Robust characterization of two distinct glutarate sensing transcription factors found in *Pseudomonas putida* L-lysine metabolism

Mitchell G. Thompson[1,2,3], Zak Costello[1,2,4], Niklas F. C. Hummel[1,7], Pablo Cruz-Morales[1,2,5], Jacquelyn M. Blake-Hedges[1,2,6],Rohith N. Krishna[1,2,6], Will Skyrud[6], Matthew R. Incha[1,2,3],Patrick M. Shih[1,7], Hector Garcia-Martin[1,2,4,8], Jay D. Keasling[1,2,9,10,11,12]

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA.

[2]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[3]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[4]DOE Agile BioFoundry, Emeryville, CA, USA

[5]Centro de Biotecnologia FEMSA, Instituto Tecnologico y de Estudios superiores de Monterrey, Mexico

[6]Department of Chemistry, University of California, Berkeley, CA 94720, USA

[7]Department of Plant Biology, University of California, Davis, Davis, CA, United States

[8]BCAM, Basque Center for Applied Mathematics, Bilbao, Spain

[9]Joint Program in Bioengineering, University of California, Berkeley/San Francisco, CA 94720, USA

[10]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

[11]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark

[12]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes for Advanced Technologies, Shenzhen, China

ABSTRACT

A significant bottleneck in synthetic biology involves screening the large genetically encoded libraries enabled by the latest technologies for desirable phenotypes such as chemical production. However, transcription factor based biosensors can be leveraged to screen thousands of genetic designs for optimal chemical production in engineered microbes. In this study we characterize two glutarate sensing transcription factors (CsiR and GcdR) from *Pseudomonas putida*. The genomic contexts of *csiR* homologs were analyzed and DNA binding sites were bioinformatically predicted. Both CsiR and GcdR were purified and shown to bind upstream of their coding sequencing *in vitro*. CsiR was shown to dissociate from DNA *in vitro* when exogenous glutarate was added confirming it acts as a genetic repressor. Both transcription factors and cognate promoters were then cloned into broad host range vectors to create two glutarate biosensors. Their respective sensing performance features were characterized, and more sensitive derivatives of the GcdR biosensor were created by manipulating the expression of the transcription factor. Sensor vectors were then reintroduced into *P. putida* and were evaluated for their ability to respond to glutarate as well as various lysine metabolites. Additionally, we developed a novel mathematical approach to describe the usable range of detection for genetically encoded biosensors which may be broadly useful to future efforts to better characterize biosensor performance.

INTRODUCTION

A rate limiting step in the design-build-test-learn cycle is often the test phase, wherein hundreds or thousands of genetic designs need to be evaluated for their productivity [1,2]. Though recent advances in analytical chemistry have dramatically increased sample throughput [3], transcription factor-based biosensors still offer multiple advantages over traditional chromatographic and mass-spectrometry based detection methods. One of the most attractive benefits is the ability to rapidly screen constructs for the production of the target compounds via either plate-based, or flow-cytometry-based assays [1–3], which increase throughput by orders of magnitude compared to mass-spectrometry based methods. Additionally, biosensors may offer unmatched sensitivity towards specific ligands, with some sensors having picomolar affinity [3]. The evolution of diverse microbial metabolism has provided researchers with a wide array of ligands capable of being sensed, ranging from complex natural products [4,5], to small central metabolites [1,6].

Diacids, polyamine, and lactams are all used to produce various polyester and nylon fibers which are currently derived from petrochemicals [7,8]. In an effort to make production of these chemicals sustainable, many groups have developed engineered microbes to synthesize these precursors [9–12]. The L-lysine metabolism of *Pseudomonas putida* has been leveraged to produce valerolactam [13], as well as the diacid glutarate both in the native host and heterologously [14,15]. Recently this utility has inspired much work to uncover missing steps in the lysine catabolism of *P. putida*. These missing steps included the discovery of an additional route of glutarate catabolism through a coA-independent route to succinate, in addition to the known coA-dependent route to acetyl-coA [14,16] (Figure 1). Recent work has also demonstrated that both

glutarate catabolic pathways are highly upregulated in the presence of glutarate [16]. While work in *Pseudomonas aeruginosa* has characterized a homolog of the regulator of the ketogenic pathway (GcdR) [17], and work in *Escherichia coli* has preliminarily characterized the glucogenic pathway regulator (CsiR), no work has directly investigated the regulation of the *P. putida* L-lysine catabolic pathways.

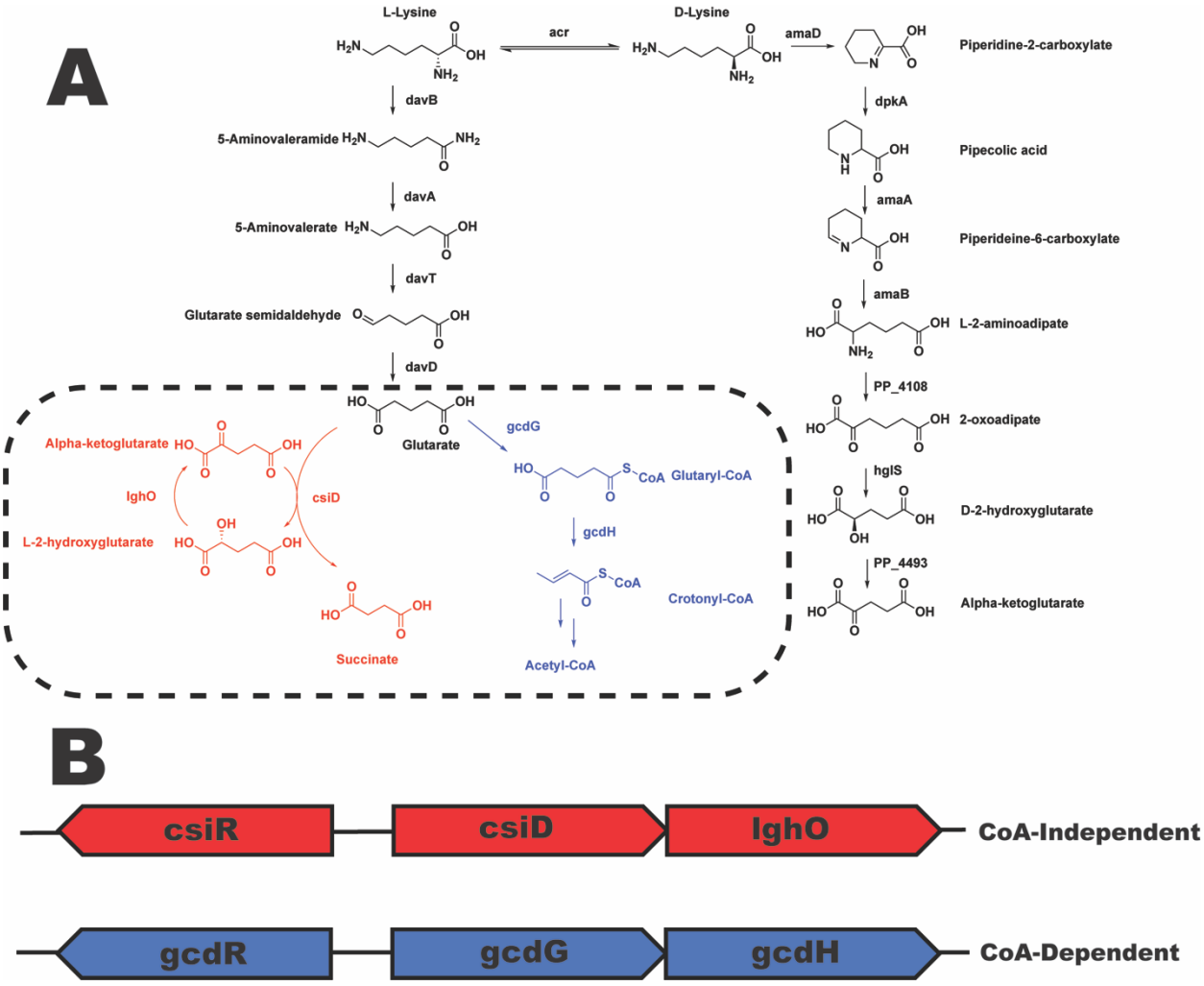

Figure 1: A) The known lysine catabolism of *P. putida*. Dashed box shows the two known pathways of glutarate catabolism in *P. putida*. Highlighted in red is the CoA-independent

route of glutarate catabolism, in blue the CoA-dependent route. B) Operonic structure of the two routes of glutarate metabolism in *P. putida*.

In this work we sought to better characterize the two putative local regulators of glutarate catabolism in *P. putida*, *csiR* and *gcdR*. First, we compared the genomic context of *csiR* homologs across bacteria to bioinformatically predict a conserved DNA binding site. We then biochemically and genetically characterized both regulators. Secondly, we developed a novel mathematical approach to rigorously determine the detection ranges for genetically encoded biosensors in a way that can be systematically compared to other biosensors. Finally, we introduced RFP transcriptional-fusions of the promoter for both catabolic pathways into *P. putida* and evaluated their induction upon the addition of various lysine metabolites.

RESULTS

Genomic contexts of *csiR* and *gcdR* homologs and prediction of *P. putida* binding sites

Work in *Pseudomonas aeruginosa* has characterized the GcdR regulation of ketogenic glutarate metabolism, and shown that the binding site is conserved across multiple bacterial species [17]. While binding sites for CsiR in *E. coli* have been identified, whether there is a conserved binding site for homologs across bacterial species has yet to be investigated [18]. In order to identify conserved binding sites of CsiR homologs, we compared the syntenic genomic contexts of 12 selected genomes that contained neighboring *csiD* and *csiR* homologs. Genes encoding *csiR* were found in two distinct genomic contexts, either transcribed divergently from *csiD* as found in *P. putida*, or transcribed as the last gene in the *csi* operon as in *E. coli* (Figure 2). The genomic regions upstream of the *csiD* homolog were extracted and Multiple EM for Motif Elicitation (MEME) was used to identify a conserved CsiR binding motif [19]. A consensus A(A/G)AAATCTAGA(C/T)ATTTT motif was identified upstream of each *csiD* homolog.

Previously, footprinting assays in *E. coli* BW25113 revealed two primary and two secondary bindings sites with the sequences TTGTN$_5$TTTT and ATGTN$_5$TTTT respectively [18]. Our consensus motif agrees closely with the secondary binding site identified, and highly suggests that the binding site of CsiR is conserved across the bacteria where homologs are present.

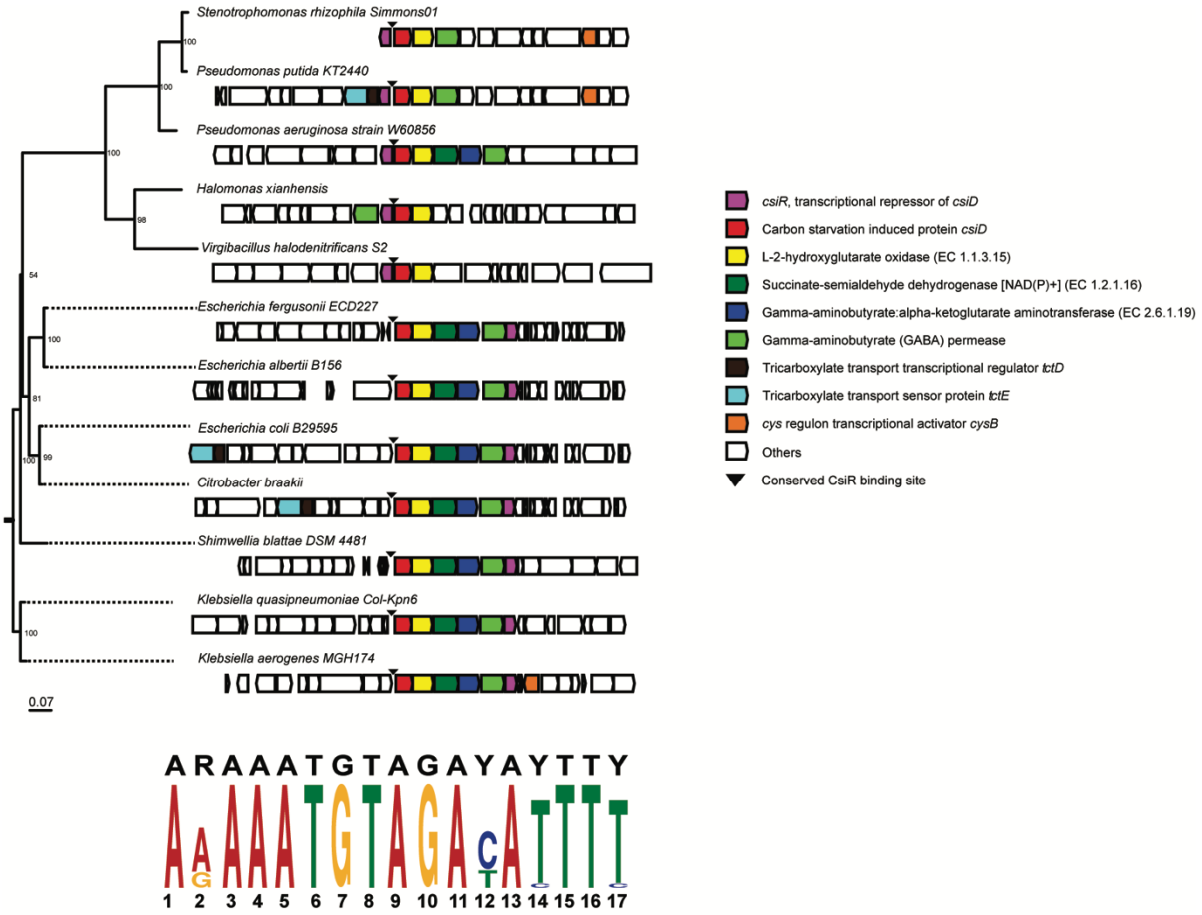

Figure 2: Genomic contexts of *csiR* homologs and predicted binding regions: Phylogeny of assorted gammaproteobacteria and the location of their *csiD* operons. MEME analysis of the intergenic regions upstream of the *csiD* operon resulted in the sequence motifs depicted below the tree.

<u>Biochemical characterization of CsiR and GcdR</u>

To determine if the *P. putida* CsiR also acts as a regulator and to identify its putative binding sites, we biochemically characterized this protein using electrophoretic mobility shift assays (EMSAs). The CsiR protein was purified and incubated with DNA probes consisting of the intergenic region between *csiR* and *csiD*. The assay showed multiple binding sites in this intergenic region, as four distinct bands appeared. These results appear to confirm the finding in *E. coli* where four binding sites of the *E. coli* CsiR homolog were observed [18]. CsiR had a high affinity for the DNA probe, with a calculated K_d of approximately 30 nM (Figure 3a), which is similar to the 10 nM CsiR/DNA K_d of the *E. coli* CsiR homolog [18]. GcdR also bound to its cognate probe, with a single distinct shift, with an estimated K_d of approximately 62.5 nM (Figure 3a) [17]. As CsiR is a GntR family transcriptional regulator, many of which act as repressors [20], we evaluated whether glutarate would decrease the DNA binding affinity of CsiR. EMSA assays were repeated in the presence of increasing glutarate concentrations. Analysis by gel electrophoresis revealed increasing quantities of free probe as glutarate concentrations increased (Figure 3b). These results confirm that the *P. putida* CsiR is a glutarate-responsive repressor of the *csiDlhgO* operon [18].

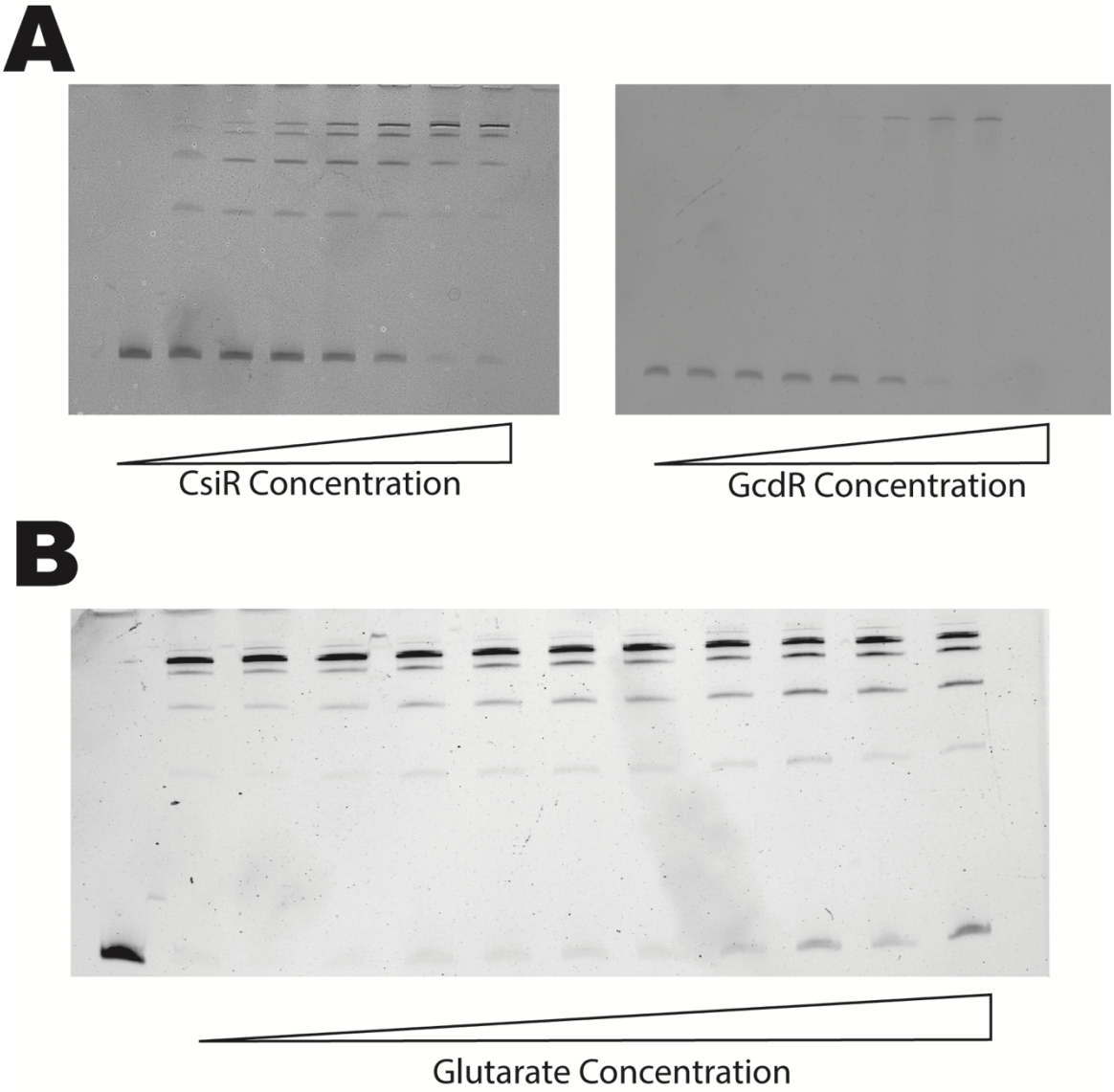

Figure 3: Biochemical characterization of CsiR and GcdR. (A) EMSA of either purified CsiR (left) or GcdR (right) incubated with fluorescent probe of intergenic region between *csiR* and *csiD*. Concentrations of CsiR protein range from 0 nM to 250 nM. (B) EMSA showing 100 nM CsiR and fluorescent probe incubated with glutarate from concentrations of 0 mM to 5 mM. Far left lane shows free probe with no CsiR.

Development of two glutarate biosensor vectors

In order to evaluate both CsiR and GcdR as biosensors we cloned both the regulator and intergenic region of both the coA-independent and coA-dependent glutarate catabolism pathways upstream of RFP on a broad host range vector pBADT. *E. coli* DH10B harboring either vector were grown for 24 hours in LB medium supplemented with concentrations of glutarate ranging from 2.5 mM to 0.00015 mM, after which both OD600 and RFP fluorescence were measured. The GcdR and CsiR vectors demonstrated a maximal induction over background at 2.5 mM glutarate of 2.07 and 2.98 times over uninduced cells (Figure 4A). Normalized RFP expression for each sensor was fitted to the Hill equation to derive biosensor performance. The GcdR system was found to have a Hill coefficient of 0.64, a K_d of 1.01 mM, and a maximum normalized RFP expression of 9511, while the CsiR system was shown to have a Hill coefficient of 0.81, a K_d of 0.39 mM, and a maximum normalized RFP expression of 9377 (Figure 4A). To test for the ability of CsiR or GcdR to sense other diacids, *E. coli* harboring either vector were grown in LB medium with 2.55 mM to 0.00015 mM of either succinate, adipate, or pimelate. While no induction was seen by any of these diacids in the CsiR system, the GcdR system was induced ~1.5x over background by 2.5 mM adipate (Figure S1).

To test whether the performance characteristics of these GcdR system could be altered, the 50 bp upstream sequence of *gcdR* in the GcdR-sensing vector was replaced by three different previously characterized constitutive promoters from the Anderson collection (J23101, J23110, and J23113), representing a high (1791 RFP AU), medium (844 RFP AU) , and low (21 RFP AU) strengths [21]. When tested against glutarate all of the engineered GcdR showed dramatically increased sensitivity to glutarate compared to when GcdR is under native regulation with each new vector showing a decreased K_d (J23113 : 0.07 mM, J23110 : 0.15 mM, J23101 : 0.04 mM) (Figure 4B). The predicted maximum RFP output of each engineered GcdR system appeared to

be inversely proportional to the predicted strength of the constitutive promoter with the strongest promoter, J23113, showing the lowest max RFP at 15,481 and the weakest predicted promoter, J23113 showing the highest max RFP at 24,190. However, all engineered GcdR vectors showed at least twice the maximal expression compared to the native system (Figure 4B). The relative induction of the J23113 engineered vector remained similar to that of the native system, while the J23110 and J23101 variants showed slightly greater induction at 2.5 mM glutarate of 3.17x and 2.65x respectively (Figure 4B).

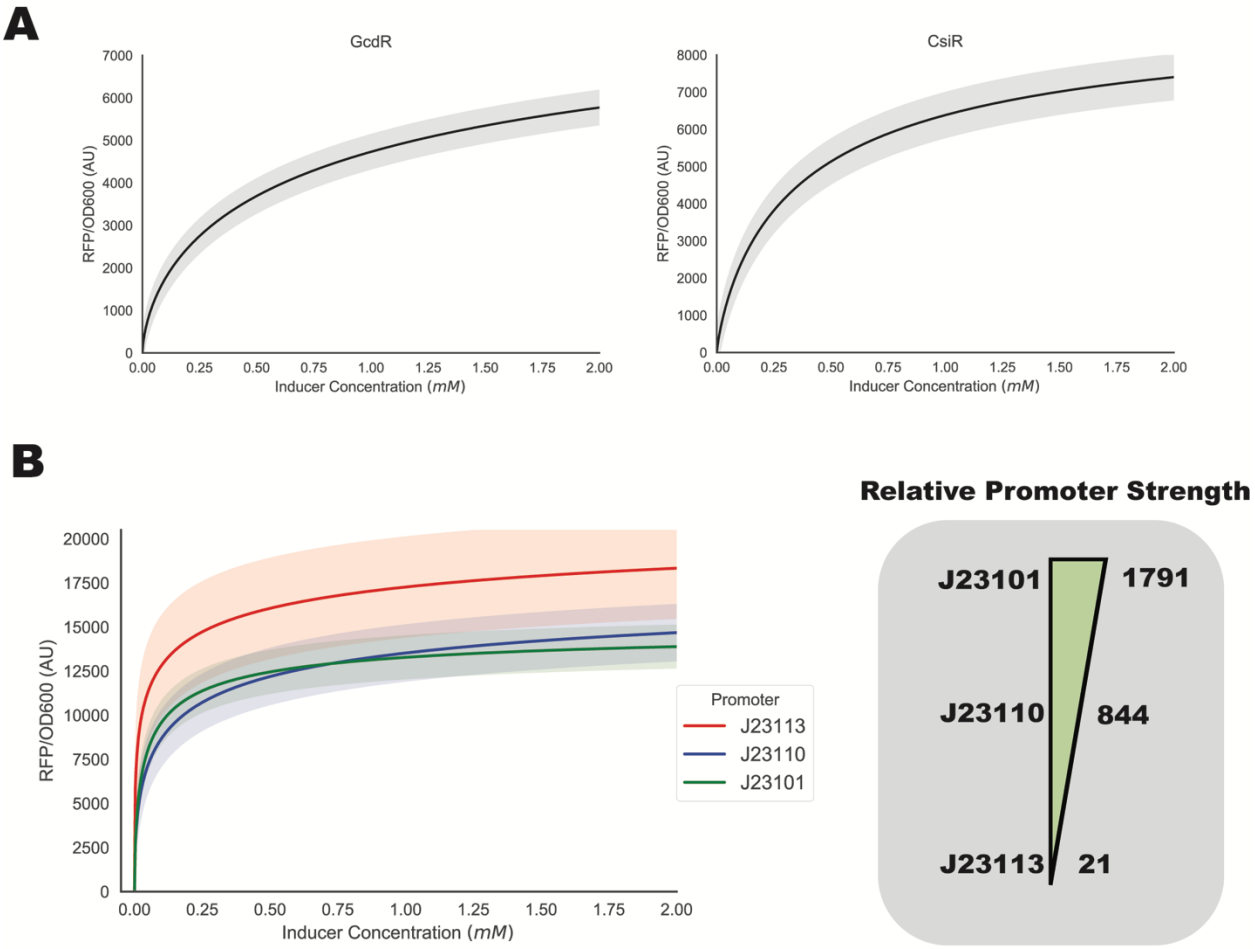

Figure 4: Development of glutarate responsive reporter vectors for *E. coli*. (A) Fluorescence data fit to the Hill equation to derive biosensor performance characteristics

for native CsiR and GcdR systems. Shaded area represents (+/-) one standard deviation, n=4. (B) Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for the engineered GcdR systems. Relative promoter strength is shown to the right. Shaded area represents (+/-) one standard deviation, n=4.

Table 1: Best fits of biosensor performance parameters

Reporter	**Hill Coef.**	**K_d (mM)**	**Max**	**Induction**
CsiR-Native	0.81	0.39	9377	2.98
GcdR-Native	0.64	1.01	9511	2.07
GcdR-J23113	0.36	0.07	24190	1.91
GcdR-J23110	0.47	0.15	19012	3.17
GcdR-J23101	0.59	0.04	15481	2.65

Developing metrics to quantify biosensor performance

In previous published work the mathematical basis for determining linear range or limits of detection has often been obscure and non-systematic [1,6,22]. Furthermore, a linear range may not be able to adequately capture the performance characteristics of a biosensor over the range of ligand concentrations where the sensor can still resolve differences. To address these deficiencies we sought to develop a mathematical method for evaluating the sensing performances of genetically encoded transcription factors fitted to the Hill equation. Our approach uses a

probabilistic model to relate inducer concentration and corresponding fluorescence measurements fit to the Hill equation assuming: (1) fluorescence measurements at a particular concentration are normally distributed, (2) the variance of fluorescence measurements is roughly constant over the range of measured values, and (3) the relationship between ligand concentrations and fluorescence can be well modeled using the Hill function. This model allows us to estimate the concentration of ligand compatible with our observed fluorescence data given the variance of the data as determined via Markov Chain Monte Carlo (MCMC) sampling [23]. A detailed methodological description can be found in the supplemental Jupyter notebook, which can also be used to analyze any other biosensor data.

By applying MCMC sampling to the model of our native GcdR biosensor we can readily produce the probability density functions (i.e. the probability that the ligand produces the observed fluorescent response) of specific ligand concentrations (Figure 5A). At glutarate concentrations of 0.25 mM, 0.68 mM, and 1.125 mM associated fluorescence values can be resolved from one another (no overlap), however the biosensor is less able to resolve ligand concentrations between 1.5 mM and 2 mM (Figure 5A). When we apply MCMC sampling to the GcdR sensor being driven by the J23101 promoter, we observe that this system possesses the resolution to distinguish between 0.003 mM and 0.0165 mM but is less able to distinguish between higher concentrations (Figure 5B). Ranges of detection can also be calculated using an MCMC based approach. The lower limit of detection is calculated by finding the smallest inducer concentration that can be statistically distinguished from a zero fluorescence measurement (Table 2). The upper limit of detection is determined by finding the largest inducer concentration that can be statistically distinguished from maximum tested inducer concentration (Table 2). A biosensor's resolution window, defined as the width of the 95% prediction interval

of inducer concentrations derived from a set of fluorescence measurements, can then be expressed as a continuous function across a range of ligand concentrations for a given biosensor (Figure 5C). Below concentrations of ~0.07 mM glutarate the J23101 GcdR biosensor has greater resolution, while at higher concentrations the native GcdR sensor system has greater resolution, as seen inFigure 5C. Another important aspect of our approach is that it allows for the resolution window to be calculated as a function of the number of replicates in a biosensor experiment. If either variance decreases or sample size increases, the resolution of a biosensor also increases. By simulating sample sizes of 1 through 100 via MCMC sampling the theoretical resolution of the native GcdR dramatically increases (Figure 4D). This “power” analysis may serve as a guide for experimental design when a certain biosensor resolution is required for a given application.

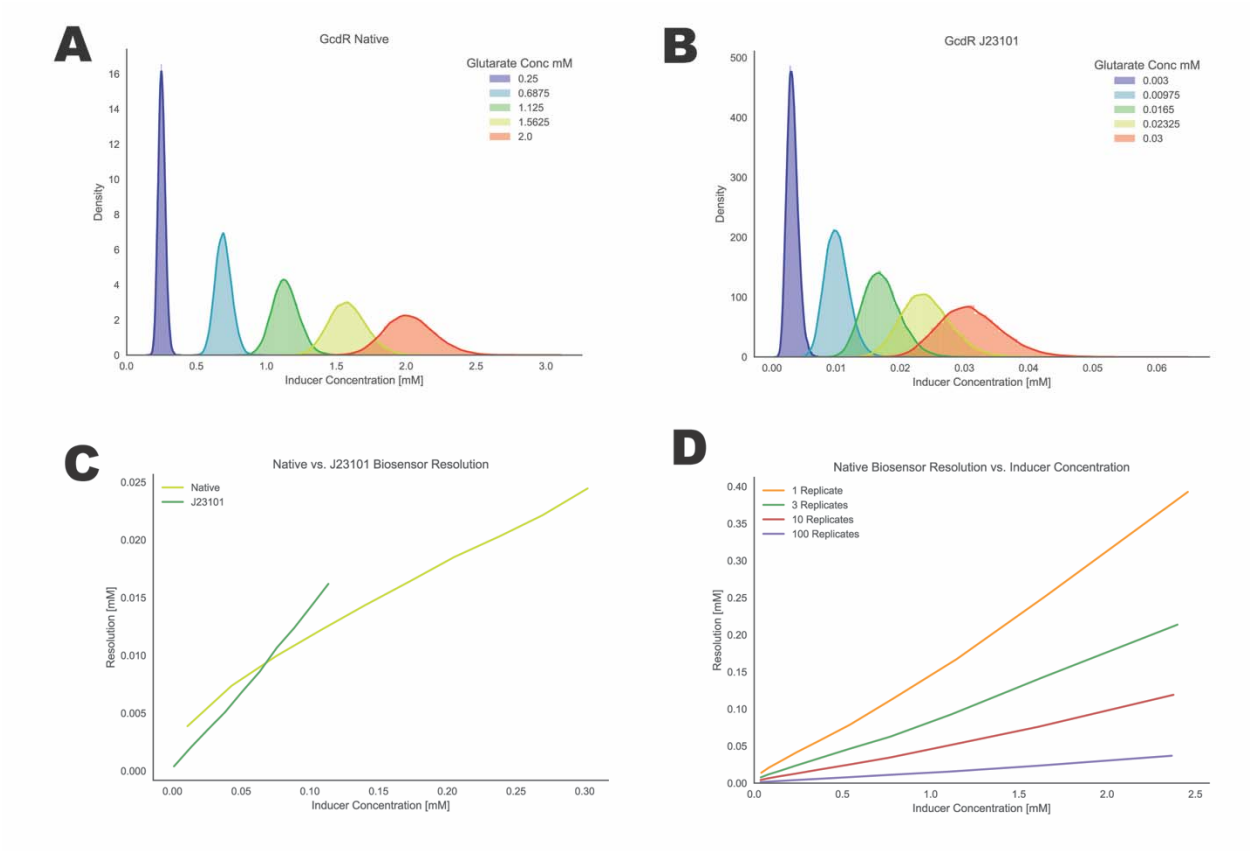

Figure 5: Development of analytics for biosensor performance. Probability density functions of fluorescence values compatible with concentrations of glutarate for selected ligand concentrations modelled to the Hill equation of the native GcdR system (A) or the GcdR J23101 system (B). (C) Resolution of native GcdR or GcdR J23101 biosensor systems over a select range of glutarate concentrations. (D) Theoretical resolution of the GcdR native biosensor with differing number of replicates.

Table 2 - Range of detection of glutarate biosensors

Reporter	Min (mM)	Max (mM)

CsiR-Native	0.031	2.74
GcdR-Native	0.018	2.19
GcdR-J23113	0.003	0.95
GcdR-J23110	0.003	1.16
GcdR-J23101	0.002	1.09

Responsiveness of Glutarate Biosensors to Lysine metabolites in *P. putida*

To assess the ability of these vectors to function in *P. putida* both plasmids were introduced into either wild type *P. putida* or a strain with both known pathways of glutarate catabolism deleted (Δ*csiD*Δ*gcdH* - referred to as Δglutarate). The resulting strains were grown in MOPS minimal medium supplemented with 10 mM glucose and glutarate ranging from 5 mM to 0.01 mM for 24 hours. Both vectors responded to increased concentrations of exogenously applied glutarate, though the GcdR vector had ~10x greater fluorescence than the CsiR vector (Figure 6A). In both CsiR and GcdR systems the Δglutarate strain showed increased levels of RFP induction, suggesting both vectors are able to sense increased levels of glutarate (Figure 6A). To further examine the ability of these vectors to probe *P. putida* lysine metabolism, both were also introduced into a Δ*davT* strain, which is unable to metabolize 5-aminovalerate to glutarate semidaldehyde (Figure 1A) and therefore precludes glutarate production. When wild type, Δ*davT*, and Δglutarate strains harboring either the CsiR or GcdR systems were grown on minimal medium supplemented with 10 mM glucose and 10 mM 5-aminovalerate both vectors in the Δ*davT* strain showed decreased fluorescence compared to wild type, while vectors in the

Δglutarate strains showed increased fluorescence output (Figure 6B). Measurement of intracellular 5-aminovalerate showed significant pools of the metabolite in the Δ*davT* strains (~1500 uM/OD600) with no detectable 5-aminovalerate in the other genetic backgrounds (Figure 6C). These results highly suggest that both GcdR and CsiR are insensitive to 5-aminovalerate, an essential feature of these sensors if they are to be used in organisms that derive glutarate from a 5-aminovalerate precursor.

To evaluate the ability of both reporter vectors to monitor the catabolism of other lysine metabolites, wild-type *P. putida* harboring either GcdR or CsiR were grown on 10 mM glucose, L-lysine, D-lysine, 5-aminovalerate, or 2-aminoadipate for 48 hours with OD_{600} and RFP fluorescence being measured continuously (Figure S2A and Figure S2B). Neither vector was induced when the bacterium was grown on glucose (Figure S2). The GcdR vector was strongly induced by 5-aminovalerate, and to a lesser extent 2-aminoadipate, D-lysine, and L-lysine (Figure S2A). Conversely, the strain harboring the CsiR vector only displayed induction of RFP above background when grown on 5-aminovalerate (Figure S2B). While 5-aminovalerate was able to induce RFP induction in the strain harboring the CsiR vector, induction was extremely limited compared to induction of RFP from the GcdR vector (Figure S2).

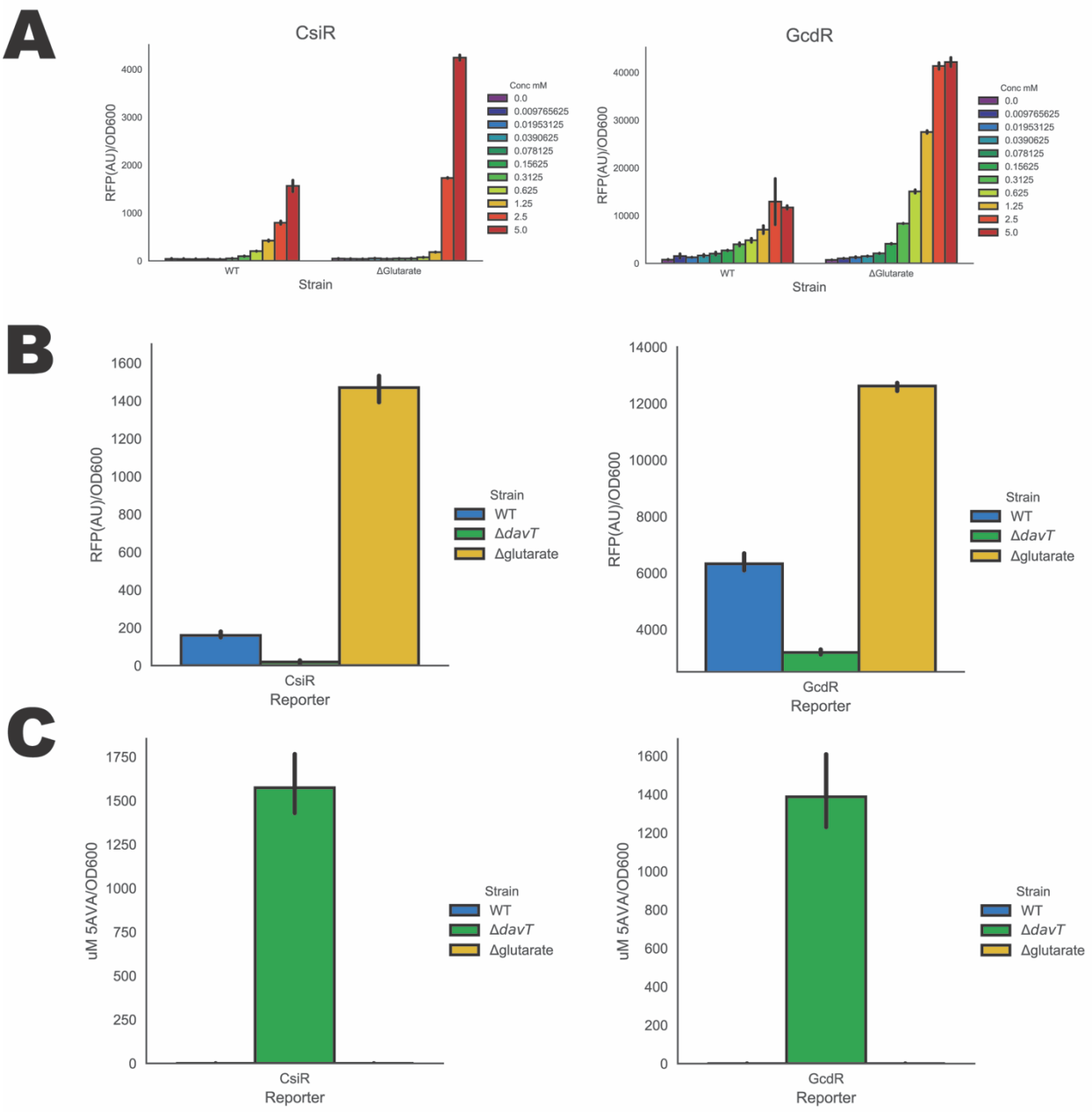

Figure 6: Performance of CsiR and GcdR biosensors in *P. putida*. (A) RFP expression of either wild type *P. putida* or *P. putida* with the ability to metabolize glutarate knocked out measured with either the CsiR or GcdR biosensor under different external glutarate concentrations. Error bars represent 95% CI, n=4 (B) RFP expression of wild type, Δ*davT*, or Δglutarate strains of *P. putida* harboring either the CsiR or GcdR biosensor when grown on 10 mM glucose and 10 mM 5-aminovalerate (5AVA). Error bars represent 95% CI,

n=3. (C) Intracellular concentration of 5-aminovalerate of wild type, Δ*davT*, or Δglutarate strains of *P. putida* harboring either the CsiR or GcdR biosensor when grown on 10 mM glucose and 10 mM 5-aminovalerate (5AVA). Error bars represent 95% CI, n=3.

DISCUSSION

Recent advances in high-throughput functional genomics have allowed researchers to rapidly identify novel metabolic pathways, and in turn infer the function of novel transcription factors [24,25]. Rigorous characterization of these regulators is a critical step to developing novel parts for synthetic biology as well as useful tools for metabolic engineering. The discovery of two distinct pathways for glutarate catabolism within *P. putida*, regulated by independent transcription factors, presents an interesting opportunity to compare and contrast the relative sensing properties of each system.

The *P. aeruginosa* GcdR was demonstrated to be a positive regulator of the *gcdHG* operon, with a binding site that was conserved in *P. putida* KT2440 among other *Pseudomonas* strains [17]. While other strains of *P. aeruginosa* have *csiD* homologs, the type strain PA01 does not. However, *csiDlghO* pathways have been rigorously characterized in both *P. putida* and *E. coli*, though only in *E. coli* has the CsiR homolog been characterized [14,18]. Our bioinformatic approach to predict the conserved binding site of CsiR largely agrees with the empirically determined binding site identified in *E. coli* (Figure 2), and our EMSA results also confirm the existence of four distinct CsiR binding sites (Figure 3). This is not surprising given the high level of conservation found in the DNA binding sites of bacterial transcription factor homologs [26]. The dissociation of CsiR from the DNA probe with the addition of glutarate *in vitro* verifies that CsiR is a negative regulator of CsiD, which is expected as many GntR-type regulator tend to act as repressors. The existence of primary and secondary series of binding sites suggests a canonical

autoregulatory negative feedback loop, found in an estimated ~60-70% of bacterial transcription factors [27]. Further work will need to be done to characterize any such autoregulation.

Both the GcdR and CsiR biosensor vectors were able to sensitively detect glutarate with no engineering of the native regulation of the transcription factor (Figure 4A). By engineering the promoter of the *gcdR* transcription factor the sensitivity of the vectors was dramatically increased, reducing the K_d of the most sensitive vector to 0.04 mM (Figure 4B). This sensitivity to glutarate may make this vector useful in prototyping novel routes to biological production of the C5 diacid. While our preliminary work with exogenously applied ligand is promising, further work remains to be done to evaluate the ability of these sensors to detect flux of glutarate in living cells. Multiple recent publications have shown that glutarate metabolism is widespread in bacteria, and evaluating the ability of the CsiR and GcdR sensors to measure this flux will require careful experimentation with C^{13} labelled substrates. Another possible confounding factor in utilizing these vectors is the presence of CsiR or GcdR binding sites in the host organism. The overexpression of either transcription factor may misregulate host metabolism. This hypothesis is supported by the differences in growth observed between GcdR and CsiR biosensor containing *P. putida* when grown on lysine metabolites (Figure S2).

The overarching goal of synthetic biology is to apply engineering principles to biological systems so that outcomes of genetic manipulation can become more predictable and repeatable [28,29]. While there has been a large body of work devoted to the characterization and application of biosensors, there has been conspicuously few attempts to rigorously describe at which ligand concentration a biosensor is useful. Often the analysis merely states limit of detection, and a ‘linear range’ with little in terms of a mathematical justification. Here we present an alternative metric that allows for the calculation of ligand resolution across the entire range of detection for

a biosensor. By leveraging a MCMC approach to predicting ligand ranges compatible with fluorescence values, researchers can more precisely describe a biosensors performance and identify whether a given biosensing system is potentially useful for a given engineering task. The MCMC approach also allows for the simulation of an increasing number replicates which may inform the researcher the number of samples may be required in an experimental design to achieve a desired level of resolution. We hope that this initial work to better characterize biosensor performance inspires other groups to develop even more sophisticated methods of analysis.

In addition to their utility as biosensors for metabolic engineering, these sensors may be a valuable tool in studying the carbon utilization in the native host *P. putida*. Work conducted here demonstrates the ability of both CsiR and GcdR sensors to distinguish between glutarate accumulating and mutants blocked in their ability to metabolize 5-aminovalerate to glutarate (Figure 6). Lysine metabolism in *P. putida* is isomer specific, with each isomer being degraded by a separate catabolic pathway [16]. While cross-feeding between the pathways has been proposed previously [30], recent work by our group has proposed a molecular mechanism for metabolite exchange between the D- and L- catabolic pathways [16]. The exchange relies on the 2-oxoacid promiscuity of the non-heme Fe(II) oxidase CsiD which normally catalyzes the hydroxylation of glutarate using 2-ketoglutarate as a cosubstrate to yield 2-hydroxyglutarate and succinate. CsiD can also use 2-oxoadipate, a D-lysine catabolic intermediate, as a 2-oxoacid cosubstrate to yield 2-hydroxyglutarate and glutarate as products. The glutarate from this reaction could then proceed down the L-lysine catabolic pathway. When 2-aminoadipate was fed into *P. putida* harboring the GcdR vector, fluorescence was observed in stationary phase. As 2-aminoadipate immediately precedes 2-oxoadipate in the D-lysine catabolic pathway, these results support the hypothesis

that CsiD could act as a bridge between the two catabolic pathways. There has been substantial interest in developing microbes to produce glutarate, with strains of *E. coli*, *P. putida*, and *Corynebacterium glutamicum* all engineered to produce high titers [14,31,32]. Further engineering of the GcdR system may be able to extend the linear range to higher concentrations furthering its utility as a tool to achieve even higher titers of glutarate. Though glutarate is a valuable commodity chemical, the C6 diacid adipate is used in much greater quantities primarily as a monomer used to make nylons [33]. While the CsiR biosensor was highly specific for only the C5 diacid, the GcdR biosensor showed some activity towards the C6 diacid adipate. This slight sensitivity toward adipic acid is especially interesting as recent work has demonstrated the effectiveness of evolving transcription factors to sense non-native ligands [4]. Such methods could be applied to GcdR in order to expand its utility in sensing other industrially important diacids.

METHODS

Media, chemicals, and culture conditions

E. coli cultures were grown in Luria-Bertani (LB) Miller medium (BD Biosciences, USA) at 37 °C while *P. putida* was grown at 30 °C. When indicated, *P. putida* and *E. coli* were grown on modified MOPS minimal medium [34]. Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), gentamicin (30 mg/L, Fisher Scientific, USA), or carbenicillin (100 mg/L, Sigma Aldrich, USA), when indicated. All other compounds were purchased through Sigma Aldrich (Sigma Aldrich, USA) .

Strains and plasmids

All bacterial strains and plasmids used in this work are listed in Table 3. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/390). All plasmids were designed using DeviceEditor and

VectorEditor software, while all primers used for the construction of plasmids were designed using j5 software [35–37]. Plasmids were assembled via Gibson Assembly using standard protocols [38], or Golden Gate Assembly using standard protocols [39]. Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Construction of *P. putida* deletion mutants was performed as described previously [16].

Expression and purification of proteins

Proteins were purified as described previously [40]. Briefly, 500 mL cultures of *E. coli* BL21 (DE3) harboring expression plasmids were grown in LB medium at 37 ^{o}C to an OD of 0.6 then induced with 1mM Isopropyl β-D-1-thiogalactopyranoside. Cells were allowed to express for 18 hours at 30 ^{o}C before being harvested via centrifugation. Cell pellets were stored at -80 ^{o}C until purification. Cell pellets were then resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, 8% glycerol, pH 7.5) and sonicated to lyse cells. Insolubles were pelleted via centrifugation (30 minutes at 40,000 x g). The supernatant was applied to a fritted column containing Ni-NTA resin (Qiagen, USA) which had been pre-equilibrated with several column volumes of lysis buffer. The resin was washed with lysis buffer containing 50 mM imidazole, then the protein was eluted using a stepwise gradient of lysis buffer containing increasing imidazole concentrations (100 mM, 200 mM, and 400 mM). Fractions were collected and analyzed via SDS-PAGE. Purified proteins were concentrated using Spin-X UF 20 (10 kDa MWCO) spin concentrators (Corning, Inc.). Concentrated protein was stored at 4 ^{o}C until *in vitro* analysis.

Plate based growth and fluorescence assays

Growth studies of bacterial strains were conducted via microplate reader kinetic assays. Overnight cultures were inoculated into 10 mL of LB medium from single colonies, and grown at 30 °C. These cultures were then washed twice with MOPS minimal medium without any added carbon and diluted 1:100 into 500 uL of MOPS medium with 10 mM of a carbon source in 48-well plates (Falcon, 353072). Plates were sealed with a gas-permeable microplate adhesive film (VWR, USA), and then optical density and fluorescence were monitored for 48 hours in a Biotek Synergy 4 plate reader (BioTek, USA) at 30 °C with fast continuous shaking. Optical density was measured at 600 nm, while fluorescence was measured using an excitation wavelength of 485 nm and an emission wavelength of 620 nm with a manually set gain of 100.

Transcriptional fusion fluorescence assays

To measure RFP production from *E. coli* fluorescence measurements were obtained from single time points of cells grown in deep-well 96-well plates as described previously [41]. Briefly, cells were grown in 500 µL of LB medium supplemented with kanamycin and a range of glutarate concentrations from 5 mM to 0 mM. Plates were sealed with AeraSeal film (Excel Scientific, AC1201-02) and grown for 22 hours at 30 °C on a 200 rpm shaker rack. After incubation, 100 µL from each well was aliquoted into a black, clear-bottom 96-well plate and fluorescence was measured with an Biotek Synergy 4 plate reader (BioTek, USA). Optical density was measured at 600 nm, while fluorescence was measured using an excitation wavelength of 485 nm and an emission wavelength of 620 nm with a manually set gain of 45.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were performed as previously described [5]. 6-Carboxyfluorescein labelled PCR products for CsiR probes were generated from the intergenic

region between PP_2908 and PP_2909 using primers csiRprobeFOR 5'-6-FAM/AGTTCGATCTGCGTAAAG-3' and csiRprobeREV 5'-CCCGCTGAATGCTGAGTT-3'), while probes for GcdR were generated from the intergenic region between PP_0157 and PP_0158 with primers gcdRprobeFOR 5'-6-FAM/CGGGTCGATCCAGTTGAAA-3' and gcdRprobeREV 5'-GCATGTACGTCAACCTCACT-3'. Primers were purchased from IDT Technologies (IDT, Coralville, IA). PCR product was then purified with a QIAquick PCR Purification Kit (Qiagen, USA), and the amount of DNA was quantified via a NanoDrop 2000C (Thermo Fisher Scientific). Binding reactions were conducted with 10 ng of labelled probe, which was added to 10 mM Tris–HCl (pH 8.0), 25 mM KCl, 2.5 mM MgCl2, 1.0 mM DTT and 2 ug salmon sperm DNA in 20 uL reactions. CsiR was added to reactions in concentrations ranging from 250 nM to ~4 nM in addition to no CsiR controls, and then allowed to incubate at 22 ^{o}C for 20 minutes. Reactions were then loaded into 10% native polyacrylamide gels buffered with 0.5x TBE. After electrophoresis gels were imaged on an Alpha Innotech MultiImage III (Alpha Innotech). Average pixel intensity over the entire band on EMSA gels were used to estimate K_d, and was done in Adobe Photoshop.

Measurement of 5-aminovalerate

To measure intracellular concentrations of 5-aminovalerate cells were quenched as previously described [42]. LC/MS analysis was performed on an Agilent 6120 single quadrupole LC/MS equipped with a Waters Atlantis Hilic 5 µM Silica column (4.6 x 150 mm). A linear gradient of 100-30% 90:10 CH3CN:H20 with 10 mM ammonium formate and 0.1% formic acid (v/v) over 20 min in 90:10 H2O:CH3CN with 10 mM ammonium formate and 0.1% formic acid (v/v) at a flow rate of 1.0 mL/min was used. Extracted ion chromatograms were integrated and

peak area was used to construct a standard curve using an authentic 5-aminovalerate standard. Concentrations of 5-aminovalerate within samples were interpolated from this curve.

Analysis of Biosensor Parameters

A model relating inducer concentrations and fluorescence measurements to characterize the performance of a biosensor was created under the following assumptions 1) the relationship between analyte concentrations and fluorescence can be well modeled using the Hill equation 2) fluorescence measurements at a particular concentration are normally distributed 3) the variance of fluorescence measurements is roughly constant over the range of measured values.

Under these assumptions we can phrase the following probabilistic model:

$$P(F|C, \theta, \alpha) = \mathcal{N}(h_\theta(C), \sigma)$$

where is the Hill function, are the parameters of the hill function, and is the estimated standard deviation. Using the probabilistic model which captures our constraints on the problem the log likelihood function is expressed as:

$$\ell(\theta, \sigma|\mathcal{D}) = \sum_{i=1}^{N} \log P(F = f_i|C = c_i, \sigma, \theta)$$

The log likelihood is used to express the maximum likelihood estimation (MLE) problem:

$$\hat{\theta}, \hat{\sigma} = \arg\max_{\theta,\sigma} \ell(\theta, \sigma|\mathcal{D}),$$

which when solved results in the optimal parameters of the model given the characterization data. In order to estimate the distribution of ligand concentrations that are compatible with experimental fluorescence data, MCMC sampling was used to solve the following MLE problem:

$$\hat{c} = \arg\max_{c} \sum_{i=1}^{N} \log P(F = f_i | C = c, \hat{\theta}, \hat{\sigma})$$

We determined biosensor resolution by solving the above maximum likelihood estimation problem iteratively over the range of observed fluorescences during the biosensor characterization process. This can determine the relationship between an inducer concentration estimate and the estimated standard deviation. The standard deviation of the estimate of inducer concentration can be interpreted as the resolution window. In the case of this work, two standard deviations is considered the resolution window of the sensor, as 95% of the compatible inducer concentration estimates fall within the interval .

To calculate the range of detection for a biosensor we assumed for given a set of normally distributed biosensor measurement replicates with sample standard deviation σ and mean measurement value f, that the associated probabilistic model for the biosensor has already been fit. To determine the range of detection, we want to find 1) the smallest possible biosensor measurement f given the sample distribution of the data that can be distinguished from 0. 2) The largest possible biosensor measurement f given the sample distribution of the data that can be distinguished from the maximum possible observation from the characterization data times an extrapolation factor For both 1 and 2, in order to distinguish a difference we calculated a prediction interval with confidence level α over the concentration space of given fluorescence measurements. If the interval does not contain the value of interest, then we say that the value can be distinguished from the chosen fluorescence level f.

A comprehensive methodological description of calculating biosensor performance parameters can be found in the supplemental Jupyter notebook which is also attached as a pdf.

<u>Bioinformatic Analysis</u>

For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder as implemented on IQ-tree [43] phylogenetic trees were constructed using IQ-tree, nodes were supported with 10,000 bootstrap replicates. The final tree figures were edited using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Orthologous syntenic regions of CsiR were identified with CORASON-BGC [44] and manually colored and annotated. DNA-binding sites were predicted with MEME [19].

Acknowledgements

The authors would like to thank Allison Pearson for help with figure aesthetics, and Nathan Hillson for constructive criticism of this manuscript. The authors would like to thank the UC Berkeley SMART program for providing support for R.K to conduct summer research. This work was part of the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. H.G.M. was also supported by the Basque Government through the BERC 2018-2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV- 2017-0718.The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any

information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Contributions

Conceptualization, M.G.T.; Methodology, M.G.T.,Z.C, J.M.B, P.C.M.,N.H.,W.S. Investigation, M.G.T.,Z.C., J.M.B, R.N.K, P.C.M, M.R.I..; Writing – Original Draft, M.G.T, M.R.I..; Writing – Review and Editing, All authors.; Resources and supervision, J.D.K,P.S.,H.M.

Competing Interests

J.D.K. has financial interests in Amyris, Lygos, Demetrix, Napigen, and Maple Bio.

Table 3. Strains and Plasmids used in this work

Strain	**JBEI Part ID**	**Reference**
E. coli DH10B		[45]
E. coli BL21(DE3)		Novagen

E.coli S17		ATCC 47055
P. putida KT2440		ATCC 47054
P. putida Δ*davT*	JPUB_013544	This work
P. putida Δ*gcdH*Δ*csiD*		[16]
Plasmids		
pET21b		Novagen
pBADT		[46]
pMQ30		[47]
pBADT-gcdR-PgcdH::RFP	JPUB_010960	This work
pBADT-csiR-PcsiD::RFP	JPUB_010962	This work
pBADT-gcdR-J23101	JPUB_013546	This work
pBADT-gcdR-J23110	JPUB_013548	This work
pBADT-gcdR-J23113	JPUB_013550	This work
pMQ30 davT	JPUB_013544	This work
pET21b-CsiR_Pput	JPUB_010964	This work
pET21b-GcdR_Pput	JPUB_010966	This work

Bibliography

(1) Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol. 2*, 47–58.
(2) Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem. 79*, 563–590.
(3) Petzold, C. J., Chan, L. J. G., Nhan, M., and Adams, P. D. (2015) Analytics for metabolic engineering. *Front. Bioeng. Biotechnol. 3*, 135.
(4) Kasey, C. M., Zerrad, M., Li, Y., Cropp, T. A., and Williams, G. J. (2018) Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic Biology. *ACS Synth. Biol. 7*, 227–239.
(5) Lei, C., Wang, J., Liu, Y., Liu, X., Zhao, G., and Wang, J. (2018) A feedback regulatory model for RifQ-mediated repression of rifamycin export in Amycolatopsis mediterranei. *Microb. Cell Fact. 17*, 14.
(6) Zhang, J., Barajas, J. F., Burdu, M., Ruegg, T. L., Dias, B., and Keasling, J. D. (2017) Development of a Transcription Factor-Based Lactam Biosensor. *ACS Synth. Biol. 6*, 439–445.
(7) Gilbert, M. (2017) Aliphatic Polyamides, in *Brydson's Plastics Materials*, pp 487–511. Elsevier.
(8) Chae, T. U., Ahn, J. H., Ko, Y.-S., Kim, J. W., Lee, J. A., Lee, E. H., and Lee, S. Y. (2019) Metabolic engineering for the production of dicarboxylic acids and diamines. *Metab. Eng.*
(9) Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shin, J. H., Kim, D. I., Ko, Y.-S., Jang, W. D., and Jang, Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal. 2*, 18–33.
(10) Li, Z., Shen, Y.-P., Jiang, X.-L., Feng, L.-S., and Liu, J.-Z. (2018) Metabolic evolution and a comparative omics analysis of Corynebacterium glutamicum for putrescine production. *J. Ind. Microbiol. Biotechnol. 45*, 123–139.
(11) Li, M., Li, D., Huang, Y., Liu, M., Wang, H., Tang, Q., and Lu, F. (2014) Improving the secretion of cadaverine in Corynebacterium glutamicum by cadaverine-lysine antiporter. *J. Ind. Microbiol. Biotechnol. 41*, 701–709.
(12) Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J., and Eirich, L. D. (1992) Metabolic engineering of candida tropicalis for the production of long–chain dicarboxylic acids. *Nat. Biotechnol. 10*, 894–898.
(13) Thompson, M., Valencia, L. E., Blake-Hedges, J., Cruz-Morales, P., Velasquez, A., Pearson, A., Sermeno, L., Sharpless, W., Benites, V., Chen, Y., Baidoo, E., Petzold, C. J., Deutschbauer, A., and Keasling, J. D. (2019) Host engineering for improved valerolactam production in Pseudomonas putida. *BioRxiv*.
(14) Zhang, M., Gao, C., Guo, X., Guo, S., Kang, Z., Xiao, D., Yan, J., Tao, F., Zhang, W., Dong, W., Liu, P., Yang, C., Ma, C., and Xu, P. (2018) Increased glutarate production by blocking the glutaryl-CoA dehydrogenation pathway and a catabolic pathway involving L-2-hydroxyglutarate. *Nat. Commun. 9*, 2114.

(15) Zhang, J., Barajas, J. F., Burdu, M., Wang, G., Baidoo, E. E., and Keasling, J. D. (2017) Application of an Acyl-CoA Ligase from Streptomyces aizunensis for Lactam Biosynthesis. *ACS Synth. Biol. 6*, 884–890.
(16) Thompson, M. G., Blake-Hedges, J. M., Cruz-Morales, P., Barajas, J. F., Curran, S. C., Eiben, C. B., Harris, N. C., Benites, V. T., Gin, J. W., Sharpless, W. A., Twigg, F. F., Skyrud, W., Krishna, R. N., Henrique Pereira, J., Baidoo, E. E. K., Petzold, C. J., Adams, P. D., Arkin, A. P., Deutschbauer, A. M., and Keasling, J. D. (2019) Massively Parallel Fitness Profiling Reveals Multiple Novel Enzymes in Pseudomonas putida Lysine Metabolism. *MBio 10*.
(17) Madhuri Indurthi, S., Chou, H.-T., and Lu, C.-D. (2016) Molecular characterization of lysR-lysXE, gcdR-gcdHG and amaR-amaAB operons for lysine export and catabolism: a comprehensive lysine catabolic network in Pseudomonas aeruginosa PAO1. *Microbiology (Reading, Engl) 162*, 876–888.
(18) Knorr, S., Sinn, M., Galetskiy, D., Williams, R. M., Wang, C., Müller, N., Mayans, O., Schleheck, D., and Hartig, J. S. (2018) Widespread bacterial lysine degradation proceeding via glutarate and L-2-hydroxyglutarate. *Nat. Commun. 9*, 5071.
(19) Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res. 34*, W369-73.
(20) Jain, D. (2015) Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *IUBMB Life 67*, 556–563.
(21) Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., and Keasling, J. D. (2010) BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng. 4*, 1.
(22) Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014) Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into Escherichia coli. *Metab. Eng. 21*, 2–8.
(23) Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.
(24) Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., Kuehl, J. V., Melnyk, R. A., Lamson, J. S., Suh, Y., Carlson, H. K., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Zane, G. M., Rubin, B. E., Wall, J. D., Visel, A., Bristow, J., Blow, M. J., and Deutschbauer, A. M. (2018) Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature 557*, 503–509.
(25) Sagawa, S., Price, M. N., Deutschbauer, A. M., and Arkin, A. P. (2017) Validating regulatory predictions from diverse bacteria with mutant fitness data. *PLoS ONE 12*, e0178258.
(26) Nitta, K. R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E. E. M., and Taipale, J. (2015) Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *elife 4*.
(27) Suvorova, I. A., Korostelev, Y. D., and Gelfand, M. S. (2015) GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS ONE 10*, e0132618.

(28) Way, J. C., Collins, J. J., Keasling, J. D., and Silver, P. A. (2014) Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell 157*, 151–161.
(29) Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T. R., and Mukhopadhyay, A. (2019) Engineering Robust Production Microbes for Large-Scale Cultivation. *Trends Microbiol. 27*, 524–537.
(30) Revelles, O., Espinosa-Urgel, M., Molin, S., and Ramos, J. L. (2004) The davDT operon of Pseudomonas putida, involved in lysine catabolism, is induced in response to the pathway intermediate delta-aminovaleric acid. *J. Bacteriol. 186*, 3439–3446.
(31) Park, S. J., Kim, E. Y., Noh, W., Park, H. M., Oh, Y. H., Lee, S. H., Song, B. K., Jegal, J., and Lee, S. Y. (2013) Metabolic engineering of Escherichia coli for the production of 5-aminovalerate and glutarate as C5 platform chemicals. *Metab. Eng. 16*, 42–47.
(32) Rohles, C. M., Gießelmann, G., Kohlstedt, M., Wittmann, C., and Becker, J. (2016) Systems metabolic engineering of Corynebacterium glutamicum for the production of the carbon-5 platform chemicals 5-aminovalerate and glutarate. *Microb. Cell Fact. 15*, 154.
(33) Hagen, A., Poust, S., Rond, T. de, Fortman, J. L., Katz, L., Petzold, C. J., and Keasling, J. D. (2016) Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth. Biol. 5*, 21–27.
(34) LaBauve, A. E., and Wargo, M. J. (2012) Growth and laboratory maintenance of Pseudomonas aeruginosa. *Curr. Protoc. Microbiol. Chapter 6*, Unit 6E.1.
(35) Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., and Keasling, J. D. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res. 40*, e141.
(36) Chen, J., Densmore, D., Ham, T. S., Keasling, J. D., and Hillson, N. J. (2012) DeviceEditor visual biological CAD canvas. *J. Biol. Eng. 6*, 1.
(37) Hillson, N. J., Rosengarten, R. D., and Keasling, J. D. (2012) j5 DNA assembly design automation software. *ACS Synth. Biol. 1*, 14–21.
(38) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343–345.
(39) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE 3*, e3647.
(40) Jackson, D. R., Tu, S. S., Nguyen, M., Barajas, J. F., Schaub, A. J., Krug, D., Pistorius, D., Luo, R., Müller, R., and Tsai, S.-C. (2016) Structural Insights into Anthranilate Priming during Type II Polyketide Biosynthesis. *ACS Chem. Biol. 11*, 95–103.
(41) Thompson, M. G., Sedaghatian, N., Barajas, J. F., Wehrs, M., Bailey, C. B., Kaplan, N., Hillson, N. J., Mukhopadhyay, A., and Keasling, J. D. (2018) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci. Rep. 8*, 1590.
(42) George, K. W., Thompson, M. G., Kim, J., Baidoo, E. E. K., Wang, G., Benites, V. T., Petzold, C. J., Chan, L. J. G., Yilmaz, S., Turhanen, P., Adams, P. D., Keasling, J. D., and Lee, T. S. (2018) Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-

producing Escherichia coli. *Metab. Eng. 47*, 60–72.
(43) Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods 14*, 587–589.
(44) Cruz-Morales, P., Ramos-Aboites, H. E., Licona-Cassani, C., Selem-Mójica, N., Mejía-Ponce, P. M., Souza-Saldívar, V., and Barona-Gómez, F. (2017) Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore functional characterization, unveil new desferrioxamine traits. *FEMS Microbiol. Ecol. 93*.
(45) Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc Natl Acad Sci USA 87*, 4645–4649.
(46) Bi, C., Su, P., Müller, J., Yeh, Y.-C., Chhabra, S. R., Beller, H. R., Singer, S. W., and Hillson, N. J. (2013) Development of a broad-host synthetic biology toolbox for Ralstonia eutropha and its application to engineering hydrocarbon biofuel production. *Microb. Cell Fact. 12*, 107.
(47) Shanks, R. M. Q., Kadouri, D. E., MacEachran, D. P., and O'Toole, G. A. (2009) New yeast recombineering tools for bacteria. *Plasmid 62*, 88–97.