methods. With only 64 possible combinations, the *P_{BAD}-gfp* expression vector was used for library cloning, and 280 total transformants were screened by microtiter-plate-based fluorescence measurement in the presence vs absence of TAL (this time at 1 mM TAL, to reduce screening stringency). The individual amino acid substitutions appear to be largely non-additive since all clones showed TAL response and background GFP expression levels similar to or worse than those of the parent clones. Results from screening and characterization of 1 variant (named AraC-TAL18) carrying substitutions from AraC-TAL12 (P11L) and AraC-TAL17 (Q54R) are presented in the Supplementary Data (Table S1).

A TAL sensor with improved specificity

Our previous studies show that our AraC-based TAL sensors also respond to one or more compounds that are structurally similar to TAL, such as 4-hydroxy-3,6-dimethyl TAL (4H36M TAL) and 2-hydroxybenzoic acid (salicylic acid) (Frei *et al.*, 2016; Frei *et al.*, 2018). We find similar ligand promiscuity for these new TAL sensors, with results summarized in Table II (response curves are given in Figs S2 and S3). Variants AraC-TAL12 through AraC-TAL17 all respond to 2-hydroxybenzoic acid (2OHBA or salicylic acid), 4H36M TAL and 2,4-dihydroxy-6-methylbenzoic acid (OA), to various extents, as does AraC-TAL1. Meanwhile, none show response to 4-hydroxybenzoic acid (4OHBA), L-arabinose or phloroglucinol. There is no apparent correlation between response to TAL and response to other inducers tested. As an example, AraC-TAL16 shows a response to 4 mM TAL that is ~1.7-fold higher than that of AraC-TAL14 (Fig. S1), but its response to 2 mM OA is ~2-fold lower than that of AraC-TAL14 (Fig. S2B). Similarly, AraC-TAL12 responds to 2 mM TAL 3.3-fold stronger than AraC-TAL1 (Fig. S1), but response to 2 mM OA is similar to that of AraC-TAL1 (Fig. S2B). These observations demonstrate how single amino acid substitutions can alter not only sensitivity but also specificity of the ligand response.

Although a sensor that responds to OA and not TAL is useful for engineering OA biosynthesis, the opposite specificity is also of interest

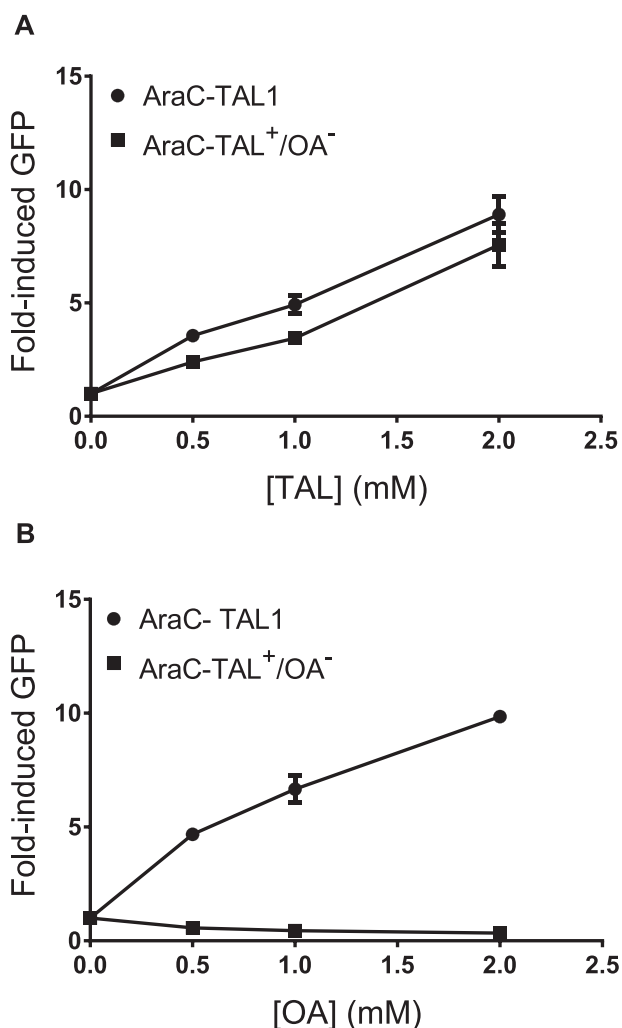
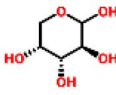
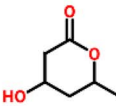
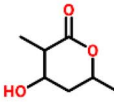
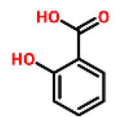
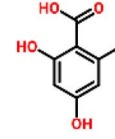
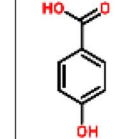
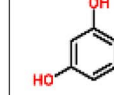
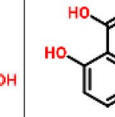


Fig. 2 Induced GFP expression of AraC-TAL1 and AraC-TAL⁺/OA⁻ induced with A) triacetic acid lactone (TAL), B) orsellinic acid (OA). Data points are the average of 2 values, and error bars represent the range.

for discriminating between the two compounds (e.g. to probe chain elongation in an OA-producing PKS). As a proof of principle, the same random mutagenesis library as above (AraC-TAL1 as parent) was again enriched using the *tetA* dual selection platform, but this time with the inclusion of 2 mM OA as a ‘decoy’ compound during negative selection steps. Following 3 rounds of selection/counterselection, one interesting variant named AraC-TAL⁺/OA⁻ was isolated and characterized. AraC-TAL⁺/OA⁻ shows slightly lowered background expression compared with AraC-TAL1, a slightly reduced response to TAL and no response to OA (Fig. 2). Further investigation revealed that OA in fact inhibits the ability of TAL to induce AraC-TAL⁺/OA⁻ (addition of 0.5 mM OA to the culture prevents GFP expression in response to TAL), whereas no such inhibition is seen with AraC-TAL1 (Fig. 1C). The addition of OA also further reduces background expression caused by AraC-TAL⁺/OA⁻, to a level near that of wild-type AraC. Interestingly, AraC-TAL⁺/OA⁻ carries four amino acid substitutions compared with parent AraC-TAL1 (Table I). The analysis of these substitutions, including the characterization of AraC-TAL1 variants carrying each single substitution, is provided in the supporting information (Fig. S4).

Table II. Analyzing ligand promiscuity of AraC-TAL1 variants

AraC-based sensors	Compounds used to test ligand promiscuity of AraC-based sensors							
	L-arabinose 	4-Hydroxy-6-methyl-2-pyrone (TAL) 	4-Hydroxy-3,6-dimethyl-2-pyrone (4H36M) 	2-Hydroxybenzoic acid (SA) 	2,4-Dihydroxy-6-methylbenzoic acid (OA) 	4-Hydroxybenzoic acid 	Phloroglucinol 	6-Methyl-2-hydroxybenzoic acid 
AraC-TAL1	×	✓	✓	✓	✓	×	×	✓
AraC-TAL12	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL13	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL14	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL15	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL16	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL17	×	✓	✓	✓	✓	×	×	N/A
AraC-TAL ⁺ /OA ⁻	N/A	✓	N/A	✓	×	N/A	N/A	✓

✓: fold-induced GFP expression > 2-fold ×: fold-induced GFP expression < 2-fold N/A: not tested

Whereas OA inhibits gene activation by AraC-TAL⁺/OA⁻, this same variant still shows a strong induced response to 6-methyl-2-hydroxybenzoic acid (differs from OA by one hydroxyl group) and 4-hydroxybenzoic acid (differs from OA by one hydroxyl and one methyl group) (Table II). The inhibition of AraC-TAL⁺/OA⁻ by OA is the reflective of inhibition of wild-type AraC by D-fucose (Doyle et al., 1972), where D-fucose differs from the native inducer L-arabinose by one methyl group. Structural analysis comparing the AraC LBD in complex with L-arabinose vs D-fucose highlights how relatively subtle differences in ligand binding can result in the inhibition of AraC rather than gene activation, which requires major conformational changes (Doyle et al., 1972). Modeling studies to investigate those molecular determinants of ligand binding which correlate with induced gene expression vs inhibition of gene expression are beyond the scope of the present study, but are anticipated to provide new insights into rational design of AraC-based sensors.

Design of a sensor with enhanced specificity toward OA

Among other PKS systems, orcinol synthase from *R. dauricum* is a type III PKS that produces OA, using acetyl-CoA and malonyl-CoA starter molecules (Taura et al., 2016). OA biosynthesis proceeds through the formation of a tetraketide intermediate, with TAL appearing as a byproduct due to lactonization of the triketide precursor (Taura et al., 2016). The type III PKS 2-pyrone synthase (2-PS) from *Gerbera hybrida* similarly uses acetyl-CoA and malonyl-CoA starter molecules to proceed through a triketide

intermediate in the biosynthesis of TAL, with no tetraketide product observed (Austin and Noel, 2003). Such control over the length of the final polyketide product is often attributed to the shape and volume of the active site cavity of the type III PKS (Austin and Noel, 2003). A biosensor that responds to OA and not TAL could therefore be a useful tool, both for engineering enhanced OA biosynthesis in a recombinant host, and for probing/engineering chain length control and substrate/product specificity in these and similar type III PKS enzymes. We therefore sought to alter the ligand specificity of an OA-responsive AraC-TAL variant.

Although using the *tetA* dual selection system with addition of OA as decoy ligand proved effective for isolating the AraC-TAL⁺/OA⁻ variant, this was deemed quite fortuitous: during the end-point screening with the fluorescence assay, only 2 out of 94 clones selected from the enriched library showed response to TAL, whereas the rest of the population was minimally responsive to TAL and OA. Both TAL-responsive clones were then identified to be identical. In retrospect, considering OA and TAL induce the parent (AraC-TAL1) to similar extents, a single library of random substitution variants was more likely to yield an OA-inhibited variant using this selection method as compared with a variant that shows little response to OA while retaining strong response to TAL. Before further directed evolution, we therefore turned to computational modeling and binding calculations for insights into improving specificity toward OA.

Computational models of AraC-TAL variants (1, 12, 14, 15 and 19) were generated using the Mutator plugin of the IPRO program (Pantazes et al., 2015; Chowdhury et al., 2020) by imposing the appropriate amino acid substitutions onto the structure of wild-type

AraC (PDB accession: 2ARC (Soisson *et al.*, 1997)), followed by a CHARMM—energy minimization step for clash-free re-packing of the amino acid side chains that lie within 10 Å of the altered residue. Initial docked conformation of OA in AraC and the list of neighboring (within 10 Å) pocket residues were obtained by using the *find_contacts* module of OptMAVEN-2.0 program (Chowdhury *et al.*, 2018). Interaction energy scores between the AraC variants and ligands TAL and OA were computed using the CHARMM energy function (Brooks *et al.*, 2009) and used as a proxy for binding affinity. Table S2 lists the CHARMM interaction energy scores (sum of van der Waals forces, electrostatics and implicit solvation effects modeled as a dielectric continuum), along with the corresponding experimentally measured fold-increases in GFP expression in response to 1 mM inducer (TAL or OA), for these AraC-TAL variants as well as wild-type AraC. The relative ratios of CHARMM interaction energy scores (OA/TAL, also presented in Table S2) correlate well with the ratios of fold-induced GFP expression (OA/TAL) ($R^2 = 0.75$), indicating that relative interaction energy scores reasonably capture ligand specificity. Of the variants, AraC-TAL14 demonstrated the strongest response to OA, whereas still showing low background GFP expression in the absence of TAL (Figs 1B, S1 and S2B). AraC-TAL14 was therefore chosen as the starting point for computational design.

OA-specific AraC variants were predicted using the IPRO suite of programs—IPRO (Pantazes *et al.*, 2015; Chowdhury *et al.*, 2020); refer to Methods section). IPRO has been previously used to successfully tune substrate and cofactor specificities of several enzymes including bacterial thioesterases (Hernández Lozada *et al.*, 2018; Grisewood *et al.*, 2017) and non-ribosomal peptide synthases (Throckmorton *et al.*, 2019). The top five designs (OA/TAL interaction energy ratios) as identified by IPRO, named AraC-OA1 through AraC-OA5, are listed in Table S3. Note that AraC-OA1, 3 and 4 each has a single amino acid substitution relative to AraC-TAL14, whereas AraC-OA2 and AraC-OA5 each have two substitutions relative to AraC-TAL14. These five variants were constructed and subsequently characterized experimentally. The experimentally determined specificity for each IPRO-designed variant, here defined as GFP expression response to 1.5 mM OA relative to that in the presence of 3 mM TAL, is given in Table S3 (fold-induced GFP expression for all variants is provided in Table S3 as well). These concentrations were chosen such that compound toxicity is not a factor, signal strengths are well above background, and the ratio indicates preference toward OA. Only AraC-OA1 and AraC-OA4 showed significant response to OA (>2-fold change in fluorescence when induced), and none of these variants showed higher specificity toward OA than AraC-TAL14.

Although none of the IPRO-predicted variants tested showed higher specificity than AraC-TAL14, we felt it prudent to screen a small library comprising all combinations of the 5 single or double amino acid substitutions returned by IPRO (those in AraC-OA1 through AraC-OA5), with AraC-TAL14 as parent (32 combinations). Using our $P_{BAD-gfp}$ reporter construct, 450 clones were screened by microtiter plate-based fluorescence assay, in the presence vs absence of 3 mM TAL or 0.5 mM OA. Screening was carried out with the intent to isolate variants with low TAL response, high OA response and low background GFP expression. Most variants screened showed either no response to OA or TAL, or showed high background GFP expression. However, four somewhat promising clones were selected for further characterization. Sequencing revealed that these four clones were identical, carrying substitutions P25G and N139G. As shown in Table III, this variant shows low background expression

and low GFP expression response to 3 mM TAL, but also a reduced response to 1.5 mM OA as compared with AraC-TAL14. Substitution of P25G or N139G (individually, in AraC-TAL14) revealed that P25G alone increases OA specificity to a level similar to that of the AraC-TAL14 parent and also reduces response to TAL (Table III). AraC-TAL14/P25G was renamed AraC-OA6 and this variant was selected as the parent for engineering further improvements in OA response and specificity via directed evolution.

Directed evolution of an OA-specific AraC variant

A library of AraC-OA6 variants carrying random substitution in the LBD was generated and screened with a microtiter plate-based fluorescence assay, using our $P_{BAD-gfp}$ reporter system. The library was first screened in the presence vs absence of 1.5 mM OA (instead of 0.5 mM OA as before, simply to reduce screening stringency), followed by the analysis of response to TAL by selected clones. From this first round of screening (513 transformants were screened and four unique clones with high response to OA were further characterized), variant AraC-OA7 showed the highest response to OA (Fig. S5A), along with a low response to TAL (6-fold increase in GFP fluorescence in the presence of 3 mM TAL, as compared with 14-fold increase in GFP fluorescence with AraC-TAL14) (Fig. S5B) and showed a background expression comparable to wild-type AraC. This resulted in a 10-fold increase in OA specificity as compared with AraC-TAL14 and a 15-fold increase compared with AraC-TAL1 (Fig. 3A). AraC-OA7 carries three amino acid substitutions relative to its parent (Table I). The analysis of the individual contributions of each of these three substitutions is presented in Table S4. We next subjected AraC-OA7 to another round of random mutagenesis and screening (513 transformants were screened and six unique clones with high response to OA were further characterized), resulting in AraC-OA8, the variant with the highest specificity toward OA (Figs 3A, S6). AraC-OA8 shows a 1.5-fold increase in OA specificity, as compared with AraC-OA7 ($P = .09$, $n = 3$), a 15-fold increase in OA specificity as compared with AraC-TAL14 ($P < .005$, $n = 3$) and a 24-fold increase in OA specificity as compared with AraC-TAL1 ($P < .005$, $n = 3$) and retains low background expression in absence of any inducer (Fig. 3A). AraC-OA8 carries two amino acid substitutions relative to its parent (Table I), and 12 total amino acid changes as compared with wild-type AraC. Dose responses of AraC-OA8 to OA are provided in Fig. 3B. We expect that variants with further improvements in OA specificity and/or sensitivity would be readily discovered through continued rounds of directed evolution.

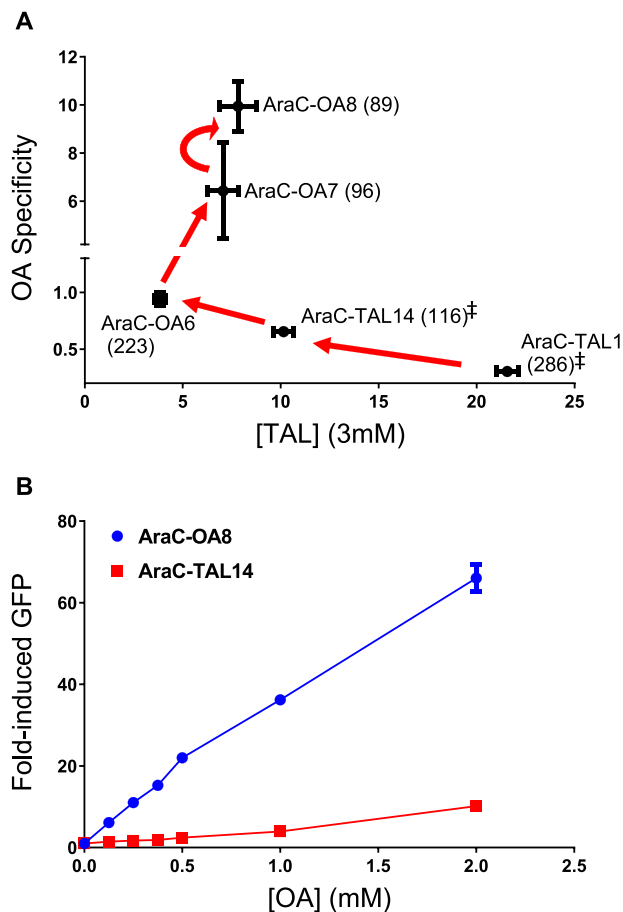
Discussion

Although engineering allosteric regulatory proteins to respond to non-native ligands for biosensor applications is relatively straightforward, attaining significant (or desired) sensitivity and specificity is less clear-cut, requiring more challenging optimization. In this work, we improved upon features of AraC-TAL1, an AraC variant that responds to both TAL and OA, resulting in new variants with greater sensitivity and specificity toward TAL or OA. Table I provides the amino acid substitutions found in all AraC variants described in this work. Four images in Fig. S7A–D show the locations of these residues in the AraC crystal structure, in relation to the ligand binding pocket. Understanding how amino acid substitutions affect gene expression by AraC is complicated by the fact that ligand binding must also induce significant conformational changes, including

Table III. OA specificities of variants identified from the combinatorial library comprising all combinations of amino acid substitutions returned by IPRO

AraC-TAL14 variants	Background GFP expression	Fold-induced GFP expression with 1.5 mM OA (N = 3, $\pm$ SD) (A)	Fold-induced GFP expression with 3.0 mM TAL (N = 3, $\pm$ SD) (B)	OA specificity (A/B)
AraC-TAL14	116 $\pm$ 13	6.6 $\pm$ 0.3	10 $\pm$ 0.5	0.7
AraC-TAL14/P25G/N139G	118 $\pm$ 14	3.8 $\pm$ 0.4	4.6 $\pm$ 0.3	0.8
AraC-TAL14/P25G (AraC-OA6)	223 $\pm$ 12	3.6 $\pm$ 0.5	3.8 $\pm$ 0.3	0.9
AraC-TAL14/N139G	84 $\pm$ 11	3.1 $\pm$ 0.2	6.9 $\pm$ 0.3	0.5

SD: standard deviation.

**Fig. 3** (A) Pathway of evolving OA specificity in AraC-based biosensors. OA specificity is defined as the ratio of fold-induced GFP by 1.5 mM OA vs fold-induced GFP by 3 mM TAL. Values in parentheses represent background GFP expression normalized by cell density (RFU/OD595) in the absence of inducer. (B) Fold-induced GFP expression of AraC-OA8, in comparison with AraC-TAL14, when induced with OA. Background GFP expression values marked with ‡ are average of 2 values. All other data points are the average of 3 values and error bars represent standard deviation.

changes in interactions between the ligand binding and DNA binding domains, before gene expression is activated (Soisson *et al.*, 1997). Below we merely highlight salient features of some substitutions that were identified.

We first isolated six AraC-TAL1 variants with reduced background expression levels and improved sensitivity to TAL from a random substitution library of the AraC-TAL1 LBD. All variants

isolated in this round bear single amino acid substitutions and display different levels of background expression, sensitivity to TAL (Fig. S1) and substrate specificities (Fig. S2). AraC-TAL12 shows the highest response to TAL amongst all existing TAL sensors (Fig. S1) and varies from AraC-TAL1 only by substitution P11L, located in the AraC N-terminal arm (comprised of residues 8–20). This same substitution reduces background expression compared with AraC-TAL1. The effect of P11L is intriguing in that most substitutions in the N-terminal arm lead to high levels of leaky expression (Berrondo *et al.*, 2010; Dirla *et al.*, 2009). Perhaps in the context of AraC-TAL1, P11L helps to stabilize the repressing conformation in the absence of TAL (strengthens interaction with the DNA binding domain), while also enhancing TAL binding (Saviola *et al.*, 1998; Soisson *et al.*, 1997). AraC-TAL14 (L72V) and AraC-TAL15 (G135W) show the lowest background expression levels, comparable with that of wild-type AraC (Fig. S1). Position 72 is in proximity to AraC ligand binding pocket (Seedorff and Schleif, 2011) (Fig. S7B); this conservative substitution may help stabilize the AraC repressing conformation, without significantly affecting inducibility by TAL. Meanwhile substitution G135W lies in the AraC dimerization interface (Fig. S7D) (Seedorff and Schleif, 2011). Substitutions in the dimerization interface have been noted to indirectly enhance AraC repression (Tang and Cirino, 2010).

As our TAL-responsive AraC variants also respond to OA, we reasoned that one of these would make a suitable starting point for engineering an OA-specific biosensor (e.g. as opposed to starting with wild-type AraC, which shows no response to OA). To this end, we started with AraC-TAL14 as the parent and identified variants with improved specificity toward OA. Variant AraC-OA6, with substitution P25G, shows reduced gene expression response to both TAL and OA, but also higher specificity toward OA (Fig. 3A). Residue 25 resides in the first β -strand of the ligand binding pocket (Fig. S7B). Substitution P25S in wild-type AraC was previously reported to reduce response to arabinose by 4-fold, possibly due to weakened interactions between the N-terminal arm and the DNA binding domain in the activating conformation (Reed and Schleif, 1999). Variant AraC-OA7 bears 3 additional amino acid substitutions (I36F, P128S and A140V). Results from characterizing each single back-substitution within AraC-OA7 are provided in Table S4. Of most significance, although I36F alone enhances response toward OA, all three new substitutions are necessary to achieve the highest specificity and lowest background expression. A final round of evolution yielded variant AraC-OA8, with two additional substitutions, H81Q and Q142L. H81Q resides in proximity to the binding pocket and introduces an extra charge (Fig. S7B). Q142 lies in the dimerization interface (Fig. S7D), which might have contributed to the further reduced background expression level of AraC-OA8.

In total, AraC-OA8 contains 12 amino acid substitutions compared with wild-type AraC and shows 24-fold improved specificity toward OA over TAL relative to AraC-TAL1. This variant retains repression in the absence of inducer to the same extent as wild-type AraC and shows a 6-fold induced GFP expression response to 125 μ M OA. This new biosensor/reporter should prove useful for engineering OA biosynthesis in *E. coli*.

Supplementary Data

Supplementary data are available at PEDS online.

Funding

We acknowledge US National Science Foundation (NSF) grants CBET-1511425 (to PCC) and CBET-1703274 (to CDM) for financial support. CDM and RC also thank the Department of Energy (DOE) - Center for Bioenergy Innovation grant DE-AC05-000R22725 for funding our computational effort.

References

- Austin, M.B. and Noel, J.P. (2003) *Nat Prod Rep*, **20**, 79–110.
- Berrondo, M., Gray, J.J. and Schleif, R. (2010) *J Mol Biol*, **398**, 462–470.
- Brooks, B.R., Brooks, C.L., Mackerell, A.D. *et al.* (2009) *J Comput Chem*, **30**, 1545–1614.
- Chen, W., Zhang, S., Jiang, P., Yao, J., He, Y., Chen, L., Gui, X., Dong, Z. and Tang, S.Y. (2015) *Metab Eng*, **30**, 149–155.
- Chowdhury, R., Allan, M.F. and Maranas, C.D. (2018) *Antibodies (Basel)*, **7**.
- Chowdhury, R., Grisewood, M.J., Boorla, V.S., Yan, Q., Pflieger, B.F. and Maranas, C.D. (2020) *Structure*.
- Collins, C.H., Leadbetter, J.R. and Arnold, F.H. (2006) *Nat Biotechnol*, **24**, 708–712.
- de los Santos, E.L., Meyerowitz, J.T., Mayo, S.L. and Murray, R.M. (2016) *ACS Synth Biol*, **5**, 287–295.
- Dietrich, J.A., McKee, A.E. and Keasling, J.D. (2010) *Annu Rev Biochem*, **79**, 563–590.
- Dirla, S., Chien, J.Y. and Schleif, R. (2009) *J Bacteriol*, **191**, 2668–2674.
- Doyle, M.E., Brown, C., Hogg, R.W. and Helling, R.B. (1972) *J Bacteriol*, **110**, 56–65.
- Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N. and Salis, H.M. (2017) *Nucleic Acids Res*, **45**, 5437–5448.
- Espah Borujeni, A., Channarasappa, A.S. and Salis, H.M. (2014) *Nucleic Acids Res*, **42**, 2646–2659.
- Espah Borujeni, A. and Salis, H.M. (2016) *J Am Chem Soc*, **138**, 7016–7023.
- Frei, C.S., Qian, S. and Cirino, P.C. (2018) *Protein Eng Des Sel*, **31**, 213–220.
- Frei, C.S., Wang, Z., Qian, S., Deutsch, S., Sutter, M. and Cirino, P.C. (2016) *Protein Sci*, **25**, 804–814.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) *Nat Methods*, **6**, 343–345.
- Grisewood, M.J., Hernandez Lozada, N.J., Thoden, J.B. *et al.* (2017) *ACS Catal*, **7**, 3837–3849.
- Hernández Lozada, N.J., Lai, R.Y., Simmons, T.R., Thomas, K.A., Chowdhury, R., Maranas, C.D. and Pflieger, B.F. (2018) *ACS Synth Biol*, **7**, 2205–2215.
- Khankal, R., Chin, J.W., Ghosh, D. and Cirino, P.C. (2009) *J Biol Eng*, **3**, 13.
- Lachmann, N., Brenig, S., Hillje, R. *et al.* (2015) *Gene Ther*, **22**, 883–892.
- Li, Y., Qian, S., Dunn, R. and Cirino, P.C. (2018) *J Ind Microbiol Biotechnol*, **45**, 789–793.
- Loew, R., Heinz, N., Hampf, M., Bujard, H. and Gossen, M. (2010) *BMC Biotechnol*, **10**, 81.
- Lönneborg, R., Varga, E. and Brzezinski, P. (2012) *PLoS One*, **7**, e29994.
- Muranaka, N., Sharma, V., Nomura, Y. and Yokobayashi, Y. (2009) *Nucleic Acids Res*, **37**, e39.
- Nomura, Y. and Yokobayashi, Y. (2007) *Biosystems*, **90**, 115–120.
- Pantazes, R.J., Grisewood, M.J., Li, T., Gifford, N.P. and Maranas, C.D. (2015) *J Comput Chem*, **36**, 251–263.
- Qian, S., Li, Y. and Cirino, P.C. (2019) *Microb Cell Fact*, **18**, 18.
- Reed, B., Blazeck, J. and Alper, H. (2012) *J Biotechnol*, **158**, 75–79.
- Reed, W.L. and Schleif, R.F. (1999) *J Mol Biol*, **294**, 417–425.
- Rogers, J.K., Taylor, N.D. and Church, G.M. (2016) *Curr Opin Biotechnol*, **42**, 84–91.
- Ryu, Y.S., Chandran, S.P., Kim, K. and Lee, S.K. (2017) *PLoS One*, **12**, e0181501.
- Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) *Nat Biotechnol*, **27**, 946–950.
- Sanchez, J.F., Chiang, Y.M., Szcwzyk, E., Davidson, A.D., Ahuja, M., Elizabeth Oakley, C., Woo Bok, J., Keller, N., Oakley, B.R. and Wang, C.C. (2010) *Mol Biosyst*, **6**, 587–593.
- Saviola, B., Seabold, R. and Schleif, R.F. (1998) *J Mol Biol*, **278**, 539–548.
- Seedorff, J. and Schleif, R. (2011) *J Bacteriol*, **193**, 5737–5746.
- Soisson, S.M., MacDougall-Shackleton, B., Schleif, R. and Wolberger, C. (1997) *Science*, **276**, 421–425.
- Stavropoulos, T.A. and Strathdee, C.A. (2000) *FEMS Microbiol Lett*, **190**, 147–150.
- Tang, S.Y. and Cirino, P.C. (2010) *Protein Sci*, **19**, 291–298.
- Tang, S.Y. and Cirino, P.C. (2011) *Angew Chem Int Ed Engl*, **50**, 1084–1086.
- Tang, S.Y., Fazelinia, H. and Cirino, P.C. (2008) *J Am Chem Soc*, **130**, 5267–5271.
- Tang, S.Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A. and Cirino, P.C. (2013) *J Am Chem Soc*, **135**, 10099–10103.
- Taura, F., Iijima, M., Yamanaka, E., Takahashi, H., Kenmoku, H., Saeki, H., Morimoto, S., Asakawa, Y., Kurosaki, F. and Morita, H. (2016) *Front Plant Sci*, **7**, 1452.
- Taylor, N.D., Garruss, A.S., Moretti, R. *et al.* (2016) *Nat Methods*, **13**, 177–183.
- Thompson, M.G., Pearson, A.N., Barajas, J.F. *et al.* (2020) *ACS Synth Biol*, **9**, 53–62.
- Throckmorton, K., Vinnik, V., Chowdhury, R., Cook, T., Chevrete, M.G., Maranas, C., Pflieger, B. and Thomas, M.G. (2019) *ACS Chem Biol*, **14**, 2044–2054.
- Wang, Z. and Cirino, P.C. (2016) *Curr Opin Biotechnol*, **42**, 159–168.
- Yeom, S.J., Kim, M., Kwon, K.K. *et al.* (2018) *Nat Commun*, **9**, 5053.
- Zhang, F. and Keasling, J. (2011) *Trends Microbiol*, **19**, 323–329.