Design and Characterization of *E. coli* Transcriptional Factors for Use as Biosensors in Synthetic Biology

by
Aarti Doshi

A dissertation submitted to the Department of Biology and Biochemistry
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biology

Chair of Committee: Patrick C. Cirino

Committee Member: Navin Varadarajan

Committee Member: Brigitte Dauwalder

Committee Member: Shaun Xiaoliu Zhang

University of Houston
August 2020

ACKNOWLEDGMENTS

Every challenging work requires self-effort along with guidance and support from others close to our hearts. I am very thankful to all the wonderful people who have helped me throughout my journey and helped me accomplish a successful thesis.

First, I thank my parents and siblings for their love and support which has always enabled me to dream and achieve big things in life. Their blessings and prayers have helped me pass through difficult times.

My special thanks to my advisor, Dr. Patrick Cirino, my academic parent, who gave me immense freedom to pursue my interests within and beyond my research. I thank him for being extremely patient with me when I did not understand certain engineering principles and appreciate the effort, he took to make sure I did not proceed work without clearly understanding the concepts. I am extremely pleased to grow as an independent thinker and a scientist under his guidance.

My special thanks to Dr Navin Varadarajan for guiding me on our collaboration project. Without his invaluable support and suggestions, my project would never have reached fruition.

I wish to thank Dr Brigitte Dauwalder for being in my committee and giving me time, despite being an expert from outside my field. Her inputs gave me a different perspective about the applications of my research. I always looked forward to her Thanksgiving dinner invite. It made me feel belonged despite being far from home.

I also wish to thank Dr Shaun Zhang for being on my committee and giving me the opportunity to rotate in his lab. It helped me build lasting collaborations with the members in the Zhang group.

My deepest gratitude goes to Dr Jeffrey Spencer and Dr Zhiqing Wang, my mentors who taught me all the lab skills I needed to know, when I joined UH and who have continued to guide me even after they have graduated. My friends and lab members, Dr Shaza Abnouf, Yixi Wang, and Nam Nguyen have always been helpful, and it is fun to discuss ideas and random thoughts with them. I thank them for the late-night media inoculations they did for me so that I could continue to be an early riser every day. My special thanks to all the members of the Varadarajan group, for always opening their lab doors for me. Through their help I could manage to carry out bacterial and mammalian culture work in parallel.

I am indebted to Bhaskar Sarkar, my friend and motivator, who pushed me to work hard and realize my abilities in lab and on the running track. His passion and love for his own research, inspired me to continue to put more effort in my work and love my research even when every project seemed to be failing. My deepest thanks and best wishes to him.

I thank my collaborators for sharing their expertise with me to help me progress in research. I thank NSF for helping me by funding my research and making PhD a less arduous task.

ABSTRACT

Biosensors are detection tools which use a biological recognition element to provide selective and quantitative information about changes in the availability of a specific input signal. Metabolite-sensing bacterial transcriptional factors (TFs) are a class of biosensors which interact with small molecule ligands (input signal) and lead to subsequent changes in a target gene expression. These bacterial TF-based biosensors are widely used in high-throughput screening (HTS) or selection, and in dynamic gene regulation for bacterial and mammalian host systems. Repurposing of the naturally existing bacterial transcription factor proteins as biosensors typically requires protein engineering to capture/retain the desirable biosensing properties.

In this thesis, I first describe the combined use of computational and evolutionary protein design to isolate AraC-based biosensors with enhanced specificity towards orsellinic acid (OA), as compared to their triacetic acid lactone (TAL)-specific parent (AraC-TAL14). Variant AraC-OA7, isolated after one round of directed evolution, showed 10-fold increase in OA specificity and retained a low background expression level, comparable to that of wild-type AraC. Further directed evolution led to variant AraC-OA8, with 15-fold increased OA specificity compared to AraC-TAL14, still retaining low background. AraC-OA8 has potential utility as biosensor for improving OA biosynthesis through engineering of related biosynthesis pathways.

I next describe the design of a salicylate (SA)-inducible, "SA_{ON}" gene expression system for regulating transgene expression in human Jurkat cell lines. The SA_{ON} system, comprising a modified MarR protein from *E. coli* along with a MarR-regulated promoter,

initially demonstrated a 1.5-fold increase in reporter protein (EGFP) expression in engineered Jurkat cells, when induced with SA. Fusion of the MarR protein with a nuclear-tagged maltose binding protein further led to a total of 2.1-fold increase in EGFP expression in presence of SA. Further optimization of this SA_{ON} system to reduce background (leaky) expression will enable its use for tight and inducible gene regulation in engineered Jurkat cell lines.

Overall, the bacterial TF-based biosensors described in this work will enable researchers to gain desirable control over gene expression for varied synthetic biology applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

BACTERIAL TRANSCRIPTION FACTOR-BASED BIOSENSORS FOR SYNTHETIC BIOLOGY APPLICATIONS

Introduction to transcription factor-based biosensors

Biosensors are detection tools which use a biological recognition element to provide selective and quantitative information about changes in the availability of a specific target signal [1]. Biosensing devices have been widely used in the food and fermentation industries, in medicine, and in metabolic engineering applications, leading to several research efforts being directed towards engineering of biosensing devices with novel recognition properties [2, 3]. Genetically encoded biosensors are *in vivo* analytical tools which propagate the recognition of a target molecule or signal into a desired biochemical response within a cell. These *in vivo* biosensor devices allow intra- and extra-cellular monitoring, of endo- or exogenous small molecules, ions, and changes in physical parameters, through measurable biological responses. Metabolite-sensing proteins (e.g., transcription factors, enzymes, or periplasmic-binding proteins), nucleic acids (DNA and RNA), and engineered whole cells are widely used as biosensors in high-throughput screening (HTS) or selection, and in dynamic gene regulation for bacterial and mammalian host systems [4].

Bacterial metabolite-sensing transcriptional factors (TF) are gene regulatory proteins which interact with small molecule ligands (ions, vitamins, amino acids, sugars, and other organic compounds) and control bacterial gene transcription by influencing the binding of RNA polymerase (RNAP) to the DNA. These TF-based biosensors are composed of two domains: a sensing domain and a regulation domain. The sensing domain or the ligand binding domain (LBD) brings about interaction between the TF and the specific small molecule ligand, while the regulation domain (DNA binding domain, DBD) controls

transcriptional activity from a specific promoter (Figure 1A). Based on their gene regulatory activity, TFs can be transcriptional repressors or transcriptional activators. The repressor proteins bind to an operator sequence within a target promoter and interfere with RNAP binding, thus inhibiting transcription in absence of an input (Figure 1B). Transcriptional activators, on the other hand, bind to enhancer sequences in or around the target promoter and recruit RNAP enzyme to enhance transcription at the target promoter (Figure 1B).

Repurposing of the naturally existing bacterial transcription factor proteins as biosensors often requires the application of protein re-design principles to ensure that the biosensor mediates a predicted outcome in bacterial cells, and to allow direct transfer of the biosensing function from bacterial cells into mammalian cells. In this chapter, I will describe some important principles underlying the design of TF-based biosensors in bacteria including the choice of reporters, design of the target promoter, and general characteristics of TF-based biosensors. I will also summarize additional design principles governing the direct transfer of TF-based biosensors from bacterial cells to mammalian cells. Such knowledge will enable engineering of predictable cell behavior for use in biomanufacturing and in cell-based therapeutics. Further, I will describe a few recent studies that demonstrate the use of TF-based biosensors for high-throughput screening (HTS) in bacteria or repurposing of bacterial TFs for use as biosensors in mammalian hosts.

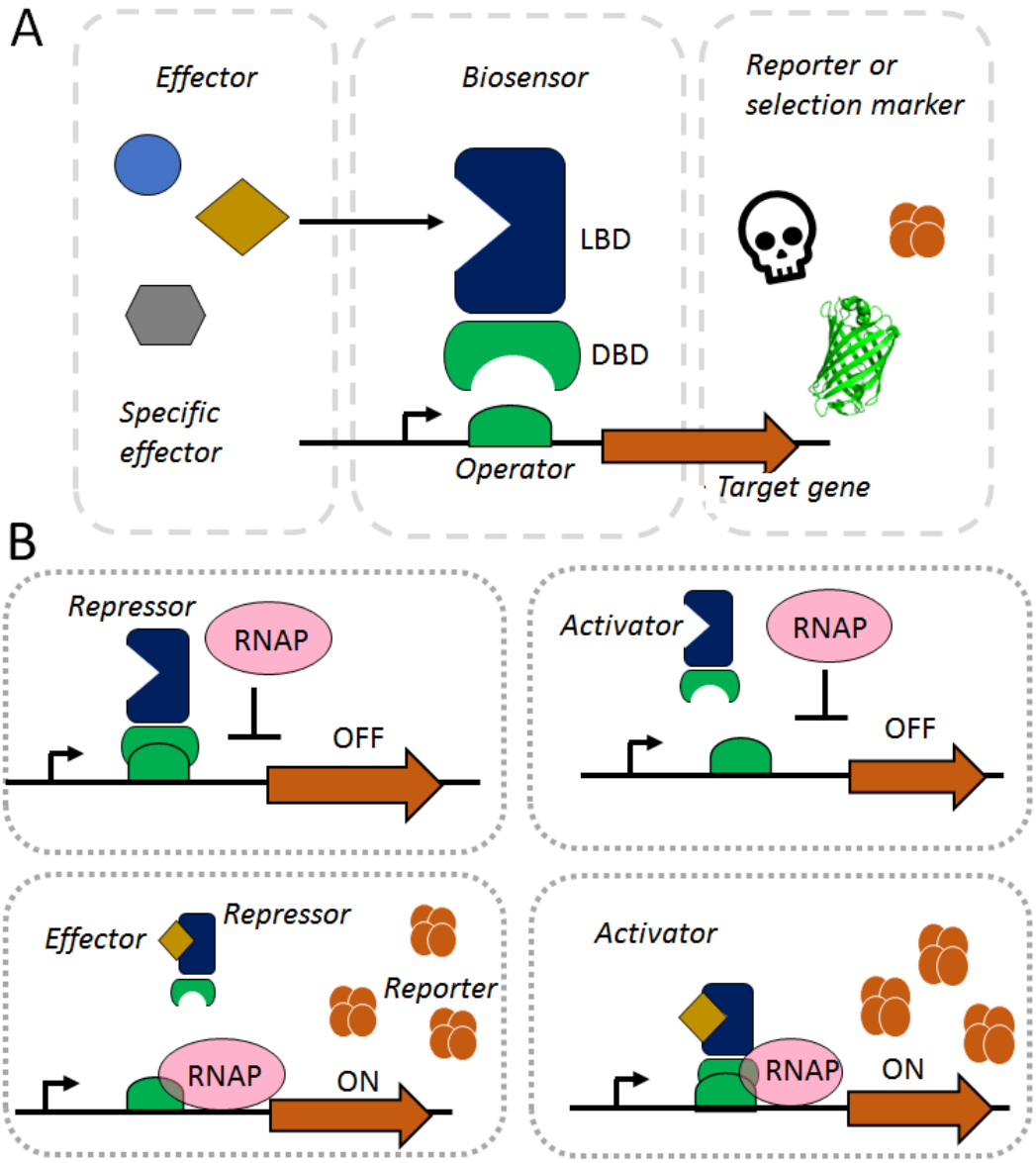

Figure 1. ***Bacterial transcription factor (TF)-based biosensors.*** *A) Bacterial TF-based biosensors consist of a ligand binding domain (LBD), which interacts with a specific ligand or effector molecule, and a DNA binding domain, which interacts with the operator sequence within the associated promoter and regulates the expression of a target gene. The target gene could be a growth selection marker, or an enzymatic or fluorescent reporter gene. B) Transcriptional repressors bind to operator DNA and physically block the RNA polymerase (RNAP) from transcribing target gene when no ligand is present. The repressor releases the blockage in presence of specific small-molecule ligands. Transcriptional activators bind to their operator sequences and enhance RNAP binding, often in presence of specific small-molecule ligands, thereby upregulating target gene expression.*

Designing TF-based biosensors for bacterial hosts

Sustainable bioproduction of chemicals, through microbial fermentations, has seen a rapid development during the past decade and recent efforts are directed towards increasing the productivity of these microbial cell factories [5-7]. Bacterial TF-based biosensors have proven to be important tools that facilitate and accelerate strain improvement by allowing real-time detection of metabolite production [8] or by enabling rapid identification of overproducers through HTS during strain evolution [9]. The development of useful and high-quality TF biosensors requires careful design considerations as discussed below:

Choice of target genes

Metabolite responsive bacterial TF-based biosensors are primarily converting a chemical signal into a gene expression output (activation or repression of transcription) at the target loci. Based on the user's application, the target gene may either be a selection marker gene, whose expression influences the bacterial host survival during a selection process, or the target gene may be reporter gene, whose expression provides a quantitative knowledge about the concentration of the specific small-molecule ligand during a screening process, without influencing the host's survival.

Antibiotic resistance genes typically serve as selection markers, where the production of a metabolite or chemical is coupled to the expression of an antibiotic resistance gene such that only producer strains are enriched in media containing the antibiotic (Figure 2A). However, the use of antibiotic resistance genes as selection markers suffers from the problem of enriching bacterial escapees [10]. Bacterial escapees are generated as a result of spontaneous mutations under the survival stress introduced by antibiotic selections. Bacterial

escapees or false positives recover normal growth phenotype without a corresponding increase in the production of a desired chemical and hence compete with desired mutants having improved chemical production. To avoid escapees, the use of genes encoding for key metabolic enzymes essential for bacterial growth in specialized conditions, is described. Liu *et al.,* (2017) developed a growth-coupled selection strategy to rapidly isolate high L-tryptophan producing strains of *E. coli* from a library of variants of anthranilate synthase, a key enzyme involved in tryptophan biosynthesis [11]. Liu *et al.,* (2017) designed a selection circuit with an L-tryptophan biosensor regulating the expression of the maltose catabolizing enzyme, amylomaltase (*malQ*), as a selection marker. In presence of L-tryptophan, the tryptophan biosensor enabled expression of amylomaltase which in turn allowed rapid growth of the tryptophan producer strains in media supplemented with maltose as the sole carbon source. Through a series of continuous growth selections, Liu *et al.,* (2017) enriched strains with up to 65% increased L-tryptophan production. Moreover, Liu *et al.,* (2017) reported the occurrence of bacterial escapees just after two rounds of selection in an antibiotic resistance-based system, while no escapees emerged even after 30 rounds of selection in the maltose utilization-based selection, demonstrating the advantage of using metabolic enzymes over antibiotic resistance genes as selection markers [11]. The use of selection markers, as target genes, enables direct elimination of unwanted protein variants or non-producer strains yet, the use of selection schemes is often limited due to the difficulty in optimizing growth conditions to avoid escapees and due to lack of obtaining a quantitative measure of the *in vivo* ligand concentration.

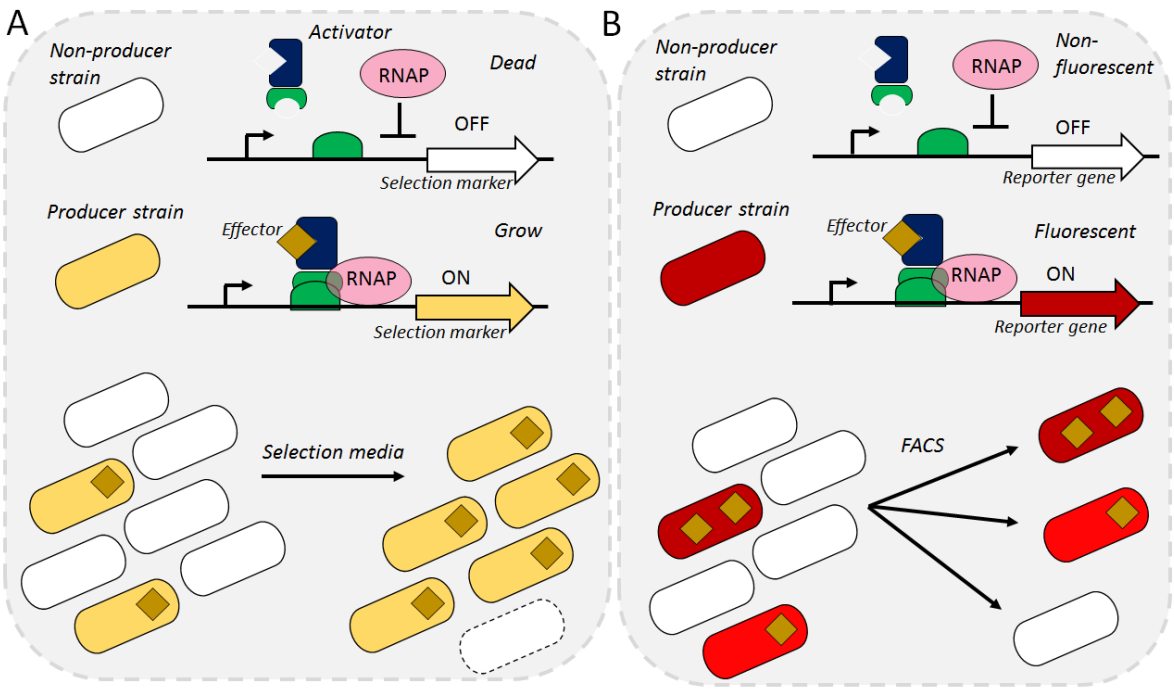

Figure 2. ***TF-based biosensors aid in selection or screening of overproducer strains****. A) Growth-coupled selection of producer strains is achieved by regulating the expression of a selection marker through a TF-based activator. In presence of specific small-molecule ligand, synthesized by a producer cell, the activator enables the RNA polymerase (RNAP) to express a selection marker essential for growth thereby enriching producer strains in a selection media. The non-producer cells are eventually eliminated due to lack of growth in the selection media. B) Overproducer can be separated from moderate-producer and non-producer strains by using fluorescent reporters as target genes. The fluorescence intensity of a cell directly correlates with the concentration of the specific small-molecule ligand. Highly fluorescent cells correspond to overproducer strains and can be separated from other strains using high-throughput screening methods like fluorescence assisted cell sorting (FACS).*

Reporter genes, unlike selection markers, primarily convert a chemical signal detected by the TF biosensor into a measurable colorimetric, fluorescent, or luminescent output and thus provide a quantitative correlation with the *in vivo* small-molecule ligand concentrations (Figure 2B). For instance, Qian *et al.,* (2019) described the use of the β-galactosidase enzyme reporter for screening mutant *E. coli* strains for enhanced salicylate production [12]. The enzyme β-galactosidase cleaves X-gal (5-bromo-4-cholo-3-indolyl β-d-galactopyranoside)

substrate to release a blue colored moiety, thereby imparting a blue color to the bacterial colony. Qian *et al.,* (2019) reported the use of a novel salicylate biosensor to regulate the expression of β-galactosidase and thus isolate high salicylate producers by correlating the intensity of blue colored colonies, developed on agar plates containing X-gal substrate, with the salicylate concentration [12]. Recently, fluorescent proteins (e.g. green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), mCherry protein) have been widely used as reporters due to their strong emissions, easy molecular manipulations, and easy coupling to ultra-throughput method of screening using fluorescence assisted cell sorting (FACS) (Figure 2B). TF sensor-based FACS screening has proven useful in rapid optimization of producer strains, for example, when an L-lysine biosensor was used to isolate L-lysine overproducers from a library of ten million mutant *E. coli* cells screened by FACS in two weeks making the process 10^4-10^5 times faster than shake flask-based selections [13]. Raghavan *et al.,* (2019) also applied a biosensor and FACS-based screening to expedite development of *E. coli* strains producing sustainable and economically viable amounts of acrylic acid after just two rounds of iterative screening [14]. In both the above studies, a direct quantification of the desired chemical (L-lysine or acrylic acid) was obtained by measuring the intensity of fluorescence reporter to identify high-yielding *E. coli* strains. To utilize the high throughput nature of selections and to gain quantitative information from reporter-based screenings, Rossum *et al.,* (2017) describe the simultaneous use of kanamycin resistance selection marker and luciferase luminescent reporter gene for detection and annotation of an L-arabinose isomerase enzyme derived from the thermophilic bacteria *Geobacillus thermodenitrificans* [15]. The kanamycin resistance selection marker enabled rapid enrichment of cells with enzyme activity over cells that do not have any enzyme activity,

while the luciferase reporter enabled separation of true positives from the false positives [15]. Such simultaneous use of selection marker and reporters can overcome the limitations of using only one type of target gene.

Design of the TF-based biosensor regulated promoter

The promoter selected for regulating the target gene expression influences the performance of a biosensor such that strong promoters will often lead to high leaky expression without a large change in target expression upon activation, as strong promoter may facilitate polymerase binding even without a transcription factor. Weak promoters will give low leaky expression but also will have low maximum expression as the polymerase may not tightly associate with a weak promoter even when a transcriptional factor is present. Most studies use naturally existing promoters to control desired target gene expression. Hanko *et al.,* (2018) identified and used the itaconate inducible promoter and its corresponding transcriptional factor from *Yersinia pseudotuberculosis* to isolate high itaconate producers from pathway engineered strains of *E. coli* [16]. Previous works from our group have used the native pBAD promoter, from the *E. coli* araBAD operon, paired with AraC TF-based biosensors, to detect novel small-molecule ligands like mevalonate, D-arabinose, triacetic acid lactone and salicylate [12, 17-19]. When using endogenous promoters care should be taken to avoid interference from the native transcriptional machinery. For instance, Tang *et al.,* (2008) deleted the native *araC* gene from *E. coli* strain BW27786 to construct *E. coli* strain HF19 (with *araC* deleted) which was used in the screening and characterization of AraC mutants without influence from the wildtype AraC protein [18].

Since the endogenous promoters are often prone to influence from the native host transcriptional machinery, or may be incompatible in non-native hosts, some groups have employed chimeric promoters for controlling target gene expression [20]. Machado *et al.,* (2019) developed a biosensor for the detection of protocatechuic acid (PCA) in *E. coli* using a PcaV TF from *Streptomyces coelicolor* and a chimeric promoter which consisted of PcaV operator sites inserted within a T7 phage promoter. The PcaV-chimeric promoter pair controlled the expression of a GFP target gene such that PcaV would bind to the operator sites and repress GFP expression when no PCA was available. Upon PCA binding, PcaV would depress reporter expression and allow generation of a fluorescent signal [20].

Characteristics of a TF-based biosensor

Bacterial TF proteins can be repurposed as biosensors in a different host or re-engineered to respond to new small-molecule ligands in the same bacterial host. For either application, the performance of a biosensor is evaluated on the basis of measurable properties like specificity of a biosensor for a certain small-molecule ligand, response curve of the biosensor relative to the small-molecule ligand concentration, and reproducibility of biosensor function to enable prediction of target gene expression. Each of these measurable properties is discussed below:

Specificity

Specificity of a biosensor is the ability of a biosensor to detect a specific small-molecule ligand in a sample that may contain several other molecules (Figure 3A). Specificity of a biosensor is calculated as the increase in the output signal upon binding a specific ligand divided by the increase in output signal when biosensor is bound by other potential ligands

molecules present in the sample. The degree of specificity desired in a biosensor depends on its application. Biosensors with a broad specificity range are often used as starting points to evolve sensor variants with new specificity. In this regard, Collins *et al.,* (2005) performed three rounds of directed evolution of a second generation LuxR transcriptional activator variant (LuxR-G2E) [21] to isolate a LuxR-G2E-R67M variant having 500-fold change in specificity for decanoyl-homoserine lactone (C10HSL) as compared to the parent G2E variant [22]. The same group demonstrated the use of the broad specificity LuxR-G2E variant for isolating a LuxR-G4E variant that exhibited an additional specificity for butanoyl-homoserine lactone (C4HSL) after two rounds of directed evolution of the LuxR-G2E variant [23]. Both these studies exemplified the use of promiscuous biosensors as intermediates for isolating specialized sensors with new specificity.

Highly specific biosensors that exhibit specificity towards one isomeric form of a molecule offer an ability to differentiate between two structurally similar compounds when present in a mixture. For example, Tang *et al.,* (2008) reported the use of site saturation mutagenesis and FACS-based dual screening to isolate an *E. coli* AraC variant that specifically responds to D-arabinose and cannot bind to the native L-arabinose inducer (small-molecule ligands that increase gene expression) [18]. The authors propose the use of the D-arabinose inducible AraC variant to replace the native AraC protein for use as inducible system in strains that consume L-arabinose [18].

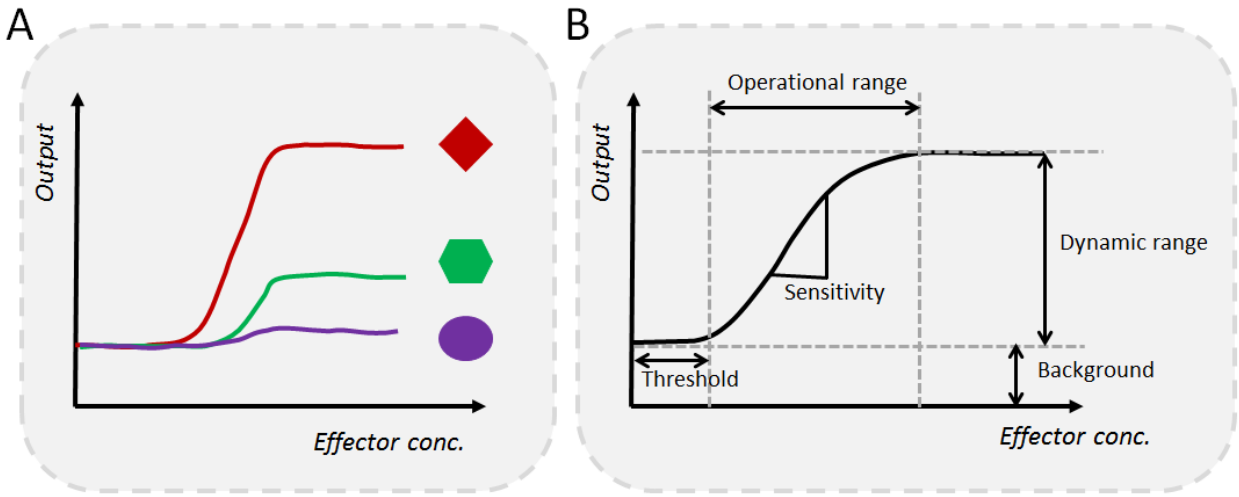

Figure 3. ***Characteristics of TF-based biosensor.*** *A) Specificity of a TF-based biosensor is defined as the ratio of output signal in presence of a specific small-molecule ligand to the output signal in presence of non-specific ligand molecules. B) Biosensor response curve determines the performance of a biosensor at varying ligand concentrations. The response curve provides information about measurable properties like operational range, dynamic range, sensitivity, threshold, and background response of a TF-based biosensor.*

Biosensor response curve

A biosensor response curve describes the changes in the target gene expression relative to varying ligand concentration [24]. The response curve itself is defined by three parameters:

1. Operational range
2. Dynamic range
3. Sensitivity or slope of the response curve

Operational range refers to the range of small-molecule ligand concentrations over which the biosensor shows a significant change in expression of the target gene (Figure 3B). Some biosensors may have a large operational range thus enabling tuning of target gene expression by simply varying the ligand concentration, while a few biosensors may only function over a short operational range essentially having a switch-like control over target

gene expression. The operational range of a biosensor can be changed by modifying the binding kinetics between the TF and its small-molecule ligand or by modifying the affinity of a TF for its promoter by either changing the promoter sequence or by varying the operator copy number and location [25]. Mannan *et al.,* (2017) developed a model based on the *E. coli* lactose inducible repressor system (LacI protein) to identify the effects of change in inducer-TF affinity on the operational range of the lac repressor. Their model predicted that, a decrease in the inducer-TF affinity increases the threshold (minimum concentration of inducer required to produce measurable reporter signal) of the LacI protein, thereby effectively changing the operational range of the LacI protein. To validate their model, Mannan *et al.,* (2017) used isopropyl β-D-1-thiogalactopyranoside (IPTG) and methyl-1-thio-β-D-galactopyranoside (TMG) as the two inducers of the lac system with TMG having ten times lower affinity for LacI than IPTG. *E. coli* cells induced with low-affinity TMG inducer required 120-times higher inducer concentration than cells induced with high-affinity IPTG inducer, to produce measurable RFP reporter fluorescence [25]. The results verified that decreasing inducer-TF affinity indeed increases the threshold for the LacI protein and provided an approach to alter the operational range of the LacI protein [25].

Dynamic range or induction ratio of a biosensor is defined as the ratio of the highest target gene expression obtained in presence of a ligand molecule to the target gene expression obtained when the ligand is not available (Figure 3B). The denominator term in the dynamic range thus demonstrates the leakiness or background expression of a biosensor. When evaluating the performance of various biosensors that may have different background expressions, the induction ratio (or dynamic range) of each biosensor provides a better comparison of the sensor's response than the direct measurement of only the induced

expression. Rogers *et al.,* (2016) developed two novel TF-based biosensors for the real-time monitoring of 3-hydroxypropionate (3-HP) production by coupling the conversion reactions for 3-HP to GFP expression [26]. 3-HP is used as a precursor in acrylate synthesis which is further used as a monomer for several plastics. Rogers *et al.,* (2016) compared the ability of the two bacterial TF-derived biosensors to monitor 3-HP production, by calculating their induction ratio in presence of varying 3-HP concentrations. They demonstrated that the *Rhodobacter sphaeroides* acrylate responsive transcriptional regulator, *acuR,* based biosensor achieved a 90-fold increase in the induction ratio at 10 mM 3-HP concentration while the *E. coli* 2-methylcitrate responsive transcriptional regulator, *prpR*, based biosensor achieved a maximum of 4-fold increase in the induction ratio even at 25 mM 3-HP. The low response of the *prpR*-based biosensor was primarily due to the high background GFP fluorescence, even in absence of the 3-HP inducer [26]. Such background or leaky expression is undesired for biosensors as it does not bring about a tight gene regulation and can accumulate target gene products even in uninduced conditions.

The sensitivity of a biosensor indicates the detection limit of the sensor and is quantified as the slope of the biosensor response curve (Figure 3B) [24, 27]. Highly sensitive biosensors can generate large measurable changes in biosensor response with small changes in the ligand concentrations. Sensitivity of a TF-based biosensor is primarily dictated by the binding affinity between the biosensor and its small-molecule ligand, and by the affinity between the biosensor and its DNA binding site [27]. Thus, the sensitivity of biosensors can be tuned by altering these TF-ligand or TF-DNA binding affinities.

Kasey *et al.,* (2018) performed a site saturation mutagenesis and sequential screening procedure to improve the sensitivity of the promiscuous macrolide sensing MphR transcriptional repressor towards the macrolide antibiotic erythromycin [28]. Five sites on the wild type MphR structure, predicted to drive macrolide recognition, were targeted for site saturation mutagenesis. MphR variants which were two- to five-fold more sensitive to erythromycin than the wild type MphR were isolated from this mutagenesis procedure. The group further used directed evolution to tailor the specificity of the MphR variants and proposed using the improved variants for high throughput screening of macrolide biosynthesis libraries [29].

In an alternate approach, Liu *et al.,* (2017) optimized the TF-promoter binding affinity to improve the sensitivity of the *E. coli* TyrR regulator used to monitor intracellular L-phenylalanine levels [30]. Liu *et al.,* (2017) introduced three mutations in the TyrR regulated promoter region which led to a 3.5-fold increase in the YFP reporter activation by TyrR. Using this improved TF-promoter pair to screen for high L-phenylalanine producers, Liu *et al.,* (2017) were able to increase the chance of a positive YFP signal from 86.2% to 99.5%, indicating the improved sensitivity of the biosensor for L-phenylalanine [31].

Reproducibility

Reproducibility is defined as the ability of identical biosensors to produce the same response to a specific small-molecule ligand. For a TF-based biosensor, reproducibility of function is also desired when combining modular domains from different TF proteins to develop chimeric biosensors. Juárez *et al.,* (2018) generated two benzoate-responsive, chimeric TF biosensors, ChTFBz01 and ChTFBz02, by fusing the benzoate interacting

periplasmic binding proteins (ABE44898 and BzdB1) to two DNA binding domain from bacterial repressor proteins (CbnR and LmrR respectively) [32]. The two chimeric benzoate biosensors repressed their target promoters in the absence of their ligand molecule, benzoate, while this repression activity was attenuated by the presence of benzoate leading to a three-fold increase in the target GFP expression with benzoate [32]. The ChTFBz01 and ChTFBz02 chimeric proteins demonstrate that the domains of TF proteins (CdnR and LmrR) retain their function even after swapping the LBD of the native TF proteins with the benzoate-specific periplasmic binding proteins. Such retained function can be considered as a demonstration of biosensor reproducibility.

Designing bacterial TF-based biosensors for mammalian hosts

Bacterial TF proteins have been successfully repurposed to gain control of mammalian gene expression [33, 34]. The characteristics that make bacterial TF proteins attractive candidates for regulating mammalian genes includes the diversity of TF proteins available that bind to various ligands, the simple TF-promoter-reporter design, and the orthogonal behavior in mammalian cells which prevents any cross-talk with the native mammalian systems [34]. To successfully transfer bacterial proteins into mammalians hosts, design principles in addition to those mentioned for bacterial hosts, need to be considered. These additional design considerations are mentioned below:

Design of mammalian promoter

Bacterial and mammalian promoters are distinct with regards to their size and complexity. To ensure sufficient expression and function of bacterial TF-based biosensors in

mammalian cells, the TF protein and target gene (reporter gene or mammalian gene target) should be expressed from a mammalian promoter (Figure 4A). Strong promoters like the promoter sequence of the human cytomegalovirus (CMV), elongation factor-1 alpha (EF1-α) are often used for expressing the biosensor, while relatively moderate promoter (e.g. promoter for 3-phosphoglycerate kinase (PGK), minimal CMV) may be preferable for expressing the reporter while keeping the background low. For the TF proteins to control the reporter expression, it is essential to insert operator sequences within the target promoter and hence build chimeric promoters similar to those used in bacteria (Figure 4B). Gossen and Bujard (1992) first reported the use of a tetracycline-responsive promoter derived from the minimal CMV promoter element and combined with tet operator sequences, for the tetracycline transactivator (tTA) mediated control of gene expression in mammalian cells [35]. Several studies have followed suite and designed chimeric promoters for regulating mammalian gene expression using bacterial TF proteins, with the target gene and TF protein expression controlled through separate promoters [36-38]. Nevozhay *et al.,* (2013) designed a negative feedback-regulator, called the 'mammalian linearizer', by expressing a tetracycline repressor (TetR) and *gfp* reporter from the same promoter such that the expression of the TetR regulator and the *gfp* target gene depended linearly on the doxycycline ligand concentration [39]. The mammalian linearizer allowed uniform *gfp* gene expression over a wide range of doxycycline concentrations as compared to a feedback devoid system which exhibited a linear response only for a very small range of doxycycline concentrations [40]. Further, Nevozhay *et al.,* (2013) optimized operator sequence positions within the mammalian promoter to increase the performance of the linearizer, indicating that proper positioning of the operator sites must be considered when designing chimeric promoters [40].

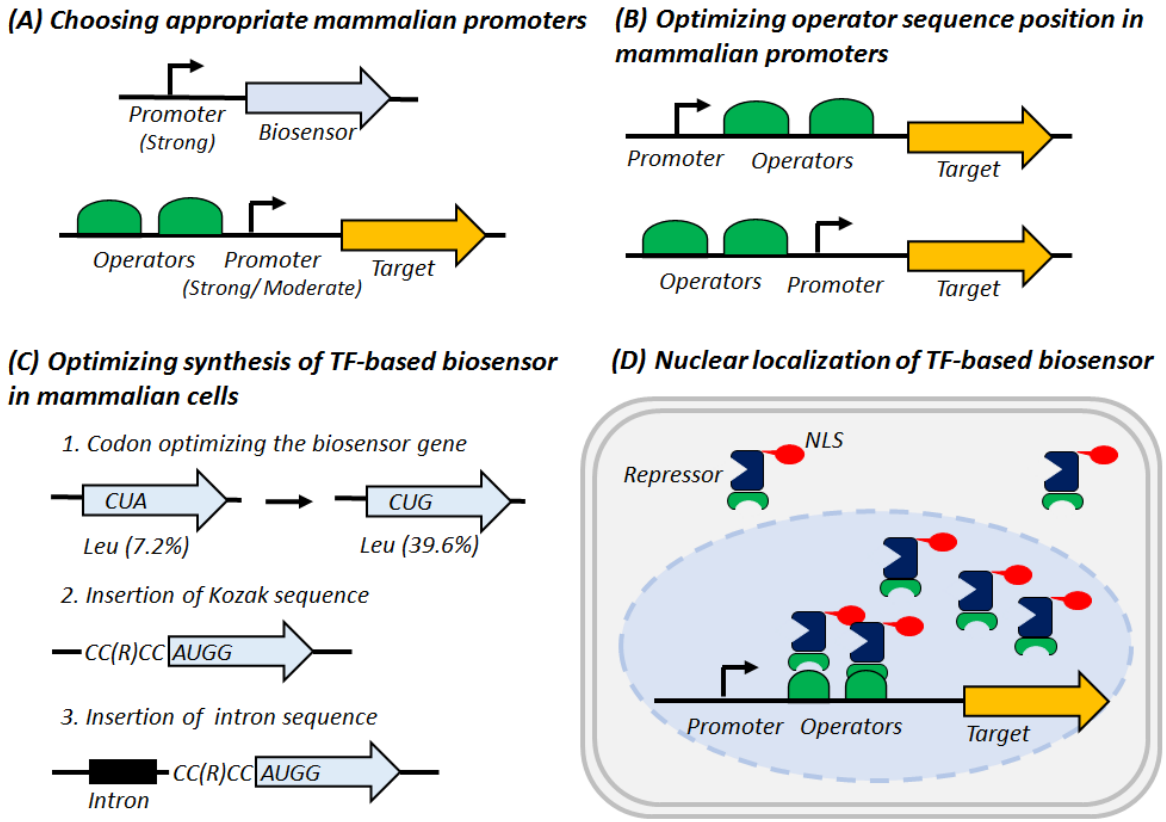

Figure 4. ***Designing bacterial TF-based biosensors for use in mammalian cells.*** *A) Choice of promoters for the biosensor and target genes determines the efficiency of gene regulation by biosensor. Strong promoters are generally used to increase the expression of the biosensor while moderately strong promoters are chosen to express target genes to keep background expression to a minimum. B) optimizing the operator position within the mammalian promoter is essential to ensure efficient regulation of target gene expression. C) Strategies used to increase the synthesis of bacterial TF protein in mammalian cells. D) Addition of a nuclear localization signal (NLS) sequence to the TF protein increases the availability of the TF in the nucleus and promotes binding to the target DNA.*

Optimizing expression of the bacterial TF- based biosensor in mammalian cells

Heterologous protein expression in mammalian cells can be enhanced by increasing the translation efficiency of the non-native genes (Figure 4C). To improve the translations efficiency of the bacterial TF proteins, the following optimization steps are commonly taken:

Codon optimization

Codon usage bias is commonly found across species and largely determines the protein and mRNA levels in an organism [41]. Rare codons in a sequence can reduce the rate of translation for that sequence while preferred codons can often speed up translation elongation [42]. Codon optimization of the bacterial TFs can be done, by either using the most abundant mammalian host codon for every amino acid of the bacterial protein or by randomly assigning codons to the bacterial TF based on the frequency distribution of codons in the host genome. The former technique can rapidly exhaust the host tRNA pool which makes the latter approach more suitable for optimization. Several online tools (GenSmart™ by GenScript, IDT Codon Optimization Tool, Codon Optimization by Genewiz, ExpOptimizer by NovoPro Labs) are available for obtaining optimized bacterial TF gene sequence for expression in a host of choice.

Addition of Kozak sequence

The CC(R)CCAUGG consensus (where R is a purine) which encompasses the AUG start codon in eukaryotes is called the Kozak sequence. This highly conserved sequence plays an important role in helping the scanning ribosome to correctly recognize the translation initiation codon [43, 44]. While a strong promoter may often be enough to increase gene expression, inserting the Kozak sequence in bacterial TF genes may further increase the protein synthesis by enhancing the protein translation efficiency.

Addition of an intron sequence

Most genes in higher eukaryotes (mammalian cells) contain an intron that must be removed to create a translatable mRNA. The intron, though not a part of the mature mRNA,

plays a primary role in increasing the translation efficiency of mRNA. Nott *et al.,* (2003) monitored the translation yields of intron-containing and intronless genes in human HeLa cell line and found that spliced mRNA from intron-containing genes yielded two to four times more protein per mRNA molecule as compared to mRNA from intronless genes [45]. Though the mechanism of intron mediated increased translation is unknown, it is postulated that the splicing machinery may be involved in actively engaging cytoplasmic ribosomal protein with the spliced mRNA. In light of these findings, addition of intron in genes coding for bacterial TF may increase the synthesis of these bacterial biosensors in mammalian cells.

Nuclear localization of TF-based biosensors

A primary difference between bacterial and mammalian cells is the presence of a well-defined nucleus in mammalian cells. The mammalian genome is harbored within the nucleus and any regulatory protein that influences transcription of the mammalian genes must gain access to the nucleus after the protein is synthesized in the cell cytoplasm. Mammalian proteins that are required in the nucleus are tagged for post-synthesis nuclear transport by a nuclear localization signal (NLS) sequence which interacts with the nuclear membrane transporters and promotes active nuclear localization of the tagged proteins [46]. To allow efficient regulation of genes in mammalian hosts, bacterial TF proteins are tagged using known NLS sequences which promote the accumulation of the bacterial TF proteins within the mammalian nucleus following their synthesis in the cytoplasm (Figure 4D). Nevozhay *et al.,* (2013) introduced the simian virus 40 large-T-antigen NLS into the TetR linearizer and found that the NLS containing linearizer construct caused a preferential translocation of the TetR protein into the nucleus of human cell lines thereby leading to a higher fold induction of the target gene (3.9 fold induction in presence of 250ng/ml doxycycline) as compared to a

TetR linearizer construct without the NLS sequence (2.4 fold induction in presence of 250ng/ml doxycycline) [40]. The authors postulated that the increased nuclear concentration of the TetR linearizer led to an increase in the effective binding rate of TetR to its operators and thus enhanced the TetR biosensor performance in the mammalian host cells.

The biosensor design principles described above should be useful for achieving efficient control of target gene expression in bacterial and mammalian hosts. For any application it is essential to characterize the biosensor response curve to identify the working range of the biosensor, and to help optimize sensor properties to expand its use.

Recent applications of TF based biosensors for High Throughput Screening

Several naturally existing bacterial-TFs have been re-engineered to respond to new ligands and repurposed to work as biosensors in native and non-native hosts. Recently, Li *et al.,* (2017) demonstrated the use of a mutated AraC protein responsive to triacetic acid lactone (AraC-TAL) [19] for developing a whole-cell biosensor to facilitate the rapid and high-throughput screening of *E. coli* strains showing hyper-production of malonyl-CoA (Figure 5) [9]. This whole-cell biosensor expressed a 2-pyrone synthase enzyme which catalyzed the synthesis of TAL using malonyl-CoA substrate. The biosensor strain also expressed a lacZ reporter gene controlled by the P_{BAD} promoter in a TAL concentration-dependent manner. Increase in the availability of the malonyl-CoA substrate led to an increase in TAL production which was reported as an increase in LacZ activity. Strains producing 4.2-fold more TAL (and hence more malonyl-CoA) than the wildtype strains were isolated after screening *E. coli* strains from a random transposon insertion library. The group further replaced the 2-PS

enzyme with another type III polyketide synthase to demonstrate screening of high malonyl-CoA producers through phloroglucinol synthesis. Thus, combining the AraC-TAL regulatory system and with a polyketide synthase of choice served to effectively screen a genome-scale library of engineered *E. coli* strains for high intracellular malonyl-CoA which is directly available for use in bioproduction of industrially important chemicals [9].

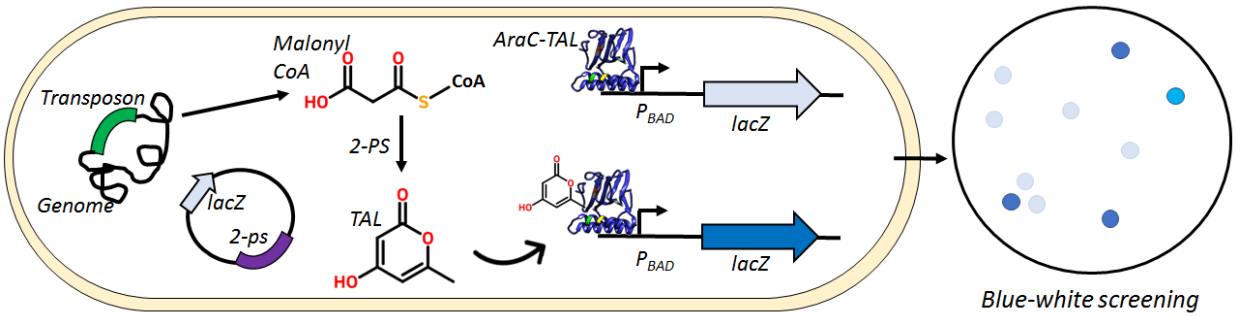

Figure 5. ***Whole-cell biosensor for screening malonyl-CoA overproducing E. coli strains.*** *A genome-scale random transposon insertion library in E. coli was screened to isolate malonyl CoA producer cells. Malonyl CoA produced by the cell was converted into triacetic acid lactone (TAL) in presence of 2 pyrone synthase (2-PS) enzyme. Increasing amounts of TAL production led to the activation of araC-TAL protein. which in turn regulates the expression of β-galactosidase (lacZ gene), such that increased expression of lacZ directly correlates with the amount of malonyl CoA available for bioproduction. The enzyme activity is measured using blue white screening in presence of X-gal substrate in the media.*

As mentioned previously, Hanko *et al.,* (2018) repurposed an itaconate-inducible, LysR type transcriptional regulator, ItcR, from *Yersinia pseudotuberculosis* (Yp) to develop an itaconate biosensor that was used in the optimization of itaconate production in *E. coli* [16]. Itaconic acid is an attractive platform chemical reported by the US DOE as one of the top 12 building block chemicals (http://www.osti.gov/bridge). It can be converted into its polymeric form to produce a superabsorbent material or even used to produce Plexiglas [47]. The *Y. pseudotuberculosis* ItcR (Yp-ItcR) was used to regulate the red fluorescent protein (RFP) reporter expression in *E. coli* in response to itaconate through a naturally existing,

putative itaconate-inducible promoter from Yp. The Yp-ItcR containing *E. coli* cells demonstrated a 39-fold increase in RFP expression after addition of the itaconate inducer. The Yp-ItcR/RFP module was further used to fine-tune the level of cis-aconitate decarboxylase (CadA) enzyme expression thus enabling rapid screening for maximal itaconate production [16].

In a similar study, Liu *et al.,* (2018) demonstrated the use of a shikimic acid inducible transcriptional regulator ShiR, for real-time monitoring of shikimic acid production in three *Corynebacterium glutamicum* strains [8]. The group used the native ShiR from *C. glutamicum* [48] to regulate green fluorescent protein (GFP) expression in response to shikimic acid (SA). The relative fluorescence intensity (RFI) showed a strong linear correlation with the intracellular SA levels (R^2 = 0.92) thus enabling real-time monitoring of SA levels in the tested strains. The ShiR biosensor was also used to screen for high yield *C. glutamicum* strains from a transketolase (tktA) RBS library designed for optimizing gene expression for SA production. The highest SA producer from this library showed a 2.4-fold improvement in SA production as compared to the negative control [8].

Recent applications of TF based biosensors for controlling gene expression in mammalian cells

Metabolite-responsive bacterial-TFs have also been re-purposed to enable inducible control of gene expression in eukaryotic systems. Recently, to minimize the side effects associated with the use of chimeric antigen receptor T-cell therapy (CAR T therapy), Gu *et al.,* (2018) developed inducible CD19CAR (iCAR19) T cells using the Tet-ON system [33]

for doxycycline (dox)-activated expression of the CD19-targeting CAR (Figure 6) [49]. The iCAR19 cells were derived from human peripheral blood mononuclear cells (PBMCs) and engineered using GMP standards for potentially safe and controlled treatment of relapsed or refractory B-cell malignancies [49]. The group reported dox-induced CAR upregulation which was brought down to background levels within 24 hours of removal of the dox inducer, thus producing an effective reversible switch. The iCAR19 cells further exhibited a dox-dependent, high cytotoxic activity (84% lysis) against Raji cells (Human Burkitt′s lymphoma cells expressing CD19) as compared to the cytotoxic activity in absence of dox (34% lysis) or against CD19-negative cells (<20% cell lysis) [49].

To complement the Tet-system, Mullick *et al.,* (2006) developed a cumate-inducible system using the cymene repressor (CymR), derived from *Pseudomonas putida* [50]. In *P. putida*, CymR regulates expression from the p-cym and p-cmt operons, involved in cumate catabolism, by binding to operators (CuO) within the promoters of the p-cym and p-cmt operons. For use in mammalian transcription control, Mullick *et al.,* (2006) placed a single CuO operator downstream of a minimal CMV promoter to generate a hybrid promoter that is regulated by a cumate-inducible CymR. The CymR regulator repressed target gene transcription from the minimal hybrid promoter, and addition of 10 μg/ml cumate was able to fully activate gene [38]. To demonstrate the modularity of the cumate-inducible system, Mullick *et al.,* (2006) also generated a cumate transactivator (cTA), by fusing a VP16 activation domain to CymR, as well as a reverse cumate transactivator (rcTA), by fusing VP16 to a CymR variant that instead binds operator in the presence of cumate. While the use of rcTA is preferable, to enable inducible activation of transcription, its performance was sub-optimal primarily due to high background expression. To improve rcTA activation, the

authors combined CymR and rcTA to tightly regulate target gene expression in the absence of cumate, resulting in up to 700-fold activation in the presence of cumate [38]. The combined CymR and rcTA system was recently used to rapidly produce large amounts of therapeutic recombinant proteins (hCD200Fc and Rituximab) in Chinese Hamster Ovary (CHO) cells [51]. Notably, the hCD200Fc yield was up to 4-fold lower when a constitutive promoter was used instead of the cumate-inducible promoter, demonstrating the advantage of using a tunable and inducible promoter system to obtain high product yields in mammalian cell [51].

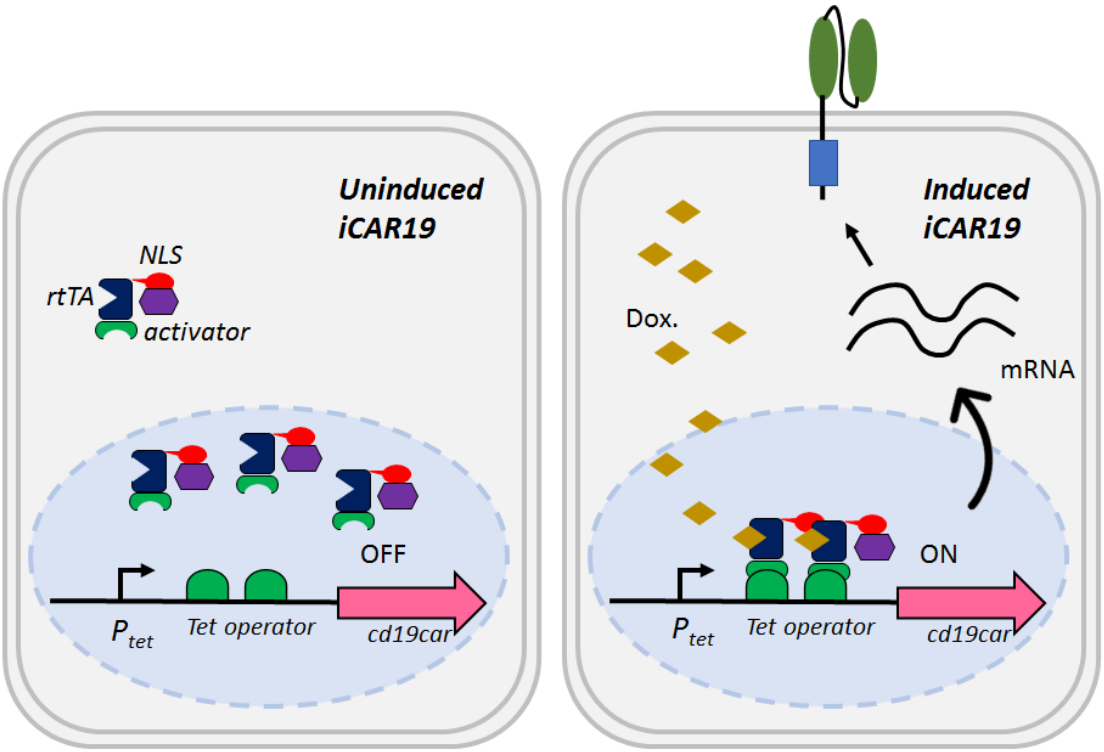

Figure 6. ***Doxycycline inducible system for controlled immunotherapy.*** *Inducible CAR19 (iCAR19) cells constitutively express the reverse tetracycline transactivator protein (rtTA) which localizes in the nucleus of the host cells. In absence of the doxycycline ligand (Dox.), rtTA cannot bind to the target locus and hence expression of the cd19car gene remains off. When dox, is available, rtTA binds to the operator sites within the Ptet promoter and upregulates expression of the CD19 antigen specific chimeric antigen receptors (CD19CAR). The induced cells are later recruited for specific cytotoxicity against cells overexpressing the CD19 antigen.*

Conclusion

Metabolite-responsive bacterial TF-based biosensors have helped to address several challenges in the fields of metabolic engineering [52, 53] and mammalian synthetic biology [54, 55]. By coupling metabolite-binding event to a readable output such as fluorescence or enzyme activity, biosensors can enable the screening of large libraries of engineered organisms for desired properties. A general challenge in the use of biosensors for screening purposes is that the numbers of metabolites that are desirable for screening often exceed the number of specific biosensors available. Bioinformatics approaches and engineering strategies can help identify or evolve novel specificities in biosensors and thus overcome the above challenge.

Bacterial TF-based biosensors have also enabled reprogramming of gene expression in mammalian cells without interfering with the normal host cell function. Successful transfer of bacterial proteins for controlling mammalian genes often requires extensive designing and optimization of the bacterial biosensors and associated regulatory sequences.

In the following chapter, I describe protein engineering efforts used in altering the specificity of an existing AraC based biosensor. A biosensor for orsellinic acid (AraC-OA) was evolved by designing a change in specificity of the existing AraC-TAL sensor. The AraC-OA sensor developed in this research can be used for identifying increased orsellinic acid production in engineered *E. coli* strains or could be used to study the mechanism of enzymes catalyzing OA synthesis.

In Chapter 3, I will describe the use of a bacterial repressor protein to design a novel salicylic acid inducible gene expression system for human cell lines. This salicylic acid

inducible system could provide a safer way to control gene expression in modified human cell lines used in therapeutics and would add a new gene regulatory tool to the existing synthetic biology toolbox to enable multilayered control over eukaryotic gene expression.

CHAPTER 2

ENGINEERING SPECIFICITY OF ARAC-BASED BIOSENSOR TO ORSELLINIC ACID

Introduction

Bacterial TF-based biosensors that respond to small molecule ligands have been used to engineer and optimize enzymes and biosynthetic pathways [56-59]. Two major applications for biosensors in metabolic engineering are: (1) to facilitate high-throughput screening/selection of large enzyme/pathway libraries; and (2) 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 TF proteins often display lowered binding affinities toward the non-native ligands and relaxed specificities, as compared to their wild-type counterparts [19, 22, 60-63]. Furthermore, many engineered TFs display elevated background expression levels [61, 62, 64]; for some applications, high background expression can be detrimental [65, 66].

Bacterial TF protein AraC was previously engineered to respond to non-native ligands, and those variants facilitated high-throughput screening of enzymes and pathways [12, 17, 19]. 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 [19] (Figure 7), as well as screening of host genome libraries (gene deletions and overexpression) [67], for enhanced TAL production by engineered *E. coli*. This TAL sensor displays elevated background (leaky) expression as compared to 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 [68]. A similar constraint was recently described by Thompson et al., (2020) for the case of engineering biosynthesis of the nylon precursor caprolactam [69]. A highly sensitive caprolactam biosensor identified from Pseudomonas putida (K_d≈5 μM) proved useful for detecting minute caprolactam levels, enabling rapid and accurate screening of novel production pathways [69]. In an attempt to obtain AraC-TAL sensors with high sensitivity and low background expression, previous studies in our lab employed a TetA-based dual-selection to isolate seven variants of AraC-TAL1 having improved sensitivity (as much as 25% increased sensitivity relative to TAL1) and/or lowered background expression (as much as 59% reduced background relative to TAL1, and comparable to that of wild-type AraC) (data not published). All seven AraC-TAL1 variants showed various degrees of relaxed inducer specificity, including response to orsellinic acid (OA) (Figure 8 and amino acid substitutions given in Table 1) [70].

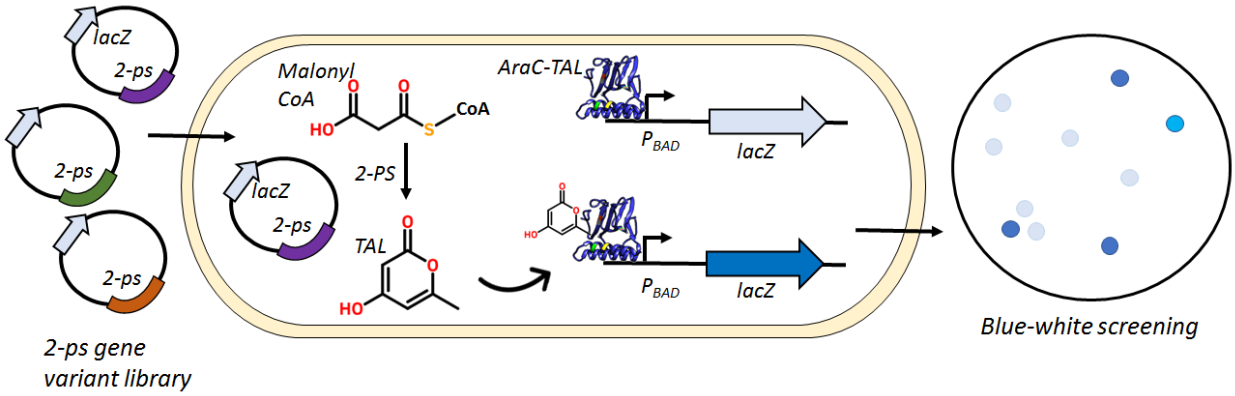

Figure 7. ***AraC-based TAL sensor*** *was used in blue/white screening of libraries of the type III polyketide synthase (PKS), 2-pyrone synthase by Tang et al., (2013) [19]*

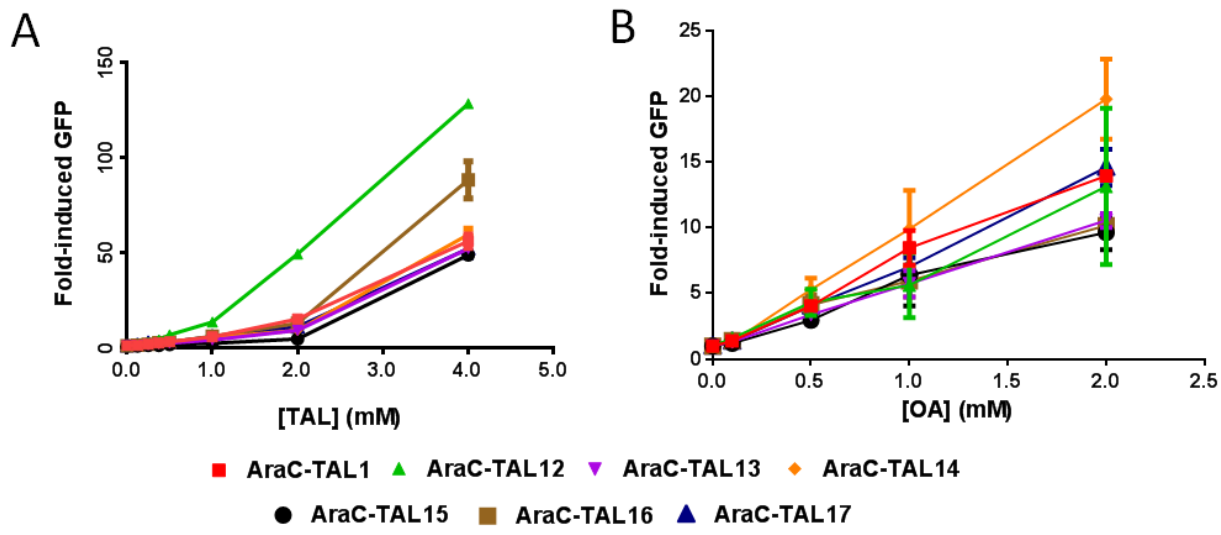

Figure 8. ***Dose response curve of AraC-TAL1 variants with varying inducer concentrations*** *A) 4-Hydroxy-6-methyl-2-pyrone (triacetic acid lactone or TAL) and B) Orsellinic acid (OA). Data points are the average of two values, and error bars represent the range.*

OA is a common tetraketide product of type I PKS systems [71], and an OA-producing type III PKS from *Rhododendron dauricum* has been described [72]. 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 [73]. 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 [73]. 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. In this study, I therefore sought to alter the ligand specificity of the OA-responsive AraC-TAL14 variant, using computational protein design followed by directed evolution. Variant AraC-OA7, with 10 total amino acid substitutions relative to wild-type AraC, showed 10-fold increase in the OA specificity as compared to AraC-TAL14 and retained a low

background expression level comparable to that of wild-type AraC. Further directed evolution of the AraC-OA7 variant led to Variant AraC-OA8, with 2 additional amino acid substitutions relative to its parent AraC-OA7. The AraC-OA8 variant showed a 15-fold increase in the OA specificity as compared to AraC-TAL14 and still retained a low background expression level 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.

Table 1. Amino acid substitutions in the AraC-based sensors 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

Materials and Methods

Plasmid Construction

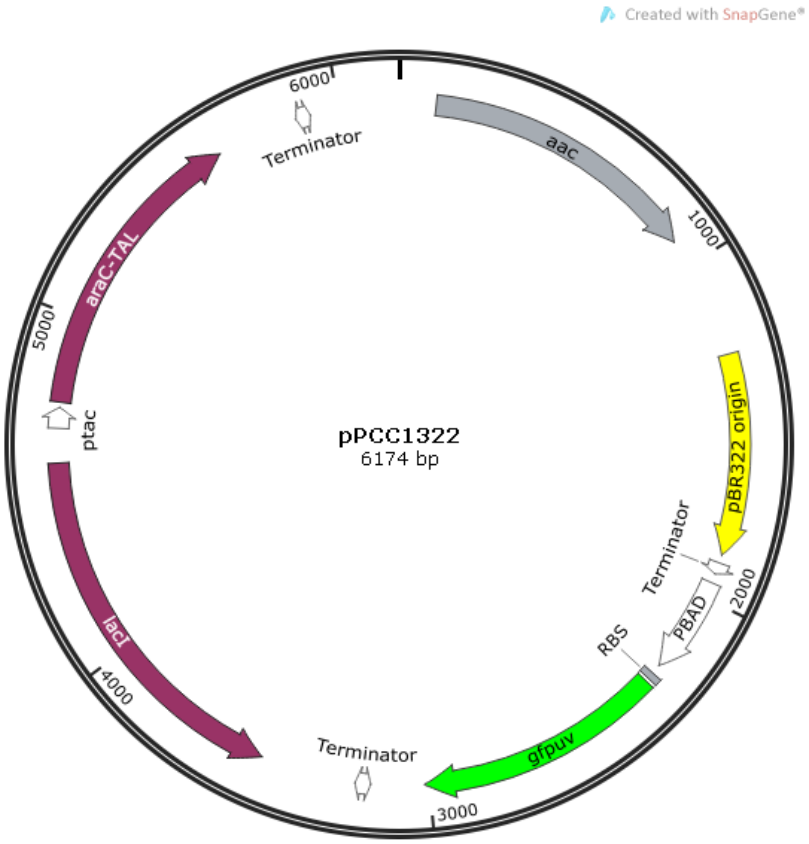

Figure 9. ***Plasmid map of pPCC1322***

Plasmid pPCC1322 (Figure 9) was constructed from pFG29-TAL [74] 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.

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 [75]. 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://tmcalculator.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 ligand binding domain (LBD) of the respective parent AraC variant using the Agilent GeneMorph® II Random Mutagenesis Kit. For libraries based on AraC-OA6 and AraC-OA7 as parents (GFP fluorescence-assay-based screening), the library insert was ligated into pPCC1322 vector. Library error rate was determined by sequencing ten random library clones: error rate was ~4.2 mutations/kb. Combinatorial assembly of mutations of the five single- or double-amino acid substitutions returned by iterative protein redesign and optimization approach (IPRO [76]), (those in AraC-OA1 through AraC-OA5) was performed by assembly of purified PCR fragments amplified from each mutant. This approach generated a library representing all 32 combinations. Primers were designed such that each fragment covers one mutation. For two mutations that are close together, these sites were instead recombined on a single PCR fragment using all four primer combinations. The fragments were then assembled with Gibson Assembly onto pPCC1322 vector for fluorescence-assay-based screening. Sequencing of randomly picked clones confirmed correct library assembly.

Iterative Protein Redesign and Optimization 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 iterative protein redesign and optimization approach (IPRO [76]), 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 [18], in addition to residue positions that appeared in new TAL-specific AraC variants (AraC-TAL12 through AraC-TAL17) 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 more than 25% of the AraC-TAL14), the top five designs with highest OA/TAL interaction energy ratios and with single or double AA 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: Colonies of fresh transformants (always using strain HF19

[18]) were inoculated into 500 µL LB supplemented with 50 µg/mL apramycin. After 6-12 hours of growth at 37 °C 250 RPM, the cultures were then diluted into fresh LB containing 50 µg/mL apramycin and 100 µM IPTG, with or without inducer(s) of interest at final concentrations as indicated in the presented results. The subcultures were grown for 4 hours at 37 °C 250 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 (OD_{595}) was measured on BMG Labtech NOVOstar Microplate reader. Fluorescence intensity was normalized by OD_{595}. Fold-induced GFP expression was calculated by dividing the normalized fluorescence intensity in presence of the inducer by the normalized fluorescence intensity without inducer.

For microtiter plate-based fluorescence screening of AraC libraries, unique colonies of fresh HF19 transformants (450-500 colonies from each library) were picked and inoculated into the wells of 96-deep-well plates containing 500 µL LB supplemented with 50 µg/mL apramycin. 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, cultures were screened thrice: with no inducer, 0.5 mM OA, and 3 mM TAL. For the random mutagenesis library screening was performed with no inducer, 1 mM OA, and 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), while the best clones from

the random mutation 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.

Results

Design of a sensor with enhanced specificity toward Orsellinic Acid

Among other PKS systems, orcinol synthase from *Rhododendron dauricum* (RdOrs) is a type III PKS that produces orsellinic acid (OA), using acetyl-CoA and malonyl-CoA starter molecules [72]. OA biosynthesis proceeds through the formation of a tetraketide intermediate, with TAL appearing as a byproduct due to lactonization of the triketide precursor [72]. 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 [73]. 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 [73]. 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. I therefore sought to alter the ligand specificity of the previously identified and OA-responsive AraC-TAL14 variant. To develop the OA sensor, I first turned to computational modeling and binding calculations for insights into improving specificity toward OA.

Computational models of existing AraC-TAL variants (1, 12, 14, 15, and 19) were generated using the Mutator program [76] by first substituting the appropriate amino acids onto the structure of wild-type AraC (PDB accession: 2ARC [77]), followed by an energy

minimization step for clash-free re-packing of the amino acid side chains that lie within 10 Å of the altered residue. Interaction energy scores between the AraC variants and ligands TAL and OA were computed using the CHARMM energy function [78] and used as a proxy for binding affinity. Table 2 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 2) correlate well with the ratios of fold-induced GFP expression (OA/TAL) ($R^2 = 0.75$), thereby indicating that relative interaction energy scores reasonably capture ligand specificity. Of the AraC-TAL variants, AraC-TAL14 demonstrated the strongest response to OA, while retaining a low background GFP expression in the absence of TAL (Table 2). The modeled structure of AraC-TAL14 was therefore chosen as the starting point for computational design.

Table 2. Relative CHARMM interaction energy scores and experimentally measured specifi city ratio of AraC-TAL1 variants with OA and TAL ligands.

Variant	**CHARMM interaction energy score with OA**	**Fold-increase in GFP with OA (1 mM) (A)**	**CHARMM interaction energy score with TAL**	**Fold - increase in GFP with TAL (1 mM) (B)**	***In silico* relative interacti on score OA/ TAL**	**OA specificity (OA/ TAL, 1 mM each) (A/B)**
Wild-type	-109.3	1.0	-99.2	0.9	1.1	1.1
AraC-TAL1	-121.8	8.5	-116.6	5.7	1.0	1.5
AraC-TAL12	-121.7	5.6	-124.4	13.5	1.0	0.3
AraC-TAL14	-123.8	9.9	-113.2	5.2	1.1	1.9
AraC-TAL15	-116.4	6.4	-102.3	2.3	1.1	2.8
AraC-TAL^{+}/OA^{-}	-98.1	0.4	-110.8	3.7	0.9	0.1

OA-specific AraC variants were obtained using an iterative protein redesign and optimization approach (IPRO [76]); refer to Methods section). The top five designs (OA/TAL interaction energy ratios) as identified by IPRO, named AraC-OA1 through AraC-OA5, are listed in Table 3. Note that AraC-OA1,3, and 4 each has a single amino acid substitution relative to AraC-TAL14, while AraC-OA2 and AraC-OA5 each have two substitutions relative to AraC-TAL14. These five variants were constructed and 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 3 (fold-induced GFP expression for all variants is provided in Table 3 as well). Only AraC-OA1 and AraC-OA4 showed any significant response to OA (>

2-fold change in fluorescence when induced), and neither of these variants showed higher specificity toward OA, as compared to AraC-TAL14.

Table 3. Relative interaction scores and specificity ratios of the computationally predicted OA-preferring variants computed using CHARMM force-field calculations and also experiments

Variant	**Amino acid substitutions in AraC-TAL14**	**CHARMM interaction energy score with OA**	**CHARMM interaction energy score with TAL**	***In silico* relative interaction score OA/ TAL**	**Fold increase in GFP with OA (1.5 mM) (N=3, ±SD)**	**OA Specificity from experiment [OA] (1.5 mM)/ [TAL] (3 mM)**
AraC-TAL14		-123.8	-113.2	1.1	6.6 ± 0.3	0.7
AraC-OA1	V8D	-185.8	-91.4	2.0	4.0 ± 0.9	0.1
AraC-OA2	I24D,P25G	-150	-100.3	1.5	0.9 ± 0.1	0.7
AraC-OA3	V72D	-140	-102.9	1.4	0.9 ± 0.1	0.8
AraC-OA4	L82M	-183.2	-85.8	2.1	2.3 ± 0.1	0.4
AraC-OA5	G135L,N139G	-173.4	-93.4	1.9	0.8 ± 0.1	0.7

I next screened a small library comprising all combinations of the five single or double AA substitutions returned by IPRO (those in AraC-OA1 through AraC-OA5), with AraC-TAL14 as parent (32 combinations). Using the P_{BAD}-*gfp* reporter construct, 450 total clones were screened by microtiter plate-based fluorescence measurement, 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 screened variants showed either no response to OA or TAL or showed very high background GFP expression. However, four somewhat promising clones were selected, and

sequencing revealed these were all identical, carrying substitutions P25G and N139G within the AraC-TAL14 parent. As shown in Table 4 this variant showed low background expression and low GFP expression response to 3 mM TAL, but also a lower response to 1.5 mM OA, as compared to AraC-TAL14. A single substitution analysis of either P25G alone or N139G alone in the AraC-TAL14 parent revealed that P25G alone increases OA specificity to a level similar to that of the AraC-TAL14 parent, while retaining reduced response to TAL (Table 4). AraC-TAL14/P25G was renamed AraC-OA6 and this variant was selected as the parent for engineering further improvements in OA response and specificity, using directed evolution.

Table 4. Summary of results from analysis of 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, ±SD) (A)**	**Fold-induced GFP expression with 3.0 mM TAL (N=3, ±SD) (B)**	**OA Specificity (A/B)**
AraC-TAL14	126 ± 23	6.6 ± 0.3	10 ± 0.5	0.7
AraC-TAL14/ P25G/N139G	118 ± 14	3.8 ± 0.4	4.6 ± 0.3	0.8
AraC-TAL14/ P25G (AraC-OA6)	223 ± 12	3.6 ± 0.5	3.8 ± 0.3	0.9
AraC-TAL14/ N139G	84 ± 11	3.1 ± 0.2	6.9 ± 0.3	0.5

Directed Evolution of an OA-Specific AraC Variant

To improve the response of AraC-OA6 towards OA I subjected a library consisting of random point mutations in the AraC-OA6 ligand binding domain to microtiter plate-based fluorescence using our P_{BAD}-gfp reporter. The library was first screened in the presence vs.

absence of 1.5 mM OA, followed by 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 specificity to OA (Figure 11A), and showed a background expression comparable to wild-type AraC. This resulted in a 10-fold increase in OA specificity as compared to AraC-TAL14, and a 15-fold increase compared to AraC-TAL1 (Figure 10). AraC-OA7 carries three amino acid substitutions relative to its parent (Table 1). Analysis of the individual contributions of each of these three substitutions is presented in Table 5. I 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 our highest specificity variant AraC-OA8 (Figure 10 and Figure 11B). AraC-OA8 shows a 15-fold increase in OA specificity as compared to AraC-TAL14 (24-fold compared to AraC-TAL1) and retains low background expression (Figure 10). Further, half-maximum dose response for AraC-OA8 occurs at ~1 mM OA thereby demonstrating an increased sensitivity to AO as compared to AraC-TAL14 variant (Figure 12). AraC-OA8 carries two amino acid substitutions relative to its parent (Table 1), and 12 total amino acid changes as compared to wild-type AraC. I expect that variants with further improvements in OA specificity and/or sensitivity would be readily discovered through continued rounds of directed evolution.

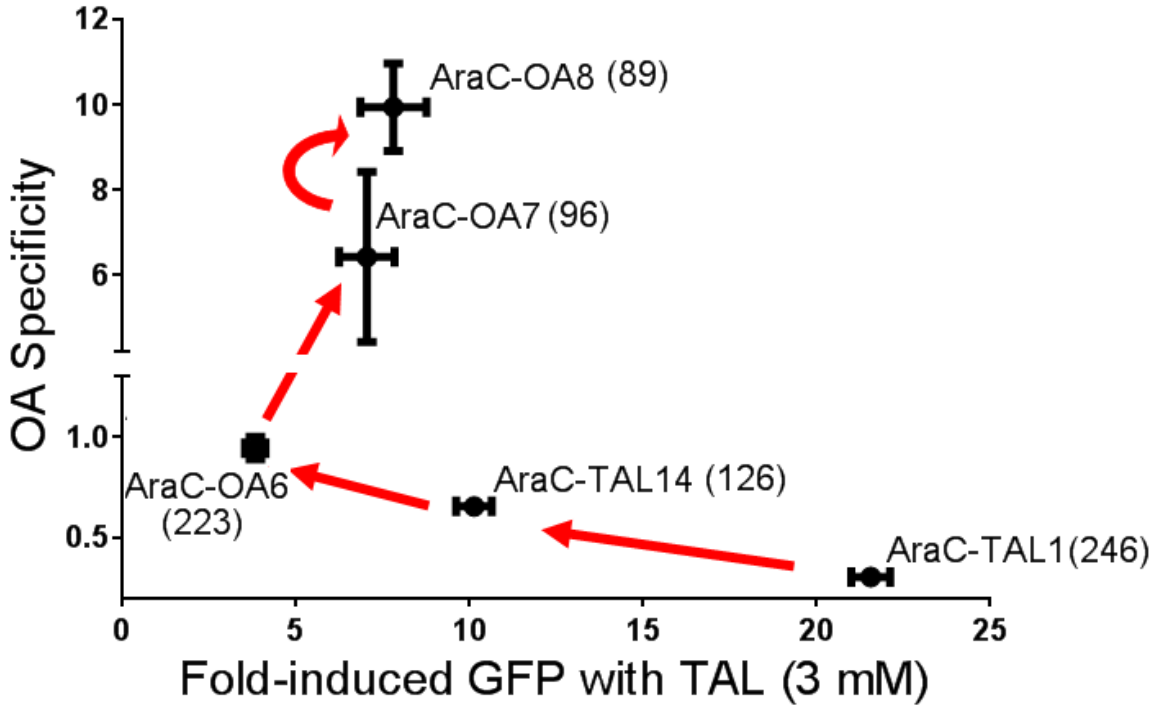

Figure 10. ***Evolving OA specificity in AraC-based biosensors****. OA specificity ratios (fold-induced GFP with 1.5 mM OA/ fold-induced GFP with 3 mM TAL) and fold-induced GFP expression with TAL (3 mM) indicate a directed increase in OA specificity of the AraC-TAL14 based variants. Data points are the average of three values, and error bars represent the standard deviation.*

Table 5. Fold-induced GFP expression values and OA specificities of AraC-TAL14 variants bearing single amino acid substitutions identified in AraC-OA7

AraC-TAL14 variant	**Background GFP**	**Fold-induced GFP with 1.5 mM OA (N=3, ±SD) (A)**	**Fold-induced GFP with 3.0 mM TAL (N=3, ±SD) (B)**	**OA Specificity (A/B)**
AraC-TAL14	126 ± 23	6.6 ± 0.3	10 ± 0.5	0.7
AraC-TAL14/P25G (AraC-OA6)	223 ± 12	3.6 ± 0.5	3.8 ± 0.3	0.9
AraC-TAL14/P25G /I36F/P128S/A140V (AraC-OA7)	96 ± 4	45 ± 14	7.1 ± 0.8	6.4
AraC-TAL14/P25G/ I36F	160 ± 14	53 ± 4	16 ± 2	3.4
AraC-TAL14/P25G/ P128S	268 ± 51	5.6 ± 0.1	7.4 ± 1.8	0.8
AraC-TAL14/P25G/ A140V	123 ± 20	2.4 ± 0	3.4 ± 0.4	0.7
AraC-TAL14/P25G/ I36F/P128S	204 ± 30	43 ± 2	15 ± 2	2.9
AraC-TAL14/P25G/ P128S/A140V	134 ± 10	2.4 ± 0	3.9 ± 0.2	0.6
AraC-TAL14/P25G/ I36F/ A140V	102 ± 4	38 ± 1	7.9 ± 1.1	4.9

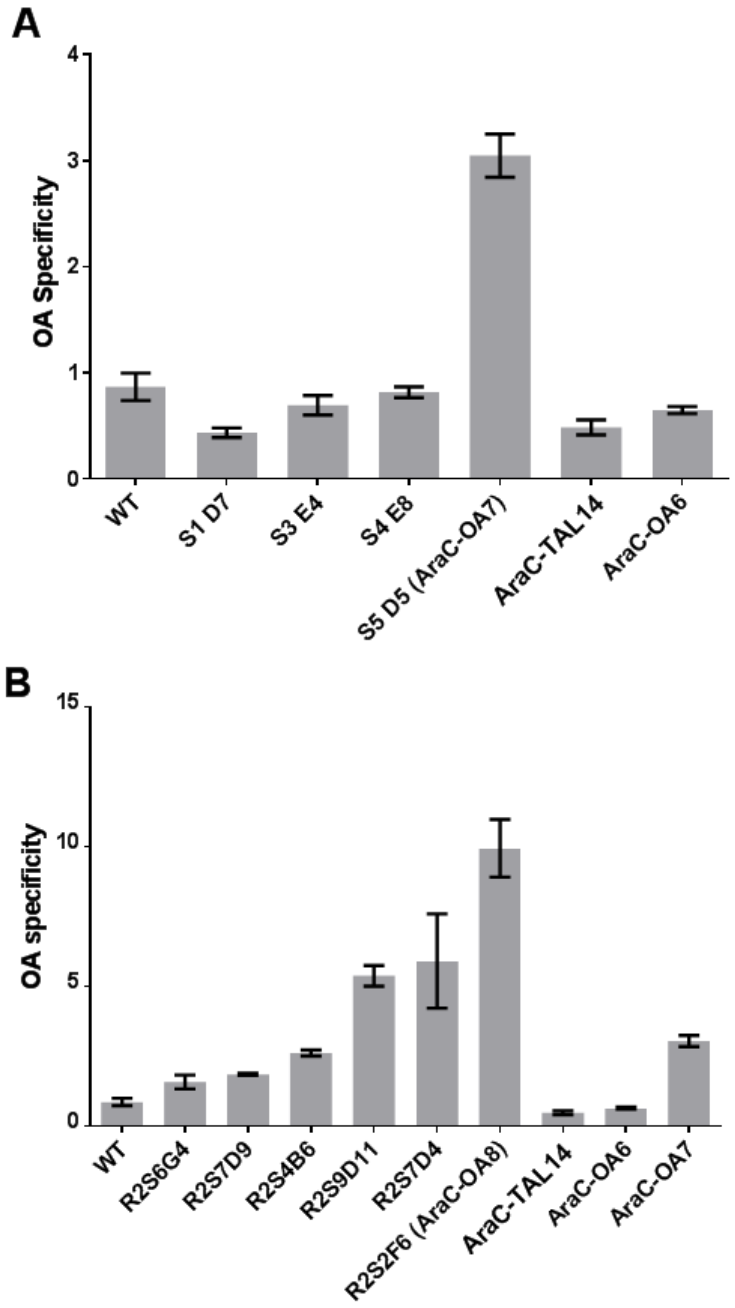

Figure 11. ***OA specificity values (ratio of induced GFP expression with 1.5 mM OA to induced GFP expression with 3 mM TAL) of unique variants isolated from two rounds of random mutagenesis on AraC-based sensor variants****. A) OA specificity of unique variants isolated from one round of RM o AraC-OA6. B) OA specificity of unique variants isolated from one round of RM on AraC-OA7. Data points are the average of three values, and error bars represent the standard deviation.*

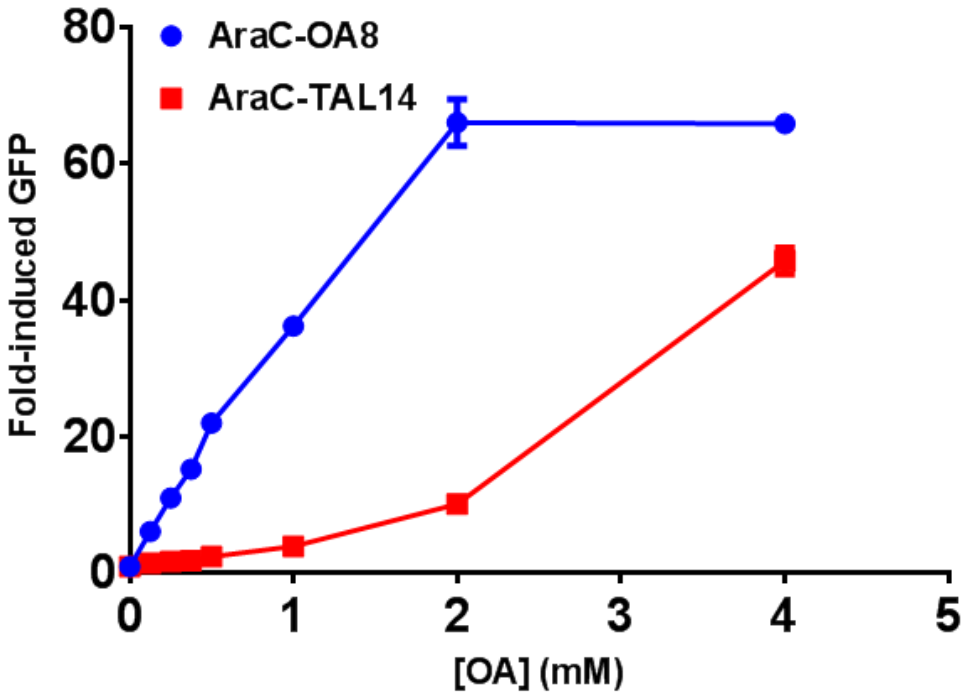

Figure 12. ***Dose responses of AraC-TAL14 and AraC-OA8 from 0-4 mM OA.*** *Data points are the average of three values, and error bars represent the standard deviation.*

Analysis of amino acid substitutions

Previous studies in our lab employed a TetA-based dual-selection to isolate six variants of AraC-TAL1 (AraC-TAL12 through AraC-TAL17) which have several improved biosensor properties as compared to the AraC-TAL1 sensor [70]. Of these existing variants, the AraC-TAL14 variant (with a L72V amino acid substitution in AraC-TAL1, Figure 13B) showed the highest specificity ratio for OA inducer. Hence in this present study, I chose AraC-TAL14 as a parent to evolve an OA sensor. AraC-OA6, derived from AraC-TAL14 after computational protein redesign, contains a single amino acid substitution (P25G) in the AraC-TAL14 variant and shows over 50% lower inducibility to TAL and OA inducers as compared to the AraC-TAL14 variant (Table 4). The P25G of the AraC-OA6 substitution

resides in the first β strand of the ligand binding pocket of AraC-OA6 and constitutes the region in the vicinity of the ligand binding pocket of the AraC protein (Figure 13C). A previous study with the wildtype AraC protein has shown that substitution at the P25 residue (P25S) led to a 4-fold reduction in the mutant protein's response to arabinose as compared to wildtype possibly due to weakened interaction between the N-terminal arm and DBD, as this interaction is believed to maintain tight repression [79]. Other studies with the wildtype AraC protein have also revealed that the preservation of the shape of the N-terminal arm and the protein's hydrophobic cluster (formed by residues L9, Y13, F15, W95, and Y97) are important for an inducible response in presence of arabinose [80, 81]. Substitution P25G may weaken these hydrophobic residue interactions, leading to lower response of AraC-OA6 towards TAL and OA. Detailed structural data will be essential to investigate the role of such amino acid mutations in influencing the sensor's properties.

Despite the inferior OA inducibility of the AraC-OA6 variant (Table 4), I chose this "best" of many inferior variants to continue evolving an OA specific sensor by inserting random point mutations in the AraC-OA6 LBD. I reasoned that using an OA-responsive variant (also having only a low response to TAL) as a template, would make a good starting point instead of the wild type AraC protein for subsequent random mutagenesis. The random mutagenesis efforts led to a notably improved AraC-OA7 variant (Figure 11A) which bears three additional amino acid substitutions in the AraC-OA6 variant: I36F, P128S, and A140V. Results from amino acid substitution analysis of the three AraC-OA7 amino acids within AraC-OA6 are given in Table 5. As seen in Table 5, while I36F (Figure 13D) alone enhances response of AraC-OA6 towards OA, all three new substitutions are needed to achieve the

highest specificity and lowest background expression. Such cooperative effect of multiple amino acid substitutions was previously also reported for other AraC-based sensors [74].

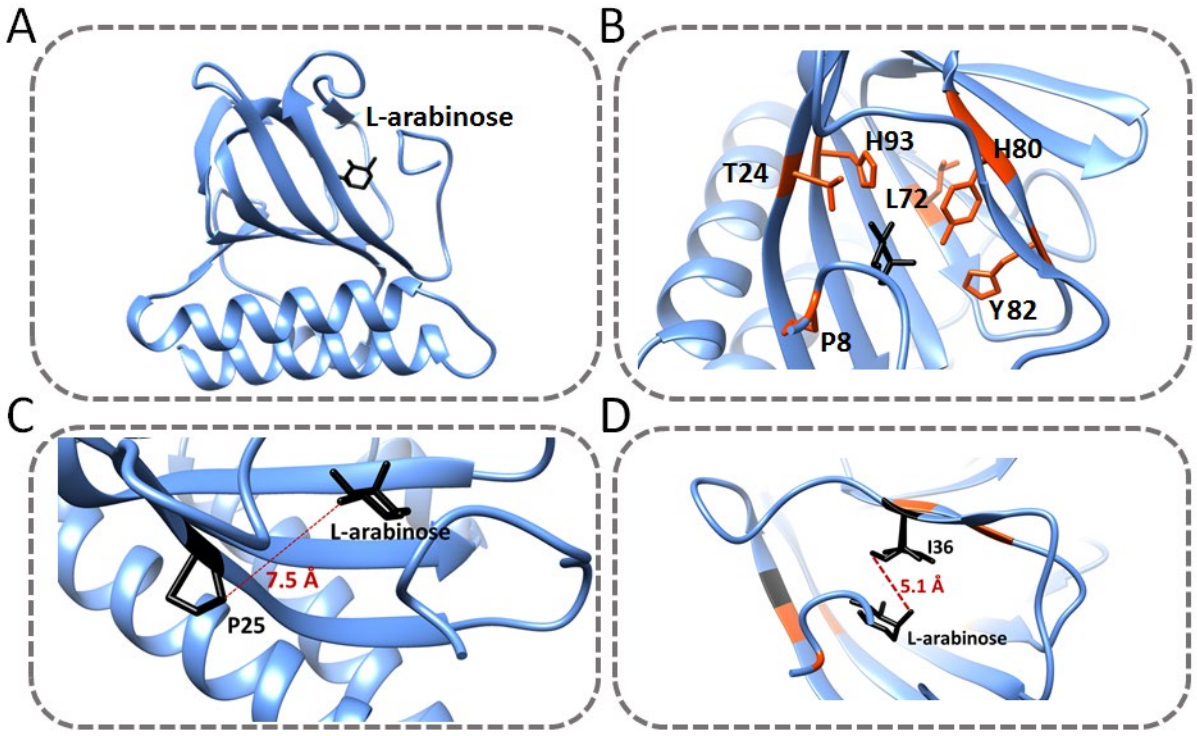

Figure 13. ***The amino acid substitutions of AraC-TAL14 variants shown on wild type AraC.*** *A) Wild type AraC crystal structure (PDBID: 2ARC). L-arabinose that is co-crystalized in WT-AraC is also shown. B) Wild type AraC structure showing the six sites that have amino acid substitutions in AraC-TAL14 variant, C) AraC-OA6 contains an additional P25G amino acid substitution in AraC-TAL14. The distance between the P25 and L-arabinose is indicated in the figure. D) The I36F amnio acid substitution in AraC-OA7 alone increases response to OA. The distance between the I36 and L-arabinose is indicated in the figure*

A second round of random mutagenesis on the AraC-OA7 variant led to AraC-OA8 (bearing H81Q and Q142L substitutions) which showed further improvement in OA inducibility and OA specificity as compared to the existing AraC-based sensor variants (Figure 10). With the addition of total 12 amino acid substitutions in AraC-OA8 as compared to wild type AraC, a detailed structural analysis of AraC-OA8 with OA and/or TAL may be

essential to understand the influence of amino acid residues on the specificity of the isolated AraC variants.

Conclusion

A biosensor that responds to OA could have potential use for engineering enhanced OA biosynthesis in a recombinant host, and for probing/engineering chain length control and substrate/product specificity in type III PKS enzymes. In this study, I altered the ligand specificity of the OA-responsive AraC-TAL14 variant, using computational protein design followed by directed evolution. Variant AraC-OA7, with three amino acid substitutions relative to AraC-TAL14, showed 10-fold increase in the OA specificity as compared to AraC-TAL14 and retained a low background expression level comparable to that of wild-type AraC. Analysis of the individual contributions of each of the three amino acid substitutions identified in AraC-OA7 reveal that the amino acid substitutions act cooperatively to increase the specificity of the AraC-OA7 variant relative to AraC-TAL14. Further directed evolution of the AraC-OA7 variant led to variant AraC-OA8, with two additional amino acid substitutions relative to its parent AraC-OA7. The AraC-OA8 variant showed a 15-fold increase in the OA specificity as compared to AraC-TAL14 and still retained a low background expression level comparable to that of wild-type AraC. Collectively, the new AraC variants described have potential utility in metabolic engineering as biosensors to facilitate engineering PKS specificity and improving OA biosynthesis in bacteria.

CHAPTER 3

DESIGN AND CHARACTERIZATION OF A SALICYLIC ACID-INDUCIBLE GENE EXPRESSION SYSTEM FOR JURKAT CELLS

Introduction

Gene regulatory systems that offer reliable and predictable control over gene expression, in mammalian cells, have broad applications that include basic biological discovery [82], epigenetic remodeling [83], and performing some therapeutic tasks [84]. A number of inducible systems regulated by steroid hormones [85], antibiotic [35], light [86], and other reagents [37, 87, 88] have been developed but their widespread application is currently limited due to several problems. For example, the use of naturally existing steroid hormones as inducers causes interference with host cellular processes [38], while using antibiotics like tetracyclines has been reported to cause changes in cell behavior [89-91] and risk of development of antibiotic resistance in host cells continuously induced with antibiotics [49]. Inducers like light, heat shock, and food and cosmetic preservatives may serve as promising new candidates for engineering therapeutic cells but the efficiency of these systems, over long-term use, is yet to be demonstrated [83]. It is thus desirable to develop additional gene control systems which would address the above-mentioned problems and add new tools, to the existing synthetic biology toolbox, allowing simultaneous and independent expression of several genes [55].

Salicylic acid (SA) or salicylate is a naturally occurring bioactive compound that functions as a major plant hormone involved in plant immune responses [92]. Free SA and its derivatives, like acetylated SA (Aspirin), have been known for their curative properties which include analgesic, anti-pyretic, and anti-inflammatory effects in humans [93]. The use of aspirin for secondary prevention of cardiovascular events has been widely reported [94-96]. A recent report suggested a 25% reduction in recurrent cardiovascular events, in a group of

patients with prior myocardial infarction, when a low dose of aspirin was taken daily [97]. Salicylates are inexpensive, easily available over the counter and are generally safe when used in low doses [98].

In this chapter, I describe the design of a SA-inducible gene expression system for regulating transgene expression in mammalian cell lines. The SA-inducible switch was designed based on the *Escherichia coli* transcriptional repressor- multiple antibiotic resistance repressor (MarR) protein. In *E. coli*, MarR represses the transcription of the multiple-antibiotic resistance *marRAB* operon by interacting as a homodimer with two palindromic DNA operator sites, called site 1 (*marO1 or O1*) and site 2 (*marO2 or O2*), located in the *marRAB* promoter region [99]. Sodium salicylate and other similar inducers bind to MarR ligand binding pockets and disrupt the MarR-*marO* (O1 and O2) interaction, resulting in the active transcription of the *marRAB* operon which mediates multidrug resistance in *E. coli* through over expression of an efflux pump [99].

By utilizing the SA-responsive MarR-*marO* interaction, I designed an inducible gene switch to regulate fluorescent reporter expression in human Jurkat cell lines. I first modified the bacterial MarR protein to ensure its expression in human cell lines (detailed description of the modifications is given in Chapter 1). Since modifications to the protein sequence may or may not alter its function, I validated the function of the modified MarR protein in *E. coli*. The SA_{ON} system, comprising the modified MarR protein along with the MarR-regulated promoter controlling the expression of an EGFP reporter, was integrated in Jurkat cells for further testing. This SA_{ON} system initially demonstrated a 1.5-fold increase in EGFP fluorescence in engineered Jurkat cells when induced with SA. Fusion of the MarR protein

with a nuclear-tagged maltose binding protein (MBP) further led to a total of 2.1-fold increase in EGFP expression in presence of salicylate inducer. The SA_{ON} system described here currently demonstrates a high background EGFP fluorescence in absence of the SA inducer, thus causing the low fold-induced EGFP expression levels when induced with SA. With further optimizations to reduce the background, I expect that the SA_{ON} system has potential to expand the existing synthetic biology toolbox to allow SA-inducible gene regulation in engineered cell lines.

Materials and Methods

Inducer and inducer concentrations

Sodium salicylate inducer (500 mM stock solution in water, Cat. No. S-3007) was purchased from Sigma-Aldrich. The solution was filter sterilized using a 0.2 μm sterile, PES syringe filter (Cat. No. 431229) and stored at 4 ^{o}C. The inducer concentration used for Jurkat cells varied from 0.5 mM, 1 mM, 2 mM, 3 mM, and 5 mM SA in the final media. Toxicity assays were carried out for upto 100 mM salicylate concentration. For bacterial studies, the inducer concentration ranged from 0.5 mM, 1 mM, 1.5 mM, 2 mM, 3 mM, 4 mM, and 5 mM in the final media.

Plasmids

All plasmids used in the Jurkat cell lines were lentiviral transfer plasmids derived from the vector LV-CD19-T2A-GZMBhis-T2A-GFP. The lentiviral plasmids contain the transgene flanked by long terminal repeat (LTR) sequences and are all replication incompetent which means that their presence should not result in the synthesis of viral proteins. A detailed list of all the plasmids used in this study are listed in Table 6.

Table 6. List of all plasmids used in this study

Plasmid	Description
LV-CD19-T2A-GZMBhis-T2A-GFP	Lentiviral vector used as backbone for all mammalian expression vectors used in this study
pPCC1624	$P_{EF1\alpha*}$ -driven EGFP expression vector used as reporter plasmid
pPCC1625	P_{hPGK} -driven MarR (N-term FLAG tag) expression vector
pPCC1625-1	P_{hPGK} -driven MarR (N-term NLS-tag) expression vector
pPCC1625-2	P_{hPGK} -driven MBP-MarR (N-term NLS-tag on MBP) fusion protein expression vector
pPCC1666-1	$P_{lacUV5*}$ -driven GFP reporter plasmid for *E. coli* experiments
pPCC1668	$P_{lacUV5*}$ -driven GFP and constitutive MarR protein (derived from *E. coli* MG1655) plasmid for *E. coli* experiments
pPCC1669-2	$P_{lacUV5*}$ -driven GFP and constitutive MarR protein (human codon optimized) plasmid for *E. coli* experiments

Table 6 continued

pPCC1669-4	$P_{lacUV5*}$ -driven GFP and constitutive N-term FLAG tagged MarR protein (human codon optimized) plasmid for *E. coli* experiments
pPCC1669-5	$P_{lacUV5*}$ -driven GFP and constitutive N-term NLS tagged MarR protein (human codon optimized) plasmid for *E. coli* experiments
pPCC1669-6	$P_{lacUV5*}$ -driven GFP and constitutive N-term NLS tagged MBP-MarR fusion protein (human codon optimized) plasmid for *E. coli* experiments
$P_{EF1\alpha*}$	$P_{EF1\alpha}$ promoter with added marO1 and marO2 operator sites
$P_{lacUV5*}$	P_{lacUV5} promoter with added marO1 and marO2 operator sites

Table 6 continued

MarR sequence (*E. coli* MG1655)	ATGAAAAGTACCAGCGATCTGTTCAATGAAATTATT CCATTGGGTCGTTTAATCCATATGGTTAATCAGAAG AAAGATCGCCTGCTTAACGAGTATCTGTCTCCGCTG GATATTACCGCGGCACAGTTTAAGGTGCTCTGCTCT ATCCGCTGCGCGGCGTGTATTACTCCGGTTGAACTG AAAAAGGTATTGTCGGTCGACCTGGGAGCACTGACC CGTATGCTGGATCGCCTGGTCTGTAAAGGCTGGGTG GAAAGGTTGCCGAACCCGAATGACAAGCGCGGCGT ACTGGTAAAACTTACCACCGGCGGCGCGGCAATATG TGAACAATGCCATCAATTAGTTGGCCAGGACCTGCA CCAAGAATTAACAAAAAACCTGACGGCGGACGAAG TGGCAACACTTGAGTATTTGCTTAAGAAAGTCCTGC CGTAA
$P_{EF1\alpha*}$ sequence	GGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCG CCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGG CAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGG TAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCC TTTTTCCCGAGGGTGGGGGAGAACCGTATATAAATA CTTGCCTGGGCAATATTATTCTTTACGTCTCATTACT TGCCAGGGCAACTAAT

Table 6 continued

codon optimized MarR sequence	ATGGGTAAATCCACCTCAGACTTGTTTAATGAAATC ATACCATTGGGGAGATTGATTCACATGGTAAACCAA AAGAAAGATAGACTCCTCAACGAATATCTGTCTCCG TTGGACATAACTGCGGCCCAGTTCAAAGTTCTTTGTT CAATCCGATGCGCGGCTTGCATAACCCCGGTGGAAC TTAAGAAGGTCTTGAGTGTCGATCTCGGTGCCTTGA CGCGCATGTTGGATCGACTGGTTTGTAAGGGATGGG TTGAACGACTCCCGAATCCAAATGACAAGCGAGGTG TACTTGTAAAGCTGACCACAGGTGGTGCCGCCATCT GTGAACAGTGCCATCAATTGGTCGGGCAGGATTTGC ACCAGGAATTGACCAAGAACTTGACGGCAGATGAA GTGGCGACATTGGAATATCTCCTCAAAAAAGTCCTC CCTTAA
$P_{lacUV5*}$ sequence	TTTACACTTTATGCTTCCGGCTCGTATAATGATACTT GCCTGGGCAATATTATCCCCTGCAACTAATTACTTGC CAGGGCAACTAAT

$P_{EF1\alpha}$-marO(2X)-EGFP

A pair of the native operator sites (*marO*1 and *marO*2) derived from *E. coli* MG1655 strain were introduced within the minimal EF1aplha promoter, a strong constitutive promoter, driving the expression of the enhanced green fluorescent protein (*egfp*) reporter gene. The following steps were used to achieve this: a unique *AgeI* and *EcoRI* site on the LV vector

were used to introduce a PCR amplified fragment containing a pair of mar operator sequences immediately downstream of the TATA box followed by the EF1α intron and the *egfp* reporter sequence. Previous studies have reported a decrease in the promoter activity when the distance between the TATA box and the initiator (Inr) site (required for enhancing the strength of the promoter through increased binding to the RNA polymerase) was increased in order to accommodate operator sites and the promoter activity was restored to normal when this distance was kept same as the wild-type [100]. Hence, the insertion of the mar operator sequences was done such that the distance between the EF1α TATA box and the Inr site was maintained as in the wild-type position. The ends of the PCR fragments were designed such that they had a minimum of 20 bp overlapping region with the *AgeI* and *EcoRI* digested vector. The PCR fragment was then inserted into the double digested vector by using Gibson assembly technique. The resulting plasmid is called pPCC1624 which contains the $P_{EF1\alpha}$-*marO*(2X)-*EGFP* reporter cassette.

P_{PGK}-MarR-IRES-mCherry

The plasmid pPCC1625 has the codon optimized repressor (MarR) expressed under the control of a constitutive human PGK promoter. The optimized *marR* sequence was ordered as a gene block from IDT Technologies and is tagged with a single FLAG sequence at its N terminus. To enable FACS based screening of Jurkat cells containing the pPCC1625 plasmid, an internal ribosome entry sequence (IRES) and mCherry fluorescent protein sequence were added downstream of the marR sequence. With this construct both the FLAG-tagged MarR and mCherry proteins are expressed from the single PGK promoter. The transgene P_{PGK}-MarR-IRES-mCherry is flanked by the LTR sequences of the LV vector.

P_{PGK}-(NLS)MarR-IRES-mCherry / P_{PGK}-(NLS)MBP-MarR-IRES-mCherry

To increase the nuclear availability of the MarR protein a nuclear localization signal (NLS) derived from the Simian Vacuolating Virus (SV40 large T antigen) was inserted at the N terminus of the codon optimized marR sequence. A previously described [101] linker sequence (NSSSNNNNNNNNNNLG) was inserted between the NLS and *marR* gene to ensure proper folding of the repressor protein without possible interference from the charged NLS sequence. Previous studies have reported an increase in the solubility of recombinant proteins, expressed in mammalian cells, when fused to the *E. coli* maltose binding protein (MBP) [102]. Hence to enhance the expression of MarR in Jurkat cells and to demonstrate fusions of the repressor protein to other functional domains, a construct containing the nuclear-tagged MBP protein fused to the N terminus of the MarR repressor protein was also developed. The MBP protein was separated from the fused MarR protein via the NSSSNNNNNNNNNLG linker. All the additions were made using PCR based methods along with the use of Gibson assembly to generate plasmids pPCC1625-1 (P_{PGK}-(NLS)MarR-IRES-mCherry) and pPCC1625-2 (P_{PGK}-(NLS)MBP-MarR-IRES-mCherry) from the pPCC1625 plasmid.

Cells and transient transfection

HEK293 Meter cells were used for transient transfections and lentiviral production. The cells were maintained in a 1:1 mixture DMEM / F-12 basal media supplemented with 10% heat inactivated FBS, 5% Penicillin-Streptomycin antibiotic solution, and 2 mM L-glutamine. Jurkat cell lines were used for stable reporter and/or repressor expression. These cells were maintained in R10 media containing RPMI basal media supplemented with 10%

heat inactivated fetal bovine serum (FBS), 5% Penicillin-Streptomycin antibiotic solution, and 2 mM L-glutamine. All reagents for media preparation were purchased from Thermofisher Scientific.

Transient transfections, to test the reporter expression, were performed in the HEK293 Meter cell lines using the Lipofectamine™ LTX Reagent with PLUS™ Reagent purchased from Thermofisher Scientific (Cat. No. 15338030) using standard protocol. Briefly, 0.5 µg of the DNA was mixed with 0.5 µl of the Plus reagent in 25 µl of the OptiMEM media (Cat no. 31985070). This DNA mix was further added to an LTX mix containing 1.5 µl of the Lipofectamine™ LTX reagent in 25 µl OptiMEM media. The final DNA:LTX concentration was optimized at 1:3 and incubated at room temperature for 30 minutes. The 50 µl DNA:LTX mix was added to respective wells of a 24-well tissue culture treated plate containing HEK cells cultured to 60%-70% confluency and were incubated at 37 °C and 5% CO_2 concentration for 24 hours. The media was later replaced with fresh media and cells were observed for reporter expression 24 hours and 48 hours post transfection using the ZEISS Axio Observer Z1 fluorescent microscope under the required emission and excitation filters.

Stable cell lines

All analysis in Jurkat cells were done using cell lines stably expressing the reporter and/or repressor protein. The stable cells were generated using standard lentiviral infection protocol as described below.

Lentivirus production

For lentiviral production, HEK293 meter cells were seeded at $6x10^6$ cells in complete DMEM-F12 media without antibiotics in T25 flask to reach around 90% confluency. The

cells were co-transfected with the transgene containing (pPCC1624/ pPCC1625/ pPCC1625-1/ pPCC1625-2) plasmid, and the psPAX2 and pMD2.g packaging plasmids in a 6:3:5 ratio using the previously described transient transfection protocol. After 24 hours the media was discarded and replaced with fresh media. The medium containing the viral particles was harvested 48 hours post transfection while adding fresh media. After 72 hours post transfection, the media was collected again and pooled with the previously collected media. The cells were discarded safely after bleaching. The viral particles in the collected media supernatant (10 ml) were concentrated using Amicon ultra 15 centrifugal filters (cutoff 100 kDa) to a final volume of 500 µl and aliquoted and stored at -80 ^{o}C until further use.

Lentiviral Infection

Jurkat-*reporter* cells were generated by infecting the Jurkat cell lines with lentiviral particles containing pPCC1624. On the day of transduction, Jurkat cells were seeded in a 6-well plate at 1×10^6 cells per well in 1 ml of R10 media without antibiotics. The cells were incubated with 10 µl of the previously prepared virus and the plate was centrifuged at 1000 rpm for 30 minutes. The plate was incubated at 37 ^{o}C and the media was changed for fresh R10 followed by additional incubation at 37 °C till the cells grow to 1×10^6 cells after which they were FACS sorted to collect the *EGFP* expressing, Jurkat-*reporter* cell line.

Jurkat-*reporter-repressor* cells were generated by infecting the *EGFP*-positive Jurkat-*reporter* cells with lentiviral particles containing pPCC1625/ pPCC1625-1/ pPCC1625-2 plasmids using the protocol described above. The infected cells were FACS sorted to collect the top 10% of *EGFP* and mCherry double positive cells with more focus on a high mCherry

expressing population. These sorted cells would thus be expressing both the reporter and repressor constructs.

Bacterial cells and culturing conditions

All modifications done on the MarR protein were first validated for proper function in *E. coli* K-12 strain JW5248-1 with marR deletion [103]. *E. coli* were grown in lysogeny broth (LB) media at 37 ^{o}C supplemented with 50 μg/ml apramycin when required. Plasmid pPCC1666-1 was used as the vector (see Table 6 for description of all bacterial plasmids used in this study) for the *E. coli* experiments. A solution of sodium salicylate was added to the media at required concentrations to analyze the repressor function.

Plate-based fluorescence screening

E. coli JW528-1 competent cells were transformed with the plasmids containing either native or modified (codon optimized, FLAG-tagged, NLS-tagged) MarR on pCC1666-1 backbone plasmid. Colonies of fresh transformants were inoculated into 5 ml LB supplemented with 50 μg/mL apramycin. After ~16 hours of growth at 37 °C, 250 RPM on Barnstead Max Q 5000 floor-model shaker, the cultures were diluted into fresh 500 μl LB containing 50 μg/mL apramycin, and SA at varying concentrations. The cells were subcultured at a final optical density (OD_{595}) of 0.2. The subcultures were grown for 4 hours at 37 °C, 900 RPM on Heidolph^{TM} Titramax 1000 vibrating platform shaker before washing with PBS buffer. Fluorescence was measured on SpectraMax® Gemini^{TM} EM Microplate Spectrofluorometer from Molecular Devices® (Excitation at 395 nm and emission at 509 nm). OD_{595} was measured on BMG Labtech NOVOstar Microplate reader. EGFP fluorescence was normalized to the OD_{595} measurements. Fold-induced GFP expression was

calculated as ratio of normalized GFP fluorescence with SA to normalized GFP fluorescence without SA. All data points reported are the average values of three independent replicates and error bars represent the standard deviation.

FACS-mediated cell sorting and Flow analysis

Jurkat cell expressing *EGFP* (488 nm laser, 530/30 emission filter) and/or mCherry (561 nm laser, 610/20 emission filter) were sorted or analyzed using a Becton Dickinson FACSAria™ III while excluding dead cells. Untreated Jurkat cells were used as negative controls. For flow analysis, Jurkat-*reporter* cells and Jurkat-*reporter-repressor* cells were induced with increasing concentrations of SA (0, 0.5 mM, 1 mM, 2 mM, 3 mM, and 5 mM) for 24 hours or 48 hours and then analyzed for increasing *EGFP* expression. All flow analyzed data points reported are values obtained from a single biological study.

Results

Validating MarR function in E.coli

To ensure successful transfer of the *E. coli* multiple antibiotic resistance repressor (MarR) protein into human Jurkat cell lines the repressor protein was codon optimized to maximize the protein expression in human cell lines. Further, several tags (FLAG, NLS, MBP) were fused to the MarR protein to enable protein detection, nuclear localization, and high-level protein expression in mammalian cells. Previous studies have reported that C-terminus MarR fusions interfere with repressor function probably by interfering with the protein dimerization [104]. Hence, all fusions were added to the N-terminus of the MarR protein. Each fusion to codon optimized MarR (hMarR) was tested for proper function in *E. coli* before transferring the constructs into Jurkat cell lines. The tests revealed that the

modified hMarR repressor was able to regulate reporter expression in *E. coli* when induced with a range of SA concentrations (Figure 14).

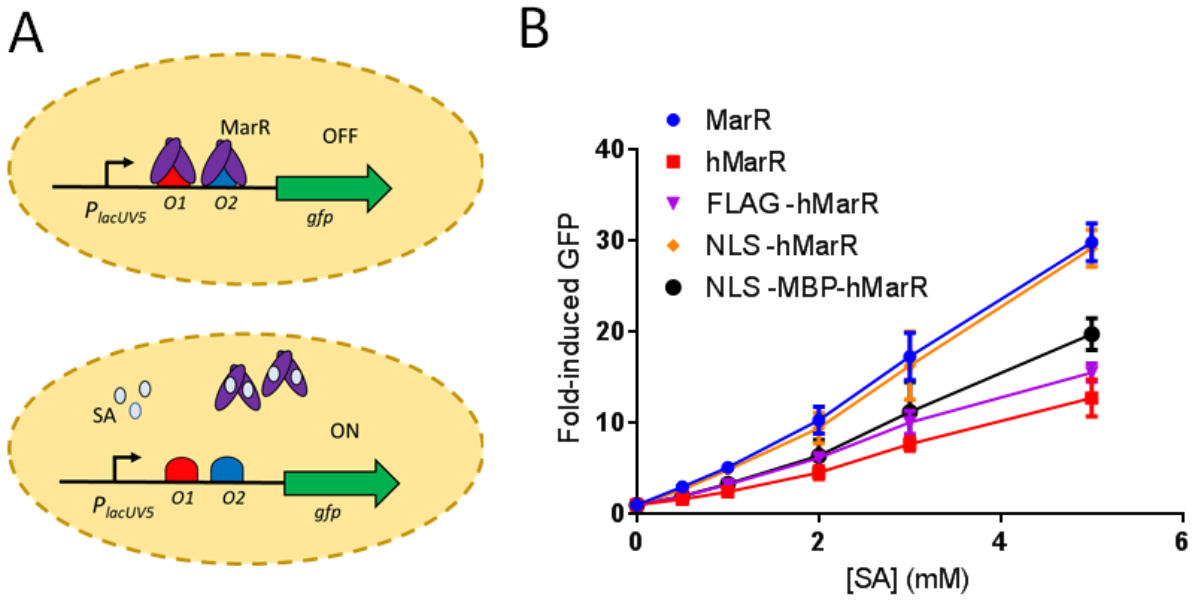

Figure 14. ***Validating MarR modification in E. coli****. A) Schematic of target gene regulation by MarR in E. coli. B) Graph of fold-induced GFP expression values with increasing salicylate concentrations when modified MarR is used to regulate GFP expression in E. coli. The modifications compared with E. coli MarR include codon optimized MarR (hMarR), hMarR with FLAG-tag on N-terminus, N-terminus NLS-tagged hMarR, and hMarR fused to NLS-tagged MBP protein. Data points are average values of three independent replicates and error bars represent the standard deviation (SD).*

Further, MarR regulates gene expression through binding two mar operator sites (marO) in the promoter region and thus blocking the RNA polymerase from transcribing the downstream genes. The marO1 (O1) and marO2 (O2) sites contain the same palindromic repeat (5'-TTGCC-3') but slightly different flanking sequences (Table 6). While previous studies have demonstrated that a single operator site is sufficient for regulating genes through MarR [105], I postulated that having multiple operator sites may prove beneficial by allowing

tight gene repression through MarR [106]. Hence, I sought to determine the best combination of marO sites, in *E. coli*, that together lead to tight gene repression (low background in absence of SA) and a high gfp expression when induced with SA. Figure 15 indicates that all combinations of the marO sites lead to similar levels of fold-induced GFP expression when induced with SA. Interestingly, the specific combination of O1 and O2 sites also retains a low background expression without SA (Figure 15). Thus, I used this combination of O1 and O2 operators for subsequent testing in *E. coli* and Jurkat cells.

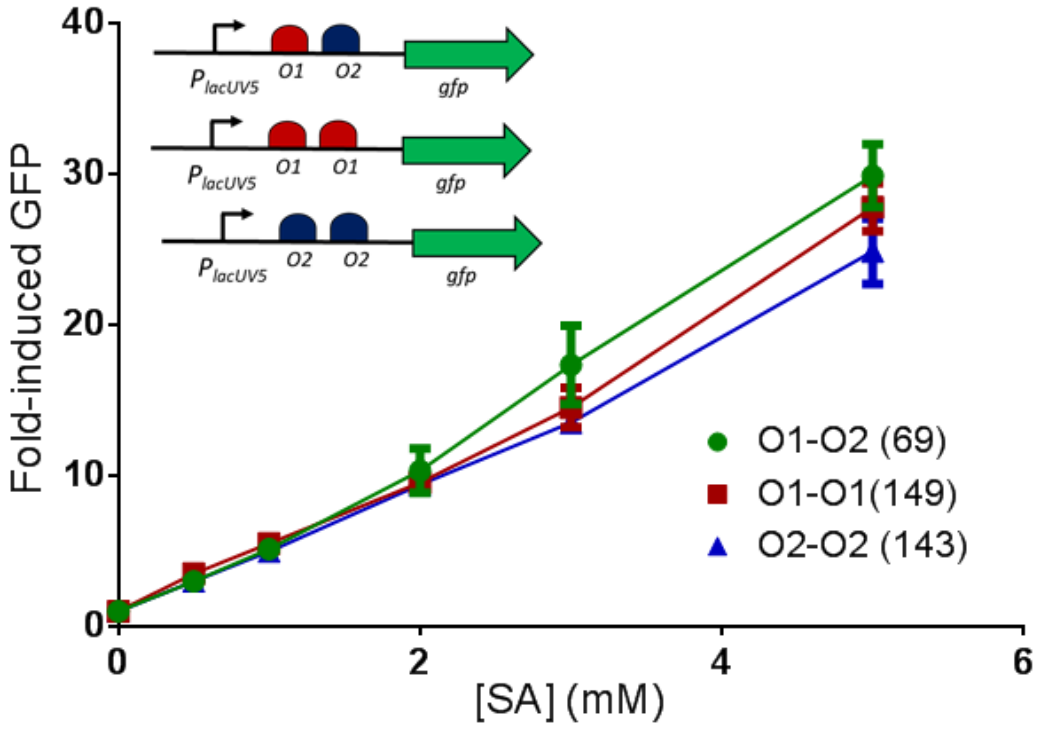

Figure 15. ***Operator combinations leading to tight and inducible gene regulation by MarR.*** *Various combinations of marO1 (O1) and marO2 (O2) sites demonstrate SA-inducible EGFP expression in E. coli. All data points are average of three values and error bars represent the SD. Background EGFP expression levels when no SA was added are given in parentheses.*

SA-inducible EGFP expression in Jurkat cell lines

Jurkat cells stably expressing the MarR protein (pPCC1625) and the EGFP reporter construct were generated using lentiviral transduction method (see Materials and Methods). The engineered Jurkat cells contained egfp gene under the control of a chimeric EF1α promoter containing the marO sites (see $P_{EF1\alpha*}$ in Table 6) while marR was constitutively expressed under the control of the human phosphoglycerate kinase 1 promoter (P_{PGK}). A high background EGFP fluorescence was seen even when no SA was added in the media and addition of 5 mM SA lead to a maximum of 1.5-fold increase in the fold-induced EGFP expression values, 24 hours post induction (Figure 16). These initial results suggest that the bacterial protein MarR can regulate genes in Jurkat cell lines in a SA-inducible manner, albeit with a high background EGFP expression even with no SA.

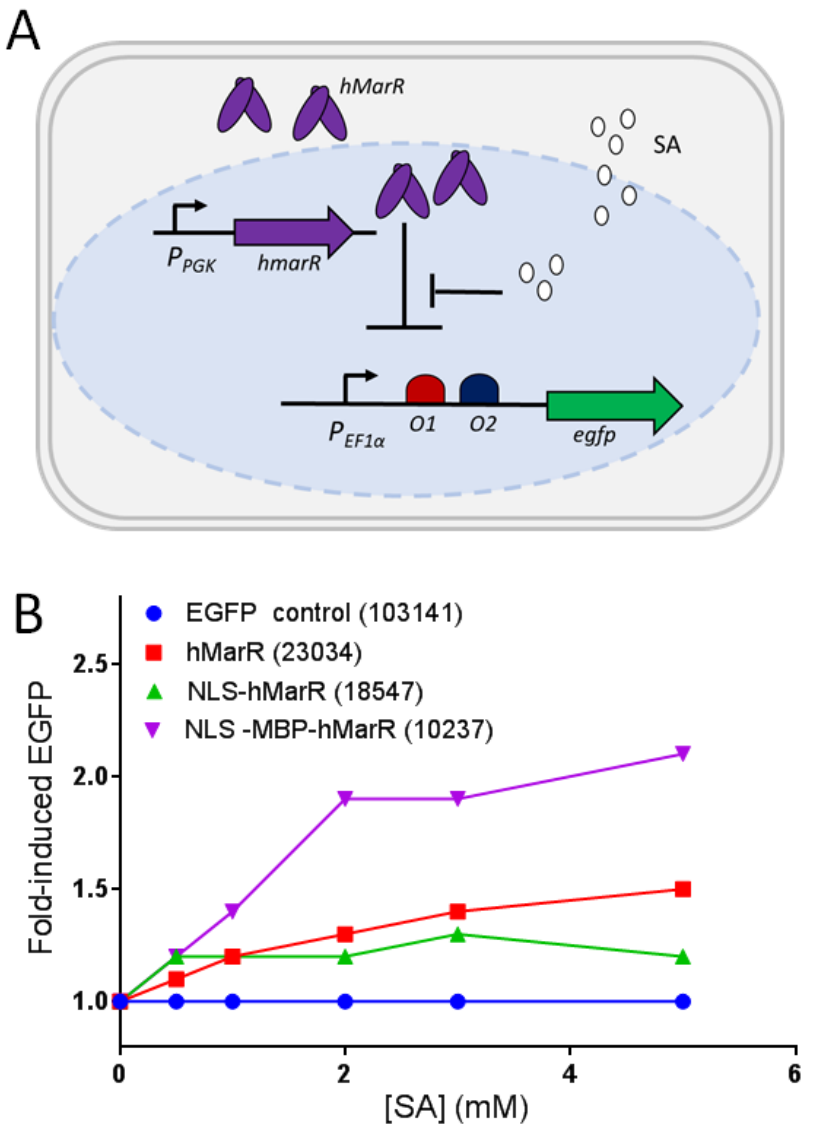

Figure 16. ***SA-inducible EGFP expression in human Jurkat cell lines****. A) Design of the* SA_{ON} *system used for SA-inducible EGFP expression in Jurkat cells. B) Fold-induced EGFP expression values in Jurkat cells incubated with increasing SA concentration for 24 hours. Background EGFP expression values, in absence of SA, is given in parentheses.*

A high background expression levels had been previously observed with the TetR-based transactivator system where Yoshida *et al*., (1997) demonstrated the use of a nuclear localization signal (NLS) to achieve tight regulation of target gene expression [107].

Nevozhay *et al*., (2013) also demonstrated a similar tight regulation with a NLS-tagged TetR repressor and reasoned that the addition of NLS increases the nuclear availability of TetR thereby increasing the effective repressor-DNA binding rate and mediating tight regulation of the target gene expression [40]. It is known that, the nuclear pore complex allows passive diffusion of proteins with a molecular mass up to 40 kDa [108]. While it is reasonable to believe that the large TetR protein (46 kDa) may not be efficiently localized in the nucleus without the NLS tag, the MarR protein used in this present study is a relatively small protein (16 kDa) and thus theoretically capable of passive diffusion across the nuclear membrane. I speculated that a probable reason for the high background EGFP expression with the MarR system, in absence of SA, may be low expression levels of the MarR repressor protein which would limit the amount of repressor available for DNA binding. High level expression of recombinant protein in mammalian cells has recently been achieved by fusion with the bacterial maltose binding protein (MBP) [109]. While the exact mechanism for the high-level expression of MBP fusion proteins in mammalian cells is not clear, it is possible that MBP fusions act indirectly by improving cell survival through increased solubility of recombinant proteins [109].

To test whether addition of NLS or a high level MarR expression can reduce the background EGFP expression I fused a simian virus 40 (SV40) NLS tag or a SV40 NLS-tagged MBP protein to the N-terminus of the codon optimized MarR (pPCC1625-1 or pPCC1625-2 respectively). While not much improvement was seen in the fold-induced EGFP expression with the NLS-hMarR as compared to the non-nuclear tagged hMarR, a slight improvement in the fold-induced EGFP expression was noted with the NLS-MBP-hMarR

(~2.1-fold increase in EGFP expression with 5 mM SA) construct (B). However, as seen in Figure 16B, the background EGFP fluorescence with the NLS-MBP-MarR fusion is still high.

Discussion

In this study I report the use of the human codon optimized, *E. coli* MarR protein to regulate target gene expression by controlling the availability of SA. However, the current design of the system shows a high background target gene expression even with no SA added to the cell media. The high background in absence of SA inducer greatly reduces the dynamic range of the SA_{ON} and limits the applicability of the system for tight and efficient gene regulation. While the use of MBP fusion protein to increase the MarR expression levels was tested to reduce the background, other modifications like use of stronger promoter for repressor expression or use of stronger-repressor domains need to be explored to improve the dynamic range and inducible response of the SA_{ON} system.

APPENDIX I

TUNING THE TRANSLATION RATE OF PLANT DERIVED ENZYMES FOR IMPROVING ORSELLINIC ACID PRODUCTION IN *E. COLI*

Introduction

Orsellinic acid (OA) is an important platform chemical used as a precursor in the synthesis of the highly potent anti-HIV agent daurichromenic acid [72, 110]. OA is the simplest tetraketide, alkylresorcylic acid and is produced by many fungi and lichens in presence of type I polyketide synthases (PKSs) [71, 111]. However, Taura *et al.,* (2016), recently reported the use of a plant type III PKS in the biosynthesis of OA [72]. Type III PKSs are structurally simple homodimeric enzymes that have been extensively studied and used in the bioproduction of important compounds like chalcones, pyrones, acridones, phloroglucinols, stilbenes, and resorcinolic lipids [112]. The simple structures of type III PKSs make their genetic manipulation easy and hence several engineering efforts have been focused on increasing the range of products that can be synthesized using the type III PKS enzymes. In this regard, Taura *et al.,* (2016), identified and demonstrated the use of the *Rhododendron dauricum* orcinol synthase enzyme (RdOrs), in the synthesis of OA *in vitro* and *in vivo* in yeast cells [72]. The enzyme Rd-ORS catalyzes the condensation of the starter acetyl-CoA molecule with two malonyl-CoA molecules to form a tetraketide intermediate which further undergoes cyclization, either spontaneously or in presence of a cyclase, to form OA and some other simple phenolic compounds [72] (Figure 17A). Since the starting compounds (acetyl-CoA and malonyl-CoA) are common *E. coli* metabolites, I sought to initially demonstrate the production of OA in *E. coli* host, by transferring the plant *RdOrs* gene into *E. coli*. To my knowledge, this is the first time OA production was sought in *E. coli*. Transformed *E. coli* BL21-DE3 cells expressing the *Rhododendron dauricum* orcinol synthase enzyme (RdOrs) and the *Cannabis sativa* olivetolic acid cyclase enzyme (CsOac), to

enhance the cyclization of the tetraketide intermediate for OA production [72]), produce OA when grown in enriched media. Further, tuning of the translation rates of the *RdOrs* and *CsOac* genes, resulted in a total of 0.09 mM of OA in cell-free supernatant, when the translation of the *RdOrs* and *CsOac* genes was driven through strong ribosomal binding sites.

Materials and Methods

DNA cloning and ribosomal binding site optimization for tuning protein expression

Three bicistronic design (bcd) ribosomal binding sites (RBS) of relative strong (bcd2), moderate (bcd11), and weak (bcd21) strengths were identified from a previous study [113] and were used to build a small library of nine combinations of the RBS sites to tune the translation rates of the *RdOrs* and *CsOac* genes. The various combinations of RBSs were generated by PCR by using RBS specific primers and the PCR fragments were assembled into the pPCC1690 backbone using Gibson assembly. For RBS characterization the strong, moderate, and weak RBSs were used to tune the translation rate of the green fluorescent protein (*gfp*) gene and relative GFP expression at varying isopropyl β-d-1-thiogalactopyranoside (IPTG) concentrations was measured using the plate-based fluorescence assay. NEBuilder® HiFi DNA Assembly Master Mix was used for Gibson Assembly. All regular PCR in this work used Phusion® High-Fidelity DNA Polymerase or Q5® High-Fidelity DNA Polymerase. PCR conditions followed NEB Tm Calculator (https://tmcalculator.neb.com/) and manufacturer's instructions for the polymerases. Zymoclean™ Gel DNA Recovery Kit was used for gel purification of DNA fragments. Translation initiation rates were calculated using RBS Calculator [114-117] (https://salislab.net/software/ .

Plate-based fluorescence assay for RBS characterization

Escherichia coli BL21(DE3) cells carrying the gene for T7 RNA polymerase under control of the lacUV5 promoter were used in this work. Colonies of fresh BL21(DE3) transformants were inoculated into 500 µL lysogeny broth (LB) medium supplemented with 100 µg/mL ampicillin. After 12-13 hours of growth at 37 °C 250 RPM, the cultures were diluted into fresh LB containing 100 µg/mL ampicillin and the IPTG inducer at varying concentrations. The subcultures were grown for 4 hours at 37 °C 900 RPM in 96-well deep well plates. The cells were washed with PBS buffer before measurements. Fluorescence was measured on SpectraMax® Gemini™ EM Microplate Spectrofluorometer from Molecular Devices®. Optical Density at 595 nm (OD_{595}) was measured on BMG Labtech NOVOstar Microplate reader. Fluorescent intensity was normalized by OD_{595}. Fold-induced GFP expression was calculated by dividing the normalized fluorescence intensity in presence of the IPTG inducer by the normalized fluorescence intensity without the inducer. All data points reported are the average values of two or more independent replicates

Culture conditions for OA production in E. coli

Codon optimized *RdOrs* and *CsOac* genes were cloned into pET vector backbone, under the P_{T7} promoter to generate recombinant plasmid pPCC1690. The cells were grown in lysogeny broth (LB) media at 37 ºC for routine culture requirements. For production of OA, fresh *E. coli* BL21(DE3) transformants carrying the recombinant plasmid pPCC1690 or the plasmids carrying the different RBS combinations for *RdOrs* and *CsOac* genes were first inoculated in 5 ml of LB medium supplemented with 100 µg/mL ampicillin. After 6-7 hours of growth at 37 °C 250 RPM, the cultures were diluted into fresh terrific broth (TB) medium (containing 1% glycerol as carbon source) with 100 µg/mL ampicillin. The subcultures were

grown in TB for 2-3 hours until OD_{595} was around 0.5, after which the cells were induced with 0.1 mM IPTG, and grown at 17 ^{o}C for 15-16 hours followed by incubation at 25 ^{o}C 250 rpm for another five hours to allow expression of the *RdOrs* and *CsOac* genes and subsequent production of OA. After five hours of growth at 25 ^{o}C, the culture was transferred into 15 ml conical tubes and centrifuged at 4 ^{o}C 4000 rpm for 50 minutes using the Sorvall™ Legend™ X1R centrifuge. The filtered supernatant was transferred to a fresh tube and used in end-point HPLC analysis for OA production along with OA standards.

HPLC analysis of amount of OA produced

The amount of extracellular OA present in the filtered supernatant obtained from culturing the BL21(DE3) transformants, carrying the *RdOrs and CsOac* genes, was measured using HPLC analysis on a Shimadzu LC-20AD HPLC system with an SPD-20A dual-wavelength UV–vis detector. A Phenomonex Luna C18 column (25 cm × 4.6 mm, 5 μm) was used for OA quantification with the following elution profile: Solvent A, 1% (v/v) acetic acid in water; solvent B, 1% (v/v) acetic acid in acetonitrile; gradient: 5% B (0-5 minutes), 5-15% B (5-18 minutes), 15-100% B (18-23 minutes), 100% B (23-30 minutes); flow rate 1.0 mL/minute; wavelength, 280 nm. The OA elution time was ~23.6 minutes, and concentrations were determined from an OA standard curve.

Results

Orsellinic acid production in E. coli by RdOrs and CsOac expression

To achieve orsellinic acid (OA) production in *E. coli*, the codon optimized orcinol synthase (*RdOrs*) and olivetolic acid cyclase (*CsOac*) genes were cloned into bacterial expression vector (pET vector) under the control of an isopropyl β-d-1-thiogalactopyranoside

(IPTG) inducible T7 promoter. The recombinant plasmid, pPCC1690 (bearing the *RdOrs* and *CsOac* genes), was used to transform *E. coli* BL21(DE3) strain. Transformed *E. coli* cells induced with IPTG were cultured under specified conditions (see Materials and Methods) post induction in terrific broth (TB) medium in a shaker flask after which the cell-free supernatant was analyzed for OA production. HPLC analysis of the cell-free supernatant revealed that cultivation of transformed *E. coli* cells under the given conditions produced orsellinic acid (Figure 17B). The results suggested that *RdOrs* and *CsOac* gene expression in *E. coli* elicits OA production.

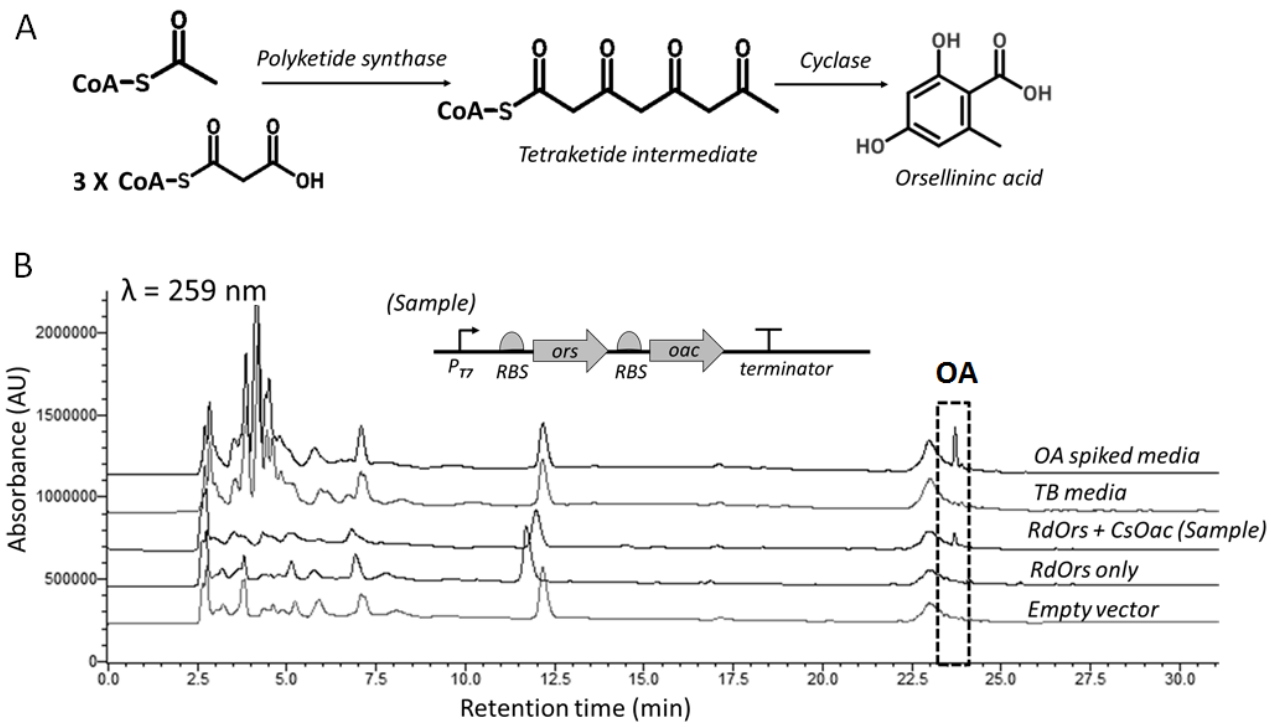

Figure 17. ***Orsellinic acid (OA) production in E. coli.*** *A) Enzyme catalyzed reactions for OA production. The plant type III polyketide synthase, orcinol synthase from Rhododendron dauricum (RdOrs gene), catalyzes synthesis of a tetraketide intermediate from acetyl coenzyme A (acetyl-CoA) and malonyl-CoA starter molecules. A cyclase, like the olivetolic acid cyclase from Cannabis sativa (CsOac gene), catalyzes the cyclization of the tetraketide intermediate to produce orsellininc acid. B) HPLC elution profile of cell-free supernatant from a culture of E. coli cells expressing the RdOrs and CsOac genes (Sample) or from cells transformed with empty vector is compared to the standard and media controls. Inset. Construct used for expressing the RdOrs and CsOac genes through promoter P_{T7} in Sample.*

Improving orsellinic acid production in E. coli by tuning RdOrs and CsOac translation rates

The efficiency of secondary metabolite production in *E. coli* is related to the enzyme expression levels and hence a common strategy to improve product yield is to overexpress the pathway enzymes [118]. However, overexpression of heterologous proteins may sometimes lead to metabolic imbalances in the host strain, causing a reduced host fitness and thereby reducing the efficiency of metabolite production [119]. Modulating the expression of pathway enzymes by promoter engineering, ribosome binding site (RBS) engineering and/or gene copy

number regulation often helps optimize protein levels to maximize product yield [119, 120]. Since the T7 promoter used for co-expression of the *RdOrs* and *CsOac* genes leads to high-level transcription and is generally recalcitrant to genetic modification [119, 121], I opted to tune the translation rates of the *RdOrs* and *CsOac* genes to maximize OA production. Three bicistronic design (bcd) RBS sequences (bcd 2, bcd 11, and bcd 21 [113]) were selected to modulate *RdOrs* and *CsOac* gene expression between relatively strong (bcd 2), moderate (bcd 11), or weak (bcd 21) translation levels (Figure 18B).

To characterize the RBS activity of the selected bcd sequences, I analyzed the green fluorescent protein (GFP) reporter expression through the three RBS constructs. The results shown in Figure 18A, indicate that the selected RBSs showed varying GFP fluorescence with the strong and moderate RBSs producing 6-fold and 3-fold more fluorescence respectively, with 0.1 mM IPTG, than the weak RBS. Further, the selected RBS reporters exhibited enhanced fluorescence in response to increasing IPTG levels, indicating that the introduced RBSs did not affect the T7 promoter functionality.

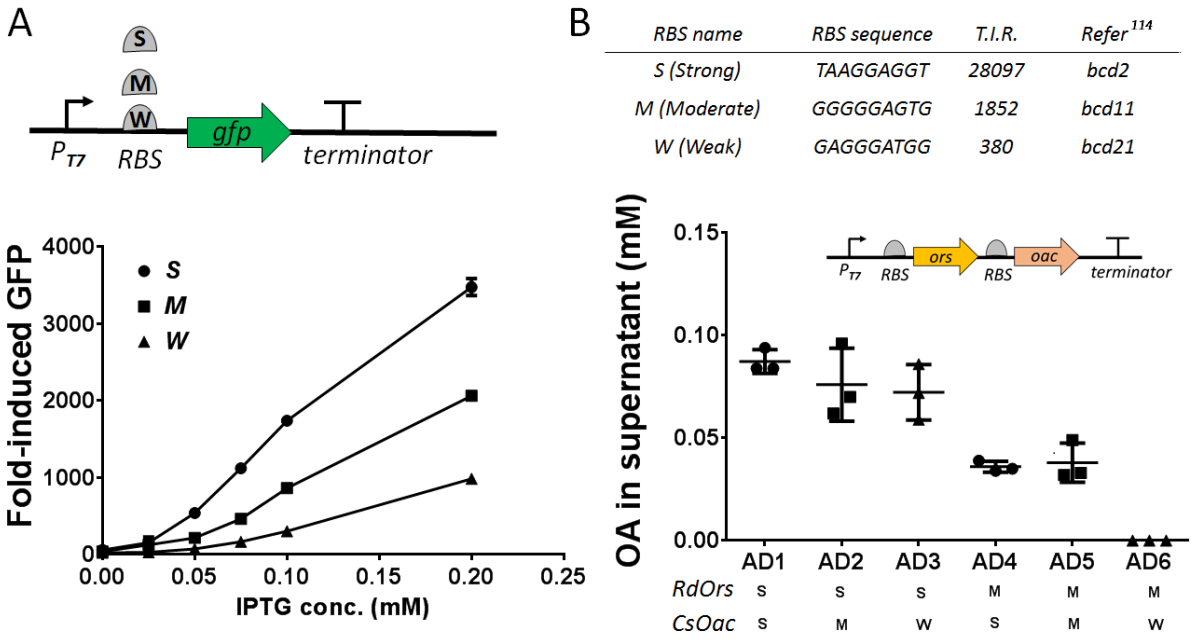

Figure 18. ***Tuning the translation rates of the RdOrs and CsOac genes to maximize OA production****. A) Tuning gfp expression through the strong (S), moderate (M), and weak (W) ribosomal binding sites (RBSs) and at varying IPTG concentrations. Fold-induced GFP, the ratio of GFP fluorescence with inducer to GFP fluorescence without the IPTG inducer, is obtained for varying IPTG concentrations and used to compare the relative strength of the used RBSs. B) Translation initiation rates (T.I.R.) calculated using RBS Calculator (https://salislab.net/software/) are shown in the table. The RdOrs and CsOac genes were expressed either under the control of strong, moderate, or weak RBS to optimize OA production in E. coli. Data points are average of three values and error bars represent the SD.*

By placing each of the *RdOrs* and *CsOac* genes under the strong, moderate, or weak RBSs, I constructed nine combinations of the two genes to identify the conditions leading to maximized OA production. As expected, the HPLC analysis of the cell-free supernatant harvested 21 hours post induction revealed that the highest OA was produced when both *RdOrs* and *CsOac* genes are expressed from the strong RBS (0.09 mM) (Figure 18B). Notably, reduced *CsOac* gene expression from a moderate or a weak RBS only slightly reduced OA production when *RdOrs* was strongly expressed while a drastic reduction in OA production was observed when *RdOrs* was expressed from moderate or weak RBS (Figure

18B). This result suggests that the *RdOrs* and *CsOac* gene products are essential for OA synthesis in *E. coli*.

Discussion

While previous studies have utilized yeast cells for producing orsellinic acid [72], to my knowledge, this is the first study that demonstrates OA production in *E. coli*. Compared to yeast cells, *E. coli* host offers the advantage of a high growth rate and ease of genetic manipulation [118]. *E. coli* cells expressing the orcinol synthase enzyme have the potential to produce orsellinic acid. However, the OA production in the *RdOrs*-expressing *E. coli* below the toxicity limit of OA (no effect on growth up to 5mM OA) a further increase in OA production in *E. coli* can be achieved by using metabolic engineering or enzyme engineering strategies. Each of these methods would generate a large, diversified library of testable variants and hence would be limited by the lack of a high-throughput screening method. Chapter 2 describes the development of an OA specific sensor which can assist screening efforts to isolate high OA producing *E. coli* strains.

APPENDIX II

CONJUGATIVE TRANSFER OF THE *E. COLI* LACI/LACO SYSTEM INTO *MARINOBACTER HYDROCARBONOCLASTICUS* SP17

Introduction

Marinobacter hydrocarbonoclasticus is a Gram-negative, aerobic, motile, non-spore forming, and rod-shaped marine bacterium which exhibits halotolerance (0.08 M to 3.5 M NaCl) [122]. The *M. hydrocarbonoclasticus* SP17 strain was originally isolated from polluted seawaters along the French Mediterranean coast, where it is known to form biofilms on oil-contaminated sediments (hydrophobic organic compounds (HOCs)) to degrade the oil (HOCs) and use it as a carbon and energy source [123]. *M. hydrocarbonoclasticus* is halotolerant and has a potential use in wastewater plants with high saline content, where *M. hydrocarbonoclasticus* acts as a denitrifier to remove nitrate and nitrite content from industrial wastes [124]. The hydrocarbonoclastic (hydrocarbon-degrading) SP17 strain may thus serve as a suitable host for developing biosensor systems to simultaneously monitor and improve seawater and wastewater quality.

As described in Chapter 1 of this thesis, biosensors are detection tools comprised of a biological recognition element, for example a bacterial transcriptional factor (TF) protein, that provides quantitative information about changes in the availability of a specific target molecule. For example the alkane responsive transcriptional activator, AlkS isolated from *Pseudomonas putida*, activates transcription from the *alkB* promoter in presence of alkanes with 6-10 carbon atoms [125] was used to develop a biosensor based on hydrocarbonoclastic bacteria *Alcanivorax borkumensis* to detect even extremely low concentrations of petrol (0.012% v/v) in seawater [126]. To develop a similar biosensing device based on SP17, a method to transfer alkane-responsive TF proteins in SP17 and a high-throughput screening platform needs to be designed.

To this end, in this study, I describe the transfer of the *E. coli* LacI/LacO transcriptional regulatory system in *M. hydrcarbonoclasticus* SP17 using conjugation. Conjugation is the unidirectional transfer of genetic material from one bacterium (donor strain) to another bacterium (recipient strain) through direct contact and is a common mode of horizontal gene transfer in Gram-negative bacteria. In this study, *E. coli* S17-1 λpir (*Escherichia coli* ATCC® 47055™) strain containing chromosomally integrated transfer gene (*tra* gene), for conjugation, was used as the donor strain to transfer a mobilizable, broad host range, pBBR1-*lacZ* plasmid into the recipient marine bacterium. Successful conjugation and recombinant protein expression led to isopropyl β-D-galactopyranoside (IPTG) inducible expression of β-galactosidase (LacZ) reporter enzyme observed as development of blue colored *Marinobacter* colonies on the induction agar plate. The results of this study demonstrate horizontal gene transfer in non-model bacteria. The ability to control gene expression in the marine bacterium will enable development of screening platform for future strain and protein engineering efforts in *M. hydrocarbonoclasticus*.

Materials and Methods

Strains and Media

M. hydrocarbonoclasticus SP17 (ATCC 49840) was received as a gift from Dr. Douglas Bartlett of University of California, San Diego. *Escherichia coli* strain ATCC® 47055™ was used as the donor strain to transfer broad host range plasmids, derived from pBBR1MCS-2 (see plasmid section), into *M. hydrocarbonoclasticus*. The *E. coli* strains MG1655, MC1061 and DH5α were used for cloning and propagating plasmid vectors. Different media formulations were used for routine culture and selection of transconjugant *M.*

hydrocarbonoclasticus cells. Halo media (Table 7) supplemented with 20 mM sodium lactate was used for routine culture of the marine bacterium at 30 ^{o}C. Transconjugant *M. hyrocarbonoclasticus* were selected on selection media supplemented with antibiotic gentamicin (50 µg/ml) (Table 7) that inhibited growth of gentamicin resistant *E. coli* donor strain. *E. coli* were grown in lysogeny broth (LB) media or Halo media at 37 ^{o}C supplemented with required antibiotics. Unless otherwise mentioned, the antibiotics for culture of *E. coli* cells were ampicillin (100 µg/ml), gentamicin (15 µg/ml).

Table 7. Composition of media used in this study

Media	**Description**
Halo media agar	Halo media broth containing 1.5% agar
Halo media broth	Peptone (5 g), yeast extract (1 g), NaCl (10 g), Na_3-citrate.$2H_2O$ (3.4 g), $MgSO_4.7H_2O$ (20 g), K_2HPO_4 (0.5 g), $Fe(NH_4)_2(SO_4)_2.6H_2O$ (0.05 g) dissolved in 1 L distilled water. Adjust pH to 7.5
LB agar	LB broth containing 1.5% agar
Lysogeny broth (LB)	Tryptone (10 g), yeast extract (5 g), NaCl (5 g) dissolved in 1 L of distilled water. Adjust pH to 7.5
Selection media agar	Prepare 1 X Solution A with 1.5% agar and then add remaining salts before pouring the agar media into plates.
Selection media broth (transconjugants)	1X Solution A (800 ml), 0.1M K2HPO4 (4 ml), 8 µM FeSO4.$7H_2O$ (2 ml), 0.1 M CaCl2.2H2O (100 ml), 0.1 M MgSO4.7H2O (100 ml), 2 M Na-DL-lactate (10 ml)
Solution A (1X)	NaCl (14.6 g), KCl (0.94 g), Tris base (7.56 g), NH_4Cl (3.75 g) dissolved in given order in 1 L distilled water. Adjust pH to 7.8

Conjugation and selection of transconjugants

Transfer of the pBBR1-*lacZ* plasmid into *M. hydrocarbonoclasticus* SP17 strain (recipient strain) was performed using conjugation as described by Mounier *et al.,* (2018). Briefly, overnight grown culture of donor and recipient bacteria were mixed in 1:1 ratio and co-cultured on a sterile filter placed on a halo media agar without any antibiotic at 30 ^{o}C for ~24 hours period. The mixed culture obtained on the filter was transferred to a 1.5 ml sterile

tube containing 1 ml of fresh selection media broth supplemented with 20 mM sodium DL-lactate. The cell suspension was centrifuged at 7000 g for five minutes to concentrate the cell pellet. The pellet was resuspended in ~100 µl of selection media and streaked or spread on the selection media agar supplemented with gentamicin (50 µg/ml) to allow growth of only transconjugant *M. hydrocarbonoclasticus*. Colonies of transconjugant marine bacteria were observed after 60 hours - 72 hours incubation at 30 ^{o}C. The presence of plasmid in *Marinobacter* was confirmed using colony PCR

Inducible gene expression

M. hydrocarbonoclasticus harboring pBBR1-*lacZ* were subcultured on induction media to demonstrate isopropyl β-D-galactopyranoside (IPTG) inducible β-galactosidase (LacZ) expression in the marine bacteria. Random colonies of the transconjugant *M. hydrocarbonoclasticus* strains were inoculated in halo media broth for overnight growth at 30^{o}C. The cell suspension from overnight growth was streaked on selection media supplemented with IPTG only (0.1 mM), selection media supplemented with X-Gal (5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside) only (40 µg/ml in dimethyl sulfoxide), or in selection media containing both IPTG and X-Gal inducer. After incubation for 48 hours – 60 hours at 30 ^{o}C the colonies were observed for development of blue color indicative of enzymatic cleavage of X-Gal by LacZ to release blue colored chromogenic compound.

Plasmids

All plasmids used in this study are listed in Table 8.

Table 8. List of all plasmids used in this study

Plasmid	Description
pBBR1MCS-2	pBBR1MCS-2 was a gift from Kenneth Peterson (Addgene plasmid # 85168 ; http://n2t.net/addgene:85168 ; RRID:Addgene_85168)
pBBR1MCS-2-Gm	pBBR1MCS-2 having kanamycin resistance coding gene swapped with a gentamicin resistance coding gene
pBBR1-*lacZ*	pBBR1MCS-2-Gm inserted with *E. coli lac* promoter controlling *lacZ* and *lac* repressor (*lacI*) genes

Cloning

All PCR reactions were carried out using Q5 polymerase and the recommended PCR conditions for Q5 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 digestion (SspI) of pBBR1MCS-2 and recovered with Zymoclean Gel Recovery Kit. Gibson Assembly was performed using Hi-Fi Assembly Master Mix, and the recommended protocol from New England Biolabs NEBuilder (http://nebuilder.neb.com/) was followed.

Results

Conjugative transfer of broad host range plasmid in M. hydrocarbonoclasticus SP17

To enable expression of *E. coli* proteins in *M. hydrocarbonoclasticus* the broad host range plasmid pBBR1MCS-2 served as a vector. The pBBR1 derived plasmids were isolated from *Bordetella bronchiseptica* and an replicate and stably maintain the genes they carry in several distant bacterial species [127]. Previous studies have reported the use of pBBR1MCS-

2 as a choice of vector to transfer desired genes in various *Marinobacter* species [128, 129] indicating that the pBBR1MCS-2 plasmid is likely to replicate in SP17. Initial attempts to transfer the pBBR1MCS-2 plasmid in SP17 through chemical and electrical transformations did not produce any transformants on selection media (data not shown). While a clear reason for failure of transformation is not currently available, a probable reason for the failure of DNA uptake through transformation may be presence of a highly impervious cell wall structure. Weakening of cell wall of the *Clostridium difficile* strain with glycine supplemented media has recently been described to increase transformation efficiency in the *C. difficile* strain [130]. Media and buffer optimizations may also be required to enable uptake of DNA in the SP17 and such optimization studies could be pursued in the future. Since another common mode of plasmid transfer in Gram negative bacteria is through conjugation [131], I decided to test the transfer the *E. coli* proteins in SP17 using conjugation. To demonstrate the usability of conjugation for horizontal gene transfer in SP17, the plasmid pBBR1MCS-2-Gm with gene coding for gentamicin (Gm) resistance was used to isolate resistant transconjugants on selection media (Figure 19). Indeed, SP17 colonies resistant to Gm were isolated after allowing overnight conjugative transfer of the plasmid between the donor and recipient strains. Control SP17 cells that did not undergo conjugation, failed to grow on the selection media indicating conjugation enables transfer of Gm resistance into SP17 cells.

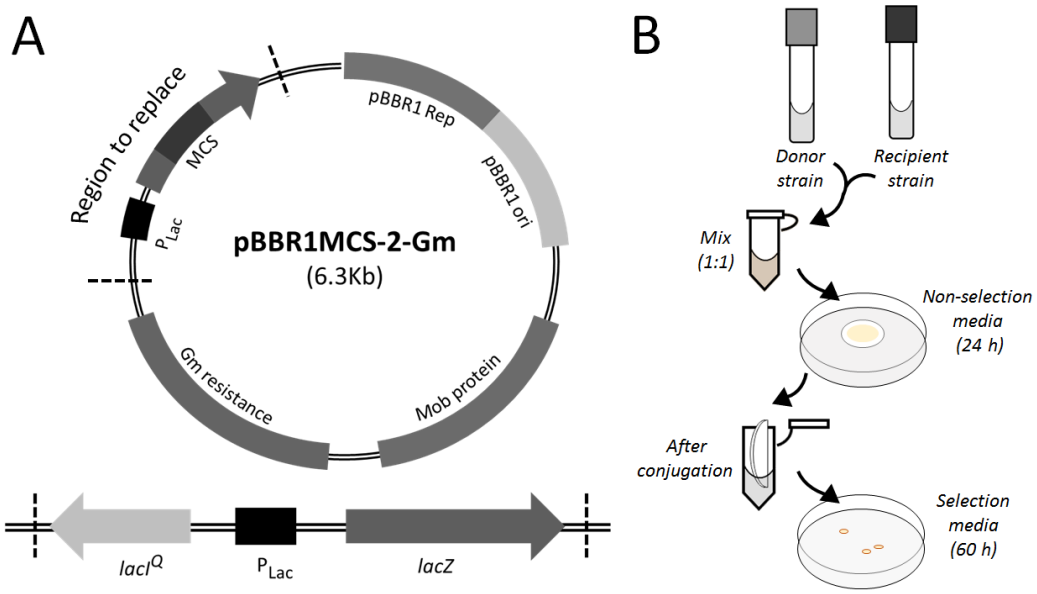

Figure 19. ***Conjugation in M. hydrocarbonoclasticus SP17****. A) The broad host range plasmid, pBBR1MCS-2-Gm, was used to transfer the E. coli LacI/LacO system in M. hydrocarbonoclasticus SP17. The lacI and lacZ genes were inserted in the region marked by dashed lines using two fragment Gibson assembly protocol. B) Schematic of the steps involved in conjugative transfer of the broad host range plasmid in M. hydrocarbonoclasticus SP17. After 60 hours of incubation, transconjugant SP17 cells were isolated on selection media containing gentamycin (50 µg/ml).*

Inducible β-galactosidase expression in M. hydrocarbonoclasticus

Inducible gene expression systems enable control over expression of target genes via tuning the availability of chemical or physical inducers. Such control over the target gene expression is highly sought for varied applications like overexpression of heterologous proteins [132], certain therapeutic tasks [133, 134], and in dynamic gene regulation through biosensor development [4]. To demonstrate the availability of *M. hydrocarbonoclasticus* SP17 as a host for inducible gene expression, I sought to transfer the *E. coli* LacI/LacO system in SP17 to regulate β-galactosidase (*lacZ* gene) expression through the lac promoter by modulating the availability of the IPTG inducer. With no IPTG in the media, the LacI repressor protein binds to the lac operator sites and inhibits downstream *lacZ* expression (Fig.

2). Media supplemented with IPTG releases the LacI inhibition and allows expression of the *lacZ* gene. SP17 transconjugants expressing LacZ enzyme produce blue colored colonies on selection media supplemented with IPTG inducer and X-Gal substrate indicating the enzymatic cleavage of the X-Gal substrate. Such inducible control over target genes in SP17, could provide means to develop biosensors to monitor alkane levels in water bodies (after accidental spills), or for high-throughput screening of efficient hydrocarbon degraders.

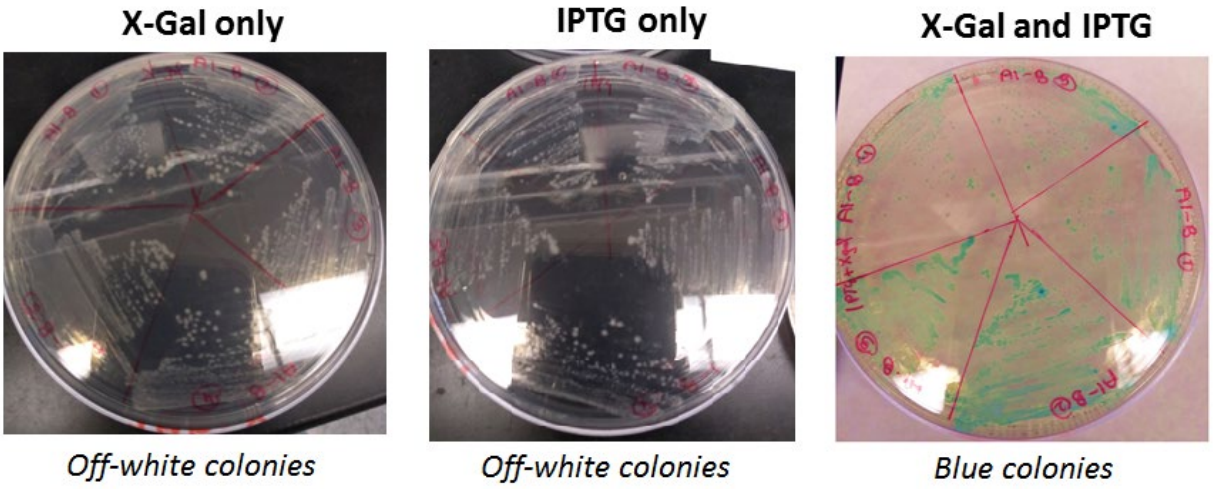

Figure 20. ***Isopropyl β-D-galactopyranoside (IPTG)-inducible β-galactosidase expression in M. hydrocarbonoclasticus SP17****. Enzymatic cleavage of the X-Gal (5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside) substrate (40 µg/ml) in selection media produces blue colored colonies of transconjugant SP17 cells only when the SP17 cells are induced with IPTG (also present in selection media). Control cells in selection media supplemented with either IPTG or X-Gal alone do not show blue colored colonies, indicating tight and inducible β-galactosidase expression.*

Plasmid pBBR1-*lacZ* carried the desired gene construct and was transferred in SP17 cells using conjugation as previously described. Transconjugant SP17 cells isolated on the selection media were re-streaked on selection media supplemented with IPTG and X-Gal and observed for development of blue colored colonies. As seen in Figure 20, blue colored SP17

colonies were observed on media containing both IPTG and X-Gal while no blue color was produced in the IPTG only and the X-Gal only controls, indicating tight and inducible control over *lacZ* expression in SP17.

Conclusion

M. hydrocarbonoclasticus SP17 develops biofilms on a variety of hydrocarbon substrates to assimilate these hydrophobic organic compounds (HOC) and utilize them as a carbon source. The assimilation and degradation of HOCs has important implications on the carbon fluxes in oceanic ecosystems and hence there is much interest in improving the efficiency of these HOC degraders. The aim of this work was to demonstrate the ability to precisely control target gene expression in SP17 using the *E. coli* LacI/LacO system with the hope to use this IPTG inducible LacI/LacO system as a basis for developing a biosensor based screening platform in the marine bacterium. Such a screening platform will enable dynamic pathway regulation and high-throughput screening of engineered protein and strain libraries for isolating efficient hydrocarbon degraders.

Broad host plasmids were used as vectors to transfer the *E. coli* proteins into SP17 using bacterial conjugation. Under the given experimental conditions, tight and inducible target gene expression was observed by varying the availability of IPTG inducer. Future experiments can be designed to engineer hydrocarbon inducibility in the LacI repressor protein or may include the use of existing hydrocarbon responsive transcriptional regulators to develop efficient HOC monitoring or degrading systems.

BIBLIOGRAPHY

[1] Thevenot, D. R., Toth, K., Durst, R. A., Wilson, G. S., Electrochemical biosensors: recommended definitions and classification. *Biosensors & Bioelectronics* 2001, *16*, 121-131.

[2] Mehrotra, P., Biosensors and their applications - A review. *Journal of Oral Biology and Craniofacial Research* 2016, *6*, 153-159.

[3] Hicks, M., Bachmann, T. T., Wang, B., Synthetic biology enables programmable cell-based biosensors. *Chemphyschem : A European Journal of Chemical Physics and Physical Chemistry* 2020, *21*, 132-144.

[4] Mahr, R., Frunzke, J., Transcription factor-based biosensors in biotechnology: current state and future prospects. *Applied Microbiology and Biotechnology* 2016, *100*, 79-90.

[5] Chu, L. L., Dhakal, D., Shin, H. J., Jung, H. J.*, et al.*, Metabolic engineering of *Escherichia coli* for enhanced production of naringenin-7-sulfate and its biological activities. *Frontiers in Microbiology* 2018, *9*, 1671.

[6] Martinelli, L., Nikel, P. I., Breaking the state-of-the-art in the chemical industry with new-to-Nature products via synthetic microbiology. *Microbial Biotechnology* 2019, *12*, 187-190.

[7] Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T.*, et al.*, A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology* 2019, *10*, 1404.

[8] Liu, C., Zhang, B., Liu, Y. M., Yang, K. Q., Liu, S. J., New intracellular shikimic acid biosensor for monitoring shikimate synthesis in *Corynebacterium glutamicum*. *ACS Synthetic Biology* 2018, *7*, 591-601.

[9] Li, H., Chen, W., Jin, R., Jin, J. M., Tang, S. Y., Biosensor-aided high-throughput screening of hyper-producing cells for malonyl-CoA-derived products. *Microbial Cell Factories* 2017, *16*, 187.

[10] Dietrich, J. A., Shis, D. L., Alikhani, A., Keasling, J. D., Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synthetic Biology* 2013, *2*, 47-58.

[11] Liu, S. D., Wu, Y. N., Wang, T. M., Zhang, C., Xing, X. H., Maltose utilization as a novel selection strategy for continuous evolution of microbes with enhanced metabolite production. *ACS Synthetic Biology* 2017, *6*, 2326-2338.

[12] Qian, S., Li, Y., Cirino, P. C., Biosensor-guided improvements in salicylate production by recombinant *Escherichia coli*. *Microbial Cell Factories* 2019, *18*, 18.

[13] Wang, Y., Li, Q., Zheng, P., Guo, Y.*, et al.*, Evolving the L-lysine high-producing strain of Escherichia coli using a newly developed high-throughput screening method. *Journal of Industrial Microbiology & Biotechnology* 2016, *43*, 1227-1235.

[14] Raghavan, S. S., Chee, S., Li, J., Poschmann, J.*, et al.*, Development and application of a transcriptional sensor for detection of heterologous acrylic acid production in *E. coli*. *Microbial Cell Factories* 2019, *18*, 139.

[15] van Rossum, T., Muras, A., Baur, M. J. J., Creutzburg, S. C. A.*, et al.*, A growth- and bioluminescence-based bioreporter for the in vivo detection of novel biocatalysts. *Microbial Biotechnology* 2017, *10*, 625-641.

[16] Hanko, E. K. R., Minton, N. P., Malys, N., A transcription factor-based biosensor for detection of itaconic acid. *ACS Synthetic Biology* 2018, *7*, 1436-1446.

[17] Tang, S. Y., Cirino, P. C., Design and application of a mevalonate-responsive regulatory protein. *Angewandte Chemie* 2011, *50*, 1084-1086.

[18] Tang, S. Y., Fazelinia, H., Cirino, P. C., AraC regulatory protein mutants with altered effector specificity. *Journal of the American Chemical Society* 2008, *130*, 5267-5271.

[19] Tang, S. Y., Qian, S., Akinterinwa, O., Frei, C. S.*, et al.*, Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter. *J Am Chem Soc* 2013, *135*, 10099-10103.

[20] L, F. M. M., Currin, A., Dixon, N., Directed evolution of the PcaV allosteric transcription factor to generate a biosensor for aromatic aldehydes. *Journal of Biological Engineering* 2019, *13*, 91.

[21] Collins, C. H., Arnold, F. H., Leadbetter, J. R., Directed evolution of *Vibrio fischeri* LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Molecular Microbiology* 2005, *55*, 712-723.

[22] Collins, C. H., Leadbetter, J. R., Arnold, F. H., Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat Biotechnol* 2006, *24*, 708-712.

[23] Hawkins, A. C., Arnold, F. H., Stuermer, R., Hauer, B., Leadbetter, J. R., Directed evolution of *Vibrio fischeri* LuxR for improved response to butanoyl-homoserine lactone. *Applied and Environmental Microbiology* 2007, *73*, 5775-5781.

[24] De Paepe, B., Peters, G., Coussement, P., Maertens, J., De Mey, M., Tailor-made transcriptional biosensors for optimizing microbial cell factories. *Journal of Industrial Microbiology & Biotechnology* 2017, *44*, 623-645.

[25] Mannan, A. A., Liu, D., Zhang, F., Oyarzun, D. A., Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synthetic Biology* 2017, *6*, 1851-1859.

[26] Rogers, J. K., Church, G. M., Genetically encoded sensors enable real-time observation of metabolite production. *Proceedings of the National Academy of Sciences of the United States of America* 2016, *113*, 2388-2393.

[27] Morales, M. A., Halpern, J. M., Guide to selecting a biorecognition element for biosensors. *Bioconjugate Chemistry* 2018, *29*, 3231-3239.

[28] Mohrle, V., Stadler, M., Eberz, G., Biosensor-guided screening for macrolides. *Analytical and Bioanalytical Chemistry* 2007, *388*, 1117-1125.

[29] Kasey, C. M., Zerrad, M., Li, Y., Cropp, T. A., Williams, G. J., Development of transcription factor-based designer macrolide biosensors for metabolic engineering and synthetic biology. *ACS Synthetic Biology* 2018, *7*, 227-239.

[30] Pittard, J., Camakaris, H., Yang, J., The TyrR regulon. *Molecular Microbiology* 2005, *55*, 16-26.

[31] Liu, Y., Zhuang, Y., Ding, D., Xu, Y.*, et al.*, Biosensor-based evolution and elucidation of a biosynthetic pathway in *Escherichia coli*. *ACS Synthetic Biology* 2017, *6*, 837-848.

[32] Juarez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D., Church, G. M., Biosensor libraries harness large classes of binding domains for construction of allosteric transcriptional regulators. *Nature Communications* 2018, *9*, 3101.

[33] Das, A. T., Tenenbaum, L., Berkhout, B., Tet-On systems for doxycycline-inducible gene expression. *Current Gene Therapy* 2016, *16*, 156-167.

[34] Auslander, S., Fussenegger, M., Engineering gene circuits for mammalian cell-based applications. *Cold Spring Harbor Perspectives in Biology* 2016, *8*.

[35] Gossen, M., Bujard, H., Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* 1992, *89*, 5547-5551.

[36] Neddermann, P., Gargioli, C., Muraglia, E., Sambucini, S.*, et al.*, A novel, inducible, eukaryotic gene expression system based on the quorum-sensing transcription factor TraR. *EMBO Reports* 2003, *4*, 159-165.

[37] Wang, H., Ye, H., Xie, M., Daoud El-Baba, M., Fussenegger, M., Cosmetics-triggered percutaneous remote control of transgene expression in mice. *Nucleic Acids Research* 2015, *43*, e91.

[38] Mullick, A., Xu, Y., Warren, R., Koutroumanis, M.*, et al.*, The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnology* 2006, *6*, 43.

[39] Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., Balazsi, G., Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 2009, *106*, 5123-5128.

[40] Nevozhay, D., Zal, T., Balazsi, G., Transferring a synthetic gene circuit from yeast to mammalian cells. *Nature Communications* 2013, *4*, 1451.

[41] Zhou, Z., Dang, Y., Zhou, M., Li, L.*, et al.*, Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences of the United States of America* 2016, *113*, E6117-E6125.

[42] Elena, C., Ravasi, P., Castelli, M. E., Peiru, S., Menzella, H. G., Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Frontiers in Microbiology* 2014, *5*, 21.

[43] Kozak, M., Initiation of translation in prokaryotes and eukaryotes. *Gene* 1999, *234*, 187-208.

[44] Wan, X., Marsafari, M., Xu, P., Engineering metabolite-responsive transcriptional factors to sense small molecules in eukaryotes: current state and perspectives. *Microbial Cell Factories* 2019, *18*, 61.

[45] Nott, A., Meislin, S. H., Moore, M. J., A quantitative analysis of intron effects on mammalian gene expression. *RNA* 2003, *9*, 607-617.

[46] Hodel, M. R., Corbett, A. H., Hodel, A. E., Dissection of a nuclear localization signal. *The Journal of Biological Chemistry* 2001, *276*, 1317-1325.

[47] Choi, S., Song, C. W., Shin, J. H., Lee, S. Y., Biorefineries for the production of top building block chemicals and their derivatives. *Metabolic Engineering* 2015, *28*, 223-239.

[48] Kubota, T., Tanaka, Y., Takemoto, N., Hiraga, K.*, et al.*, Identification and expression analysis of a gene encoding a shikimate transporter of Corynebacterium glutamicum. *Microbiology* 2015, *161*, 254-263.

[49] Gu, X., He, D., Li, C., Wang, H., Yang, G., Development of inducible CD19-CAR T cells with a Tet-On system for controlled activity and enhanced clinical safety. *International Journal of Molecular Sciences* 2018, *19*.

[50] Eaton, R. W., p-Cymene catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA encoding conversion of p-cymene to p-cumate. *Journal of Bacteriology* 1997, *179*, 3171-3180.

[51] Poulain, A., Perret, S., Malenfant, F., Mullick, A.*, et al.*, Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch. *Journal of Biotechnology* 2017, *255*, 16-27.

[52] Liu, Y., Liu, Y., Wang, M., Design, optimization and application of small molecule biosensor in metabolic engineering. *Frontiers in Microbiology* 2017, *8*, 2012.

[53] Lee, J. H., Wendisch, V. F., Biotechnological production of aromatic compounds of the extended shikimate pathway from renewable biomass. *Journal of Biotechnology* 2017, *257*, 211-221.

[54] Kallunki, T., Barisic, M., Jaattela, M., Liu, B., How to choose the right inducible gene expression system for mammalian studies? *Cells* 2019, *8*.

[55] Stanton, B. C., Siciliano, V., Ghodasara, A., Wroblewska, L*., et al.*, Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synth Biol* 2014, *3*, 880-891.

[56] Rogers, J. K., Taylor, N. D., Church, G. M., Biosensor-based engineering of biosynthetic pathways. *Curr Opin Biotechnol* 2016, *42*, 84-91.

[57] Dietrich, J. A., McKee, A. E., Keasling, J. D., High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem* 2010, *79*, 563-590.

[58] Zhang, F., Keasling, J., Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol* 2011, *19*, 323-329.

[59] Wang, Z., Cirino, P. C., New and improved tools and methods for enhanced biosynthesis of natural products in microorganisms. *Curr Opin Biotechnol* 2016, *42*, 159-168.

[60] Reed, B., Blazeck, J., Alper, H., Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *J Biotechnol* 2012, *158*, 75-79.

[61] Chen, W., Zhang, S., Jiang, P., Yao, J*., et al.*, Design of an ectoine-responsive AraC mutant and its application in metabolic engineering of ectoine biosynthesis. *Metab Eng* 2015, *30*, 149-155.

[62] Lönneborg, R., Varga, E., Brzezinski, P., Directed evolution of the transcriptional regulator DntR: isolation of mutants with improved DNT-response. *PLoS One* 2012, *7*, e29994.

[63] Taylor, N. D., Garruss, A. S., Moretti, R., Chan, S.*, et al.*, Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods* 2016, *13*, 177-183.

[64] de los Santos, E. L., Meyerowitz, J. T., Mayo, S. L., Murray, R. M., Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: developing a vanillin sensor. *ACS Synth Biol* 2016, *5*, 287-295.

[65] Loew, R., Heinz, N., Hampf, M., Bujard, H., Gossen, M., Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol* 2010, *10*, 81.

[66] Lachmann, N., Brennig, S., Hillje, R., Schermeier, H.*, et al.*, Tightly regulated 'all-in-one' lentiviral vectors for protection of human hematopoietic cells from anticancer chemotherapy. *Gene Ther* 2015, *22*, 883-892.

[67] Li, Y., Qian, S., Dunn, R., Cirino, P. C., Engineering *Escherichia coli* to increase triacetic acid lactone (TAL) production using an optimized TAL sensor-reporter system. *J Ind Microbiol Biotechnol* 2018, *45*, 789-793.

[68] Yeom, S. J., Kim, M., Kwon, K. K., Fu, Y.*, et al.*, A synthetic microbial biosensor for high-throughput screening of lactam biocatalysts. *Nat Commun* 2018, *9*, 5053.

[69] Thompson, M. G., Pearson, A. N., Barajas, J. F., Cruz-Morales, P.*, et al.*, Identification, characterization, and application of a highly sensitive lactam biosensor from. *ACS Synth Biol* 2020, *9*, 53-62.

[70] Wang, Z., Doshi, A., Chowdhury, R., Wang, Y.*, et al.*, In Preparation.

[71] Sanchez, J. F., Chiang, Y. M., Szewczyk, E., Davidson, A. D.*, et al.*, Molecular genetic analysis of the orsellinic acid/F9775 gene cluster of Aspergillus nidulans. *Mol Biosyst* 2010, *6*, 587-593.

[72] Taura, F., Iijima, M., Yamanaka, E., Takahashi, H.*, et al.*, A novel class of plant type III polyketide synthase involved in orsellinic acid biosynthesis from *Rhododendron dauricum*. *Front Plant Sci* 2016, *7*, 1452.

[73] Austin, M. B., Noel, J. P., The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* 2003, *20*, 79-110.

[74] Frei, C. S., Wang, Z., Qian, S., Deutsch, S.*, et al.*, Analysis of amino acid substitutions in AraC variants that respond to triacetic acid lactone. *Protein Sci* 2016, *25*, 804-814.

[75] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C.*, et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009, *6*, 343-345.

[76] Pantazes, R. J., Grisewood, M. J., Li, T., Gifford, N. P., Maranas, C. D., The iterative protein redesign and optimization (IPRO) suite of programs. *J Comput Chem* 2015, *36*, 251-263.

[77] Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., Wolberger, C., Structural basis for ligand-regulated oligomerization of AraC. *Science* 1997, *276*, 421-425.

[78] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L.*, et al.*, CHARMM: the biomolecular simulation program. *J Comput Chem* 2009, *30*, 1545-1614.

[79] Reed, W. L., Schleif, R. F., Hemiplegic mutations in AraC protein. *J Mol Biol* 1999, *294*, 417-425.

[80] Saviola, B., Seabold, R., Schleif, R. F., Arm-domain interactions in AraC. *J Mol Biol* 1998, *278*, 539-548.

[81] Damjanovic, A., Miller, B. T., Schleif, R., Understanding the basis of a class of paradoxical mutations in AraC through simulations. *Proteins* 2013, *81*, 490-498.

[82] Knopf, F., Schnabel, K., Haase, C., Pfeifer, K.*, et al.*, Dually inducible TetON systems for tissue-specific conditional gene expression in zebrafish. *Proc Natl Acad Sci U S A* 2010, *107*, 19933-19938.

[83] Yin, J., Yang, L., Mou, L., Dong, K.*, et al.*, A green tea-triggered genetic control system for treating diabetes in mice and monkeys. *Sci Transl Med* 2019, *11*.

[84] van der Sluis, T. C., Sluijter, M., van Duikeren, S., West, B. L.*, et al.*, Therapeutic peptide vaccine-induced CD8 T cells strongly modulate intratumoral macrophages required for tumor regression. *Cancer Immunol Res* 2015, *3*, 1042-1051.

[85] Belteki, G., Haigh, J., Kabacs, N., Haigh, K.*, et al.*, Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res* 2005, *33*, e51.

[86] Yamada, M., Suzuki, Y., Nagasaki, S. C., Okuno, H., Imayoshi, I., Light control of the Tet gene expression system in mammalian cells. *Cell Rep* 2018, *25*, 487-500.e486.

[87] Bojar, D., Scheller, L., Hamri, G. C., Xie, M., Fussenegger, M., Caffeine-inducible gene switches controlling experimental diabetes. *Nat Commun* 2018, *9*, 2318.

[88] Gitzinger, M., Kemmer, C., Fluri, D. A., El-Baba, M. D.*, et al.*, The food additive vanillic acid controls transgene expression in mammalian cells and mice. *Nucleic Acids Res* 2012, *40*, e37.

[89] Michaels, Y. S., Barnkob, M. B., Barbosa, H., Baeumler, T. A.*, et al.*, Precise tuning of gene expression levels in mammalian cells. *Nat Commun* 2019, *10*, 818.

[90] Kuzin, I. I., Snyder, J. E., Ugine, G. D., Wu, D.*, et al.*, Tetracyclines inhibit activated B cell function. *Int Immunol* 2001, *13*, 921-931.

[91] Ahler, E., Sullivan, W. J., Cass, A., Braas, D.*, et al.*, Doxycycline alters metabolism and proliferation of human cell lines. *PLoS One* 2013, *8*, e64561.

[92] Malakar, S., Bioactive food chemicals and gastrointestinal symptoms: a focus of salicylates. *J Gastroenterol Hepatol* 2017, *32 Suppl 1*, 73-77.

[93] Gaziano, J. M., Greenland, P., When should aspirin be used for prevention of cardiovascular events? *JAMA* 2014, *312*, 2503-2504.

[94] Chi, N. F., Wen, C. P., Liu, C. H., Li, J. Y.*, et al.*, Comparison between aspirin and clopidogrel in secondary stroke prevention based on real-world data. *J Am Heart Assoc* 2018, *7*, e009856.

[95] Pagidipati, N. J., Navar, A. M., Pieper, K. S., Green, J. B.*, et al.*, Secondary prevention of cardiovascular disease in patients with type 2 diabetes mellitus: international insights rrom the TECOS trial (trial evaluating cardiovascular outcomes with sitagliptin). *Circulation* 2017, *136*, 1193-1203.

[96] Patrono, C., Baigent, C., Role of aspirin in primary prevention of cardiovascular disease. *Nat Rev Cardiol* 2019, *16*, 675-686.

[97] Fanaroff, A. C., Roe, M. T., Contemporary reflections on the safety of long-term aspirin treatment for the secondary prevention of cardiovascular disease. *Drug Saf* 2016, *39*, 715-727.

[98] Sehgal, M., Wood, S. K., Ouslander, J. G., Hennekens, C. H., Aspirin in older adults: need for wider utilization in secondary prevention and individual clinical judgments in primary prevention. *J Cardiovasc Pharmacol Ther* 2017, *22*, 511-513.

[99] Duval, V., McMurry, L. M., Foster, K., Head, J. F., Levy, S. B., Mutational analysis of the multiple-antibiotic resistance regulator MarR reveals a ligand binding pocket at the interface between the dimerization and DNA binding domains. *J Bacteriol* 2013, *195*, 3341-3351.

[100] Nevozhay, D., Zal, T., Balázsi, G., Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat Commun* 2013, *4*, 1451.

[101] Seoane, A. S., Levy, S. B., Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in Escherichia coli. *J Bacteriol* 1995, *177*, 3414-3419.

[102] Bokhove, M., Sadat Al Hosseini, H., Saito, T., Dioguardi, E.*, et al.*, Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. *J Struct Biol* 2016, *194*, 1-7.

[103] Baba, T., Ara, T., Hasegawa, M., Takai, Y.*, et al.*, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006, *2*, 2006.0008.

[104] Notka, F., Linde, H. J., Dankesreiter, A., Niller, H. H., Lehn, N., A C-terminal 18 amino acid deletion in MarR in a clinical isolate of *Escherichia coli* reduces MarR binding properties and increases the MIC of ciprofloxacin. *J Antimicrob Chemother* 2002, *49*, 41-47.

[105] Martin, R. G., Rosner, J. L., Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proc Natl Acad Sci U S A* 1995, *92*, 5456-5460.

[106] Murphy, K. F., Balázsi, G., Collins, J. J., Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* 2007, *104*, 12726-12731.

[107] Yoshida, Y., Hamada, H., Adenovirus-mediated inducible gene expression through tetracycline-controllable transactivator with nuclear localization signal. *Biochem Biophys Res Commun* 1997, *230*, 426-430.

[108] Freitas, N., Cunha, C., Mechanisms and signals for the nuclear import of proteins. *Curr Genomics* 2009, *10*, 550-557.

[109] Reuten, R., Nikodemus, D., Oliveira, M. B., Patel, T. R.*, et al.*, Maltose-binding protein (MBP), a secretion-enhancing tag for mammalian protein expression systems. *PLoS One* 2016, *11*, e0152386.

[110] Kang, Y., Mei, Y., Du, Y., Jin, Z., Total synthesis of the highly potent anti-HIV natural product daurichromenic acid along with its two chromane derivatives, rhododaurichromanic acids A and B. *Org Lett* 2003, *5*, 4481-4484.

[111] Schroeckh, V., Scherlach, K., Nützmann, H. W., Shelest, E.*, et al.*, Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci U S A* 2009, *106*, 14558-14563.

[112] Yu, D., Xu, F., Zeng, J., Zhan, J., Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* 2012, *64*, 285-295.

[113] Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C.*, et al.*, Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 2013, *10*, 354-360.

[114] Espah Borujeni, A., Channarasappa, A. S., Salis, H. M., Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* 2014, *42*, 2646-2659.

[115] Espah Borujeni, A., Salis, H. M., Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *J Am Chem Soc* 2016, *138*, 7016-7023.

[116] Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A.*, et al.*, Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res* 2017, *45*, 5437-5448.

[117] Salis, H. M., Mirsky, E. A., Voigt, C. A., Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 2009, *27*, 946-950.

[118] Okamoto, S., Chin, T., Hiratsuka, K., Aso, Y.*, et al.*, Production of itaconic acid using metabolically engineered *Escherichia coli*. *J Gen Appl Microbiol* 2014, *60*, 191-197.

[119] Xu, P., Gu, Q., Wang, W., Wong, L.*, et al.*, Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. *Nat Commun* 2013, *4*, 1409.

[120] Li, C., Swofford, C. A., Sinskey, A. J., Modular engineering for microbial production of carotenoids. *Metab Eng Commun* 2020, *10*, e00118.

[121] Studier, F. W., Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 1986, *189*, 113-130.

[122] Grimaud, R., Ghiglione, J. F., Cagnon, C., Lauga, B.*, et al.*, Genome sequence of the marine bacterium Marinobacter hydrocarbonoclasticus SP17, which forms biofilms on hydrophobic organic compounds. *Journal of Bacteriology* 2012, *194*, 3539-3540.

[123] Oude Vrielink, A. S., Vance, T. D., de Jong, A. M., Davies, P. L., Voets, I. K., Unusually high mechanical stability of bacterial adhesin extender domains having calcium clamps. *PloS One* 2017, *12*, e0174682.

[124] Carreira, C., Mestre, O., Nunes, R. F., Moura, I., Pauleta, S. R., Genomic organization, gene expression and activity profile of Marinobacter hydrocarbonoclasticus denitrification enzymes. *PeerJ* 2018, *6*, e5603.

[125] Grund, A., Shapiro, J., Fennewald, M., Bacha, P.*, et al.*, Regulation of alkane oxidation in Pseudomonas putida. *Journal of Bacteriology* 1975, *123*, 546-556.

[126] Sevilla, E., Yuste, L., Rojo, F., Marine hydrocarbonoclastic bacteria as whole-cell biosensors for n-alkanes. *Microbial Biotechnology* 2015, *8*, 693-706.

[127] Jain, A., Srivastava, P., Broad host range plasmids. *FEMS Microbiology Letters* 2013, *348*, 87-96.

[128] Bonis, B. M., Gralnick, J. A., Marinobacter subterrani, a genetically tractable neutrophilic Fe(II)-oxidizing strain isolated from the Soudan iron mine. *Frontiers in Microbiology* 2015, *6*, 719.

[129] Clark, I. C., Melnyk, R. A., Youngblut, M. D., Carlson, H. K.*, et al.*, Synthetic and evolutionary construction of a chlorate-reducing *Shewanella oneidensis* MR-1. *MBio* 2015, *6*, e00282-00215.

[130] Bhattacharjee, D., Sorg, J. A., Factors and conditions that impact electroporation of *Clostridioides difficile* strains. *MSphere* 2020, *5*.

[131] Grohmann, E., Muth, G., Espinosa, M., Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews : MMBR* 2003, *67*, 277-301, table of contents.

[132] Brautaset, T., Lale, R., Valla, S., Positively regulated bacterial expression systems. *Microbial Biotechnology* 2009, *2*, 15-30.

[133] Hwang, I. Y., Koh, E., Wong, A., March, J. C.*, et al.*, Engineered probiotic *Escherichia coli* can eliminate and prevent *Pseudomonas aeruginosa* gut infection in animal models. *Nature Communications* 2017, *8*, 15028.

[134] Lim, D., Song, M., Development of bacteria as diagnostics and therapeutics by genetic engineering. *Journal of Microbiology* 2019, *57*, 637-643.