Title

A song of lysine and Pseudomonas putida

Author

Thompson, Mitchell G

Publication Date

2019

A song of lysine and *Pseudomonas putida*

By

Mitchell Gregory Thompson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D Keasling, Chair
Professor Adam P Arkin
Professor Rasmus Nielsen
Professor Matthew Traxler

Summer 2019

A song of lysine and *Pseudomonas putida*

Abstract

A song of lysine and *Pseudomonas putida*

by

Mitchell Gregory Thompson

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Jay D Keasling, Chair

Synthetic biology attempts to make inherently chaotic biological systems both predictable and useful. To achieve this, well characterized genetic parts are necessary both to control genetic circuitry as well as to catalyze specific chemical reactions. While the discipline is maturing, the number of chemical reactions we can catalyze and the number of means to control the expression of these enzymes is still extremely limited in the context of what occurs in nature. It therefore behooves the discipline to rapidly explore nature for biological novelty so we can then leverage it for biological programming. Herein I utilize multiple functional genomics techniques to uncover novel enzymatic and regulatory elements within the lysine catabolism of the bacterium *Pseudomonas putida*. From these data we are able to develop novel tools for synthetic biology, improve *P. putida* as a chassis for metabolic engineering of bioproducts, as well as make fundamental discoveries about previously unknown enzymatic steps of the lysine catabolism of both bacteria and plants.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation is first and foremost dedicated to my parents, Penny and Jeff. Their unending support has made all of my scientific career possible. I love you both dearly. I would also like to thank my sister, Mary, has been a constant friend and supporter since forever. This dissertation would also not be possible without my fiancé Cat Adams. While many on this page have made me a better scientist, Cat has made me a better person.

My advisor, Jay Keasling, created an environment that was unparalleled in which to learn in. He showed infinite patience in allowing me to explore my many ideas that were well outside the lab norm. I am forever indebted for his enthusiasm and support during my time in his lab. It was Great! I would also like to thank Adam Deutschbauer who allowed me unfettered access to his lab and provided critical support throughout my graduate career. I would also like to thank my committee Adam Arkin, Matt Traxler, and Rasmus Nielsen for their help and input during my time at Berkeley.

This work would never be possible without my teachers and PIs before I came to graduate school Allison Miller, Anthony Hay, Faustino Toba, Dan Zurawski, and Chad Black. I stand on their shoulders.

I want to thank my mentors in the Keasling Lab, Kevin George, Maggie Brown, and Tristan deRond, who taught me how to become a better scientist.

My thesis would be a shell of what it is without my scientific partner in crime Jackie Blake-Hedges. She was the biochemical Ying to my genetics Yang, and working with her made science fun. Pablo Cruz-Morales, Jesus Barajas, Jose Henrique Pereira, Zak Costello, John Hangasky, Morgan Price, and Kelly Wetmore all provided scientific input and expertise that I never could and allowed me to do science that would not have otherwise been possible.

I want to thank the undergraduates who have worked with me Nima Sedaghatian, Ron Yadin, Janel Weatherly, Will Sharpless, Allie Pearson, Rohith Krishna, and Allie Velasquez. Teaching them has been the most rewarding part of my PhD.

My other co-authors Bobbo Haushalter, Sam Curran, Chris Eiben, Nick Harris, Veronica Benites, Jennifer Gin, Fred Twigg, Will Skyrud, Ed Baidoo, Chris Petzold, Paul Adams, Niklas Hummel, Matt Incha, Pat Shih, Hector Garcia-Martin, Luis Valencia, Lauren Sermeno, Yan Chen, Joonhoon Kim, George Wang, Leanne Chan, Suzan Yilmaz, Petri Turhanen, Taek Soon Lee, Maren Wehrs, Constance Bailey, Nurgul Kaplan, Nathan Hillson, Aindrila Mukhopadhyay, Michael Belcher, William Moore, Tyler Barnum[4]Yuzhong Lui, Michael A. Marletta, and Megan E. Garber all provided assistance that made this thesis possible.

Chapter 1. A World of Genes and Chemicals

1.1. Functional genomics to inform synthetic biology

The gene is the fundamental unit of biological information, though millions of genes exist in nature for which we have no known function. This represents a potentially tremendous reservoir of untapped enzymatic reactions, chemical sensors, and other biological functions we have yet to comprehend. One of the great tasks left in biological research is to rapidly assign function to these genes, which are often referred to as domains of unknown function (DUFs). This task often seems Sisyphean, given the tremendous pace at which we are now able to sequence the environment. Currently the pace at which we uncover novel proteins via next-generation sequencing far outstrips the pace at which we can figure out what exactly they do. However, recent advances in functional genomics (RNAseq, Proteomics, Metabolomics, Transposon Sequencing) coupled with and increasing ability to manipulate the genomes of non-model organisms (CRISPR-Cas9) provides hope that as researchers we can put a dent in the ~3000 DUFs that we are currently aware of. Microorganisms contain within their genomes the majority of these DUFs, which is unsurprising given their wide environmental dispersal and their diverse metabolic capabilities. In addition to answering fundamental questions about how life works, these unknown proteins may have valuable applications.

Synthetic biology is the field of converting inherently chaotic biological organisms into predictable systems capable of doing useful work. The words that we write this code in are the transcription factors that control gene expression, the enzymes that convert sugar into jet fuel, and the efflux pumps that keep cells safe from toxic molecules. However, relatively few enzymes are used to interconvert biochemicals, and even fewer molecules are used to control the expression of our genetic circuits. This seems at times an almost criminal underutilization of the tremendous genetic bounty that has evolved on earth. The rate limiting step, of course, is associating ligands to sensors, DNA-binding domains to promoters, and substrates and products to enzymes. In this dissertation, I hope to demonstrate on a small scale how we can rapidly increase our synthetic biology tool box by leveraging the omics levels techniques currently available to us to answer straight-forward metabolic questions in microbial hosts.

.

1.2. Dissertation organization

In **Chapter 2** we leverage Random Barcode Transposon Sequencing (RB-TnSeq), to uncover the missing gaps of L- and D-lysine metabolism in the bacterium *Pseudomonas putida*. We discover a 3 enzyme pathway that converts 2-aminoadipate to D-2-hydroxyglutarate. One of these enzymes, PP_5260, is a member of the DUF1338 family and previously had no known function. We demonstrate that it catalyzes a novel 2-oxoadipate to D-2-hydroxyglutarate decarboxylation. Our data also identified the recently discovered coA-independent route of glutarate catabolism in the bacterium. We expand on previous work by showing that the enzyme CsiD is promiscuous in its 2-oxoacid selectivity, and demonstrate a plausible biochemical connection between the L-and D-lysine catabolic pathways. Finally, we demonstrate that the expression of enzymes within *P. putida* lysine catabolism are highly sensitive to specific pathway metabolites implying significant regulation.

In **Chapter 3** we build upon the work of chapter 2 by characterizing two transcription factors, CsiR and GcdR, that control the expression of two distinct routes of glutarate catabolism in *P. putida*. After bioinformatics and biochemical characterization of the binding site and affinities of each regulator we develop a suite of glutarate responsive biosensors. We demonstrate the utility of these biosensors by using them to measure the putative flux to glutarate in *P. putida*. Importantly we also develop a novel mathematical means of defining the range of detection for genetically encoded biosensors using Markov-Chain Monte Carlo sampling.

In **Chapter 4** we identify the route of valerolactam catabolism in *P. putida* using a combination of shotgun proteomics and RB-TnSeq. We identify the predicted amidase OplBA as the enzyme putatively responsible for the hydrolysis of both valerolactam and caprolactam. By eliminating *oplBA* and other loci that compete for the precursor 5-aminovalerate we are able to increase titers of valerolactam from 0 mg/L to 90 mg/L after 48 hours of fermentation.

In **Chapter 5** we show that the expression of *oplBA* is regulated by a transcription factor that is highly sensitive to both valerolactam and caprolactam. We create plasmid based biosensors based on the regulator, OplR, which can be engineered to detect concentrations of the lactams from 12 nM to 1 mM, making this lactam biosensor over 1000x more sensitive that previously published vectors. Using one of these vectors we are able to detect the production of caprolactam in *E. coli* at titers as low as 0.01 mg/L.

In **Chapter 6** we use a combination of structural biology and biochemical analyses to identify the mechanism of action and specificity of the DUF1338 containing enzyme PP_5260. We show that enzyme catalyzes and oxygen and Fe (II) dependent decarboxylation followed by an intramolecular hydroxylation to convert 2-oxoadipate to 2-hydroxyglutarate. We then show that the specificity of the enzyme for 2-oxoadipate is mediated by a highly conserved arginine within the active site. Homologs from plants, while having low sequence identity compared to bacterial copies, were found to be highly co-expressed with the known enzymes of lysine catabolism. Structural and biochemical analysis of the *Oryzae sativa* homolog FLO7 revealed that plant homologs also catalyzed a 2-oxoadipate the 2-hydroxyglutarate reaction and represents that final missing step of plant lysine catabolism.

Chapter 2. A Game of Lysine: Massively parallel fitness profiling reveals multiple novel enzymes in *Pseudomonas putida* lysine metabolism

Including material from published work from **MG Thompson**, JM Blake-Hedges, P Cruz-Morales, JF Barajas, SC Curran, CB Eiben, NC Harris, VT Benites, JW Gin, WA Sharpless, FF Twigg, W Skyrud, RN Krishna, J Henrique Pereira, EEK Baidoo, CJ Petzold, PD Adams, AP Arkin, AM Deutschbauer, JD Keasling. "Massively parallel fitness profiling reveals multiple novel enzymes in *Pseudomonas putida* lysine metabolism", mBio, 2019

2.1. Abstract

Despite intensive study for 50 years, the biochemical and genetic links between lysine metabolism and central metabolism in *Pseudomonas putida* remain unresolved. To establish these biochemical links, we leveraged Random Barcode Transposon Sequencing (RB-TnSeq), a genome-wide assay measuring the fitness of thousands of genes in parallel, to identify multiple novel enzymes in both L- and D-lysine metabolism. We first describe three pathway enzymes that catabolize L-2-aminoadipate (L-2AA) to 2-ketoglutarate (2KG), connecting D-lysine to the TCA cycle. One of these enzymes, PP_5260, contains a DUF1338 domain, a family with no previously described biological function. Our work also identified the recently described CoA independent route of L-lysine degradation that metabolizes to succinate. We expanded on previous findings by demonstrating that glutarate hydroxylase CsiD is promiscuous in its 2-oxoacid selectivity. Proteomics of select pathway enzymes revealed that expression of catabolic genes is highly sensitive to particular pathway metabolites, implying intensive local and global regulation. This work demonstrates the utility of RB-TnSeq for discovering novel metabolic pathways in even well-studied bacteria, as well as a powerful tool for validating previous research.

2.2. Significance statement

P. putida lysine metabolism can produce multiple commodity chemicals, conferring great biotechnological value. Despite much research, connecting lysine catabolism to central metabolism in *P. putida* remained undefined. Herein we use Random Barcode Transposon Sequencing to fill in the gaps of lysine metabolism in *P. putida*. We describe a route of 2-oxoadipate (2OA) catabolism in bacteria, which utilizes DUF1338 containing protein PP_5260. Despite its prevalence in many domains of life, DUF1338 containing proteins had no known biochemical function. We demonstrate PP_5260 is a metalloenzyme which catalyzes an unusual 2OA to D-2HG decarboxylation. Our screen also identified a recently described novel glutarate metabolic pathway. We validate previous results, and expand the understanding of glutarate hydroxylase CsiD by showing can it use either 2OA or 2KG as a cosubstrate. Our work demonstrates biological novelty can be rapidly identified using unbiased experimental genetics, and that RB-TnSeq can be used to rapidly validate previous results.

2.3. Introduction

Pseudomonas putida is an ubiquitous saprophytic soil bacterium and is a model organism for bioremediation [1]. Interest in utilizing *P. putida* KT2440 as a chassis organism for metabolic engineering has recently surged due to the existence of well-established genetic tools and its robust metabolism of aromatic compounds that resemble lignin hydrolysis products [2–4]. As lignin valorization remains essential for the economic feasibility of cellulosic bioproducts, a nuanced and predictable understanding of *P. putida* metabolism is highly desirable [5].

Although its aromatic metabolism has garnered much attention, the lysine metabolism of *P. putida* has also been rigorously studied for over fifty years [6]. An understanding of lysine metabolism has had biotechnological value, as it has been used to produce glutarate, 5-aminovalerate (5AVA), as well as valerolactam in *P. putida* and in the other bacteria [7–10]. Our current understanding of lysine catabolism however, remains incomplete. In particular, the connection between D-lysine metabolism and central metabolism in *P. putida* is unclear and has not been fully characterized.

P. putida employs bifurcating pathways to catabolize lysine, separately metabolizing the L- and D-isomers (Figure 2-1a) [11]. The L-lysine degradation pathway proceeds to glutarate, which can then be either be degraded to acetyl-CoA via a glutaryl-CoA intermediate, or to succinate without a CoA bound intermediate (Figure S1a) [9]. The final steps of D-lysine catabolism remain more elusive. The initial steps of D-lysine catabolism are well described, but the genetic basis stops at 2AA [12]. Furthermore, ^{13}C labeling experiments by Revelles et al. demonstrated a putative metabolic connection between the D- and L-lysine pathways at 2AA [11]. Subsequent steps to central carbon metabolism have never been fully validated. [6,11–13]. Given the importance of lysine metabolism, and recent availability of high-throughput genetic tools, we sought to identify the missing steps in D-lysine metabolism that have remained despite 50 years of research.

Random barcode transposon sequencing (RB-TnSeq) is a genome-wide approach that measures the importance of each gene to growth (or fitness) in a massively parallel assay [14]. RB-TnSeq can identify phenotypes for thousands of previously uncharacterized genes [14,15], including the levulinic acid degradation pathway in *P. putida* KT2440 [16]. In this study, we applied RB-TnSeq to uncover multiple novel genes implicated in L- and D-lysine metabolism in *P. putida*. We first describe a three enzyme route connecting L-2AA to 2KG (Figure 2-1b). Within this pathway, D-lysine metabolism connects to central metabolism through a 2HG intermediate, which is directly produced from 2OA in a reaction catalyzed by a DUF1338-containing protein. This protein family, widely distributed across many domains of life, previously had no known function. Subsequently, we further characterize the glutarate hydroxylase CsiD, by demonstrating its 2-oxoacid promiscuity during the hydroxylation of glutarate. Finally, we show the expression of all newly discovered enzymes changes significantly in response to specific metabolites within the two catabolic pathways.

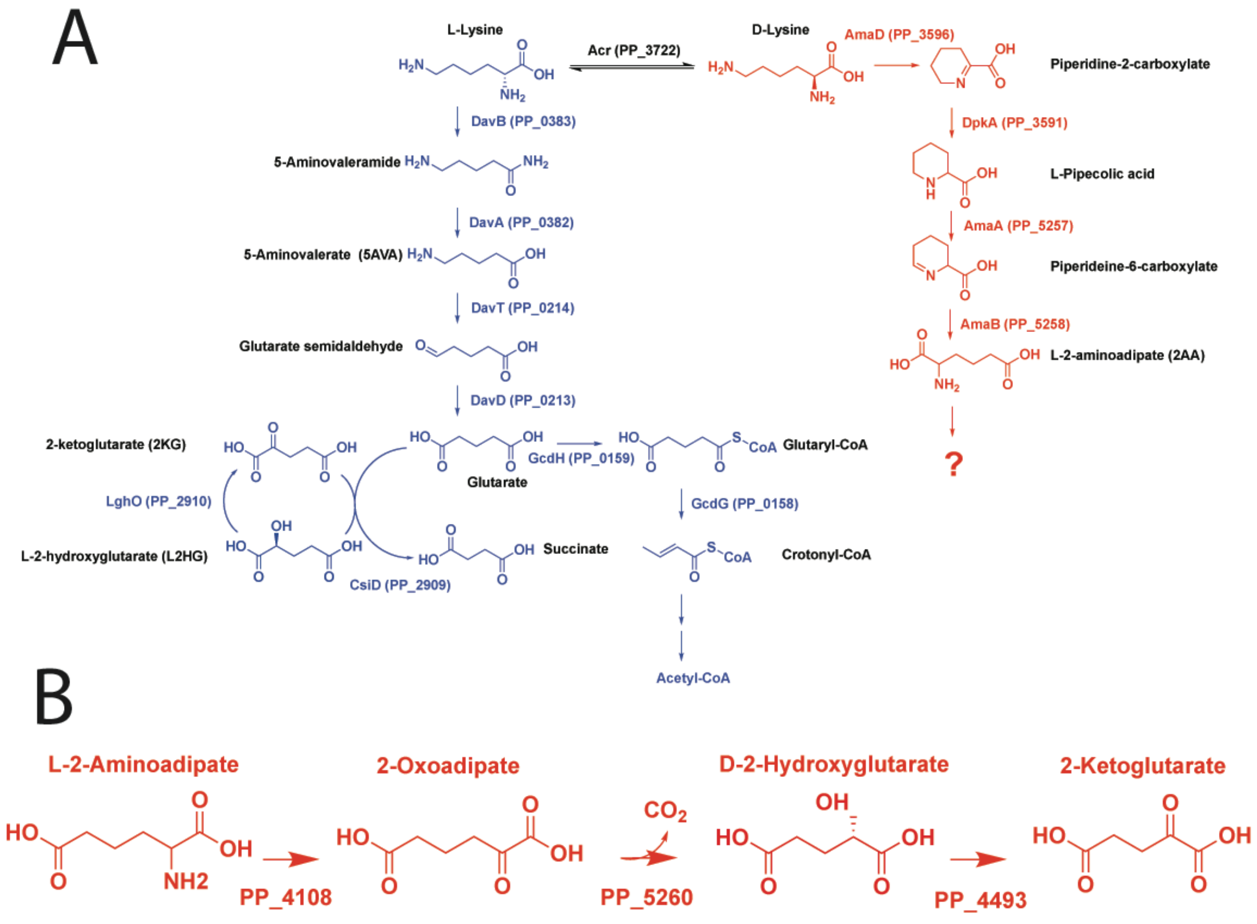


Figure 2-1 Metabolic pathways of lysine catabolism in *P. putida* KT2400. A) L-lysine metabolic pathway is shown in blue, while the known steps of D-lysine metabolism are shown in red. B) Proposed route of 2AA metabolism in *P. putida*.

2.4. Results

2.4.1. Identification of lysine catabolism genes via RB-TnSeq

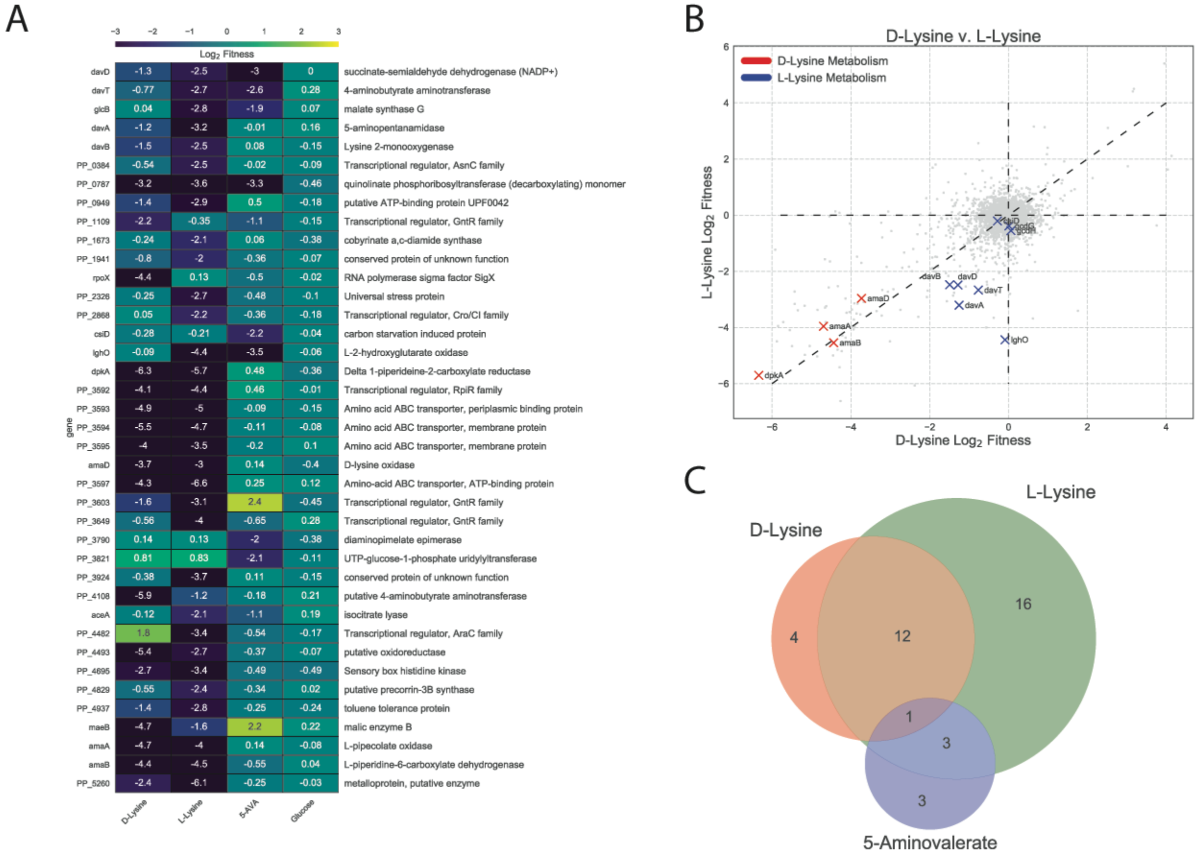
To identify mutants defective in lysine catabolism in *P. putida* KT2440, an RB-TnSeq library of this bacterium [16] was grown on minimal medium supplemented with either D-lysine or L-lysine as the sole carbon source. To evaluate whether D-lysine metabolism was required for the metabolism of other downstream metabolites of L-lysine, the library was also grown on 5AVA. As a control, we also grew the library on glucose. Fitness was calculated as the $\log_2$ ratio of strain and gene abundance at the end of selective growth relative to initial abundance [14]. Fitness profiling revealed 39 genes with significant fitness values of less than -2 for 5AVA, D-lysine, or L-lysine, and no less than -0.5 fitness for glucose (Figure 2-2a). Within this set, 10 of the 12 known lysine degradation genes were identified, with the exception of the two enzymes in the CoA-dependent route of glutarate degradation (*gcdH* and *gcdG*), which both had significant fitness values ($t < -4$) but whose fitness was greater than -2. Instead, we identified the recently-characterized genes involved in the CoA independent pathway (*csiD* and *lghO*) [9].

The fitness data corroborated previous work showing a functional D-lysine pathway is required for L-lysine catabolism [6,11]. None of the known L-lysine catabolic genes showed fitness defects for growth on D-lysine, but transposon insertions in all previously-identified D-lysine genes showed negative fitness scores when grown on L-lysine (Figure 2-2b). No known D-lysine catabolic enzymes showed fitness defects when grown on 5AVA, suggesting the D-lysine dependence of L-lysine catabolism may only occur for early catabolic steps (Figure 2-2c).

In addition to catabolic enzymes, lysine transporters and multiple transcriptional regulators were identified (Figure 2-2a). The putative lysine amino acid ABC transporter system (PP_3593, PP_3394, and PP_3395) showed significant fitness defects when grown with either isomer of lysine. Some of the transcriptional regulators were located near known catabolic or transport enzymes (PP_0384, PP_3592, and PP_3603), while others were not clustered with any obviously related genes (PP_1109, PP_2868, PP_3649, and PP_4482). Two known global regulators were identified in our screen: *cbrA* (PP_4695), a histidine kinase sensor that showed fitness defects on both lysine isomers, and the alternative sigma factor *rpoX* (PP_2088) which only had fitness defects when grown on D-lysine.

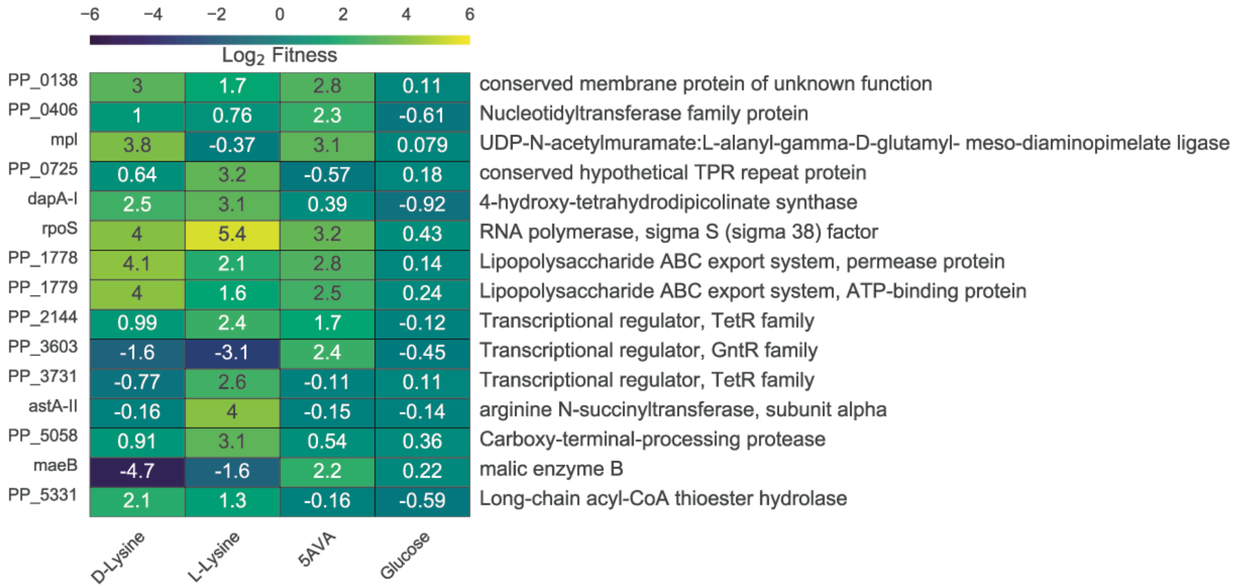
Additionally, there were 15 genes which, when disrupted, displayed fitness advantages greater than 2 on 5AVA, D-lysine, or L-lysine and less than 0.5 fitness when grown on glucose. This positive fitness value indicates these mutations confer a competitive advantage compared to other strains when grown on these carbon sources. Most striking amongst these genes were the sigma factor *rpoS* and the LPS export system (PP_1778/9), which when disrupted, both displayed fitness benefits on all three non-glucose carbon sources (Figure 2-3).

Only one gene (PP_0787, a quinolinate phosphoribosyltransferase) showed fitness defects on all three non-glucose carbon sources (Figure 1c). However, disruption of PP_0787 also showed a significant fitness defect when grown on levulinic acid, suggesting it is unlikely to be uniquely important to lysine metabolism [16]. Only 3 genes shared fitness defects between 5AVA and L-lysine (*davT*, *davD*, and *lghO*), all of which have been previously implicated in 5AVA metabolism (Figure 1c) [9].



gene	D-Lysine	L-Lysine	5-AVA	Glucose	
davD	-1.3	-2.5	-3	0	succinate-semialdehyde dehydrogenase (NADP+)
davT	-0.77	-2.7	-2.6	0.28	4-aminobutyrate aminotransferase
glcB	0.04	-2.8	-1.9	0.07	malate synthase G
davA	-1.2	-3.2	-0.01	0.16	5-aminopentanamidase
davB	-1.5	-2.5	0.08	-0.15	Lysine 2-monooxygenase
PP_0384	-0.54	-2.5	-0.02	-0.09	Transcriptional regulator, AsnC family
PP_0787	-3.2	-3.6	-3.3	-0.46	quinolinate phosphoribosyltransferase (decarboxylating) monomer
PP_0949	-1.4	-2.9	0.5	-0.18	putative ATP-binding protein UPF0042
PP_1109	-2.2	-0.35	-1.1	-0.15	Transcriptional regulator, GntR family
PP_1673	-0.24	-2.1	0.06	-0.38	cobyrinate a,c-diamide synthase
PP_1941	-0.8	-2	-0.36	-0.07	conserved protein of unknown function
rpoX	-4.4	0.13	-0.5	-0.02	RNA polymerase sigma factor SigX
PP_2326	-0.25	-2.7	-0.48	-0.1	Universal stress protein
PP_2868	0.05	-2.2	-0.36	-0.18	Transcriptional regulator, Cro/CI family
csiD	-0.28	-0.21	-2.2	-0.04	carbon starvation induced protein
lghO	-0.09	-4.4	-3.5	-0.06	L-2-hydroxyglutarate oxidase
dpkA	-6.3	-5.7	0.48	-0.36	Delta 1-piperideine-2-carboxylate reductase
PP_3592	-4.1	-4.4	0.46	-0.01	Transcriptional regulator, RpiR family
PP_3593	-4.9	-5	-0.09	-0.15	Amino acid ABC transporter, periplasmic binding protein
PP_3594	-5.5	-4.7	-0.11	-0.08	Amino acid ABC transporter, membrane protein
PP_3595	-4	-3.5	-0.2	0.1	Amino acid ABC transporter, membrane protein
amaD	-3.7	-3	0.14	-0.4	D-lysine oxidase
PP_3597	-4.3	-6.6	0.25	0.12	Amino-acid ABC transporter, ATP-binding protein
PP_3603	-1.6	-3.1	2.4	-0.45	Transcriptional regulator, GntR family
PP_3649	-0.56	-4	-0.65	0.28	Transcriptional regulator, GntR family
PP_3790	0.14	0.13	-2	-0.38	diaminopimelate epimerase
PP_3821	0.81	0.83	-2.1	-0.11	UTP-glucose-1-phosphate uridylyltransferase
PP_3924	-0.38	-3.7	0.11	-0.15	conserved protein of unknown function
PP_4108	-5.9	-1.2	-0.18	0.21	putative 4-aminobutyrate aminotransferase
aceA	-0.12	-2.1	-1.1	0.19	isocitrate lyase
PP_4482	1.8	-3.4	-0.54	-0.17	Transcriptional regulator, AraC family
PP_4493	-5.4	-2.7	-0.37	-0.07	putative oxidoreductase
PP_4695	-2.7	-3.4	-0.49	-0.49	Sensory box histidine kinase
PP_4829	-0.55	-2.4	-0.34	0.02	putative precorrin-3B synthase
PP_4937	-1.4	-2.8	-0.25	-0.24	toluene tolerance protein
maeB	-4.7	-1.6	2.2	0.22	malic enzyme B
amaA	-4.7	-4	0.14	-0.08	L-pipecolate oxidase
amaB	-4.4	-4.5	-0.55	0.04	L-piperidine-6-carboxylate dehydrogenase
PP_5260	-2.4	-6.1	-0.25	-0.03	metalloprotein, putative enzyme

Figure 2-2 Results of RB-TnSeq Screen. A) Genes that showed less than -2 $\log_2$ fitness on either D-lysine, L-lysine, or 5AVA but showed no less than -0.5 $\log_2$ fitness defect when grown on glucose. B) Plot of genome wide fitness values of libraries grown on either L-lysine or D-lysine. Genes encoding for enzymes known to be involved in D-lysine metabolism are shown in red, while those known to be involved in L-lysine metabolism are shown in blue. C) Venn diagram of genes with significant fitness defects when grown on either D-lysine, L-lysine, or 5AVA.



	D-Lysine	L-Lysine	5AVA	Glucose	
PP_0138	3	1.7	2.8	0.11	conserved membrane protein of unknown function
PP_0406	1	0.76	2.3	-0.61	Nucleotidyltransferase family protein
mpl	3.8	-0.37	3.1	0.079	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl- meso-diaminopimelate ligase
PP_0725	0.64	3.2	-0.57	0.18	conserved hypothetical TPR repeat protein
dapA-I	2.5	3.1	0.39	-0.92	4-hydroxy-tetrahydrodipicolinate synthase
rpoS	4	5.4	3.2	0.43	RNA polymerase, sigma S (sigma 38) factor
PP_1778	4.1	2.1	2.8	0.14	Lipopolysaccharide ABC export system, permease protein
PP_1779	4	1.6	2.5	0.24	Lipopolysaccharide ABC export system, ATP-binding protein
PP_2144	0.99	2.4	1.7	-0.12	Transcriptional regulator, TetR family
PP_3603	-1.6	-3.1	2.4	-0.45	Transcriptional regulator, GntR family
PP_3731	-0.77	2.6	-0.11	0.11	Transcriptional regulator, TetR family
astA-II	-0.16	4	-0.15	-0.14	arginine N-succinyltransferase, subunit alpha
PP_5058	0.91	3.1	0.54	0.36	Carboxy-terminal-processing protease
maeB	-4.7	-1.6	2.2	0.22	malic enzyme B
PP_5331	2.1	1.3	-0.16	-0.59	Long-chain acyl-CoA thioester hydrolase

Figure 2-3 Results of RB-TnSeq Screen with Fitness Benefit. A) Genes that showed greater than 2 log2 fitness on either D-lysine, L-lysine, or 5AVA but showed no less than 0.5 log2 fitness defect when grown on glucose.

2.4.2. PP_4108 is an L-2AA aminotransferase

In humans and other animals, L-lysine degradation proceeds through a 2AA intermediate, which a transaminase converts to 2OA [9,11,17]. Yet, no such transaminase has been identified in *P. putida*. We identified a candidate aminotransferase, PP_4108, for which gene inactivation showed a significant growth defect on D-lysine (-5.9) and a relatively minor defect on L-lysine (-1.2). To corroborate our RB-TnSeq fitness data, we constructed a deletion mutant of PP_4108 that failed to grow in a plate reader assay on 10 mM DL-2AA (Figure 2-4a). The mutant showed a severe growth defect on 10 mM D-lysine and an increased lag time when grown on 10 mM L-lysine (Figure 2-5).

To further validate this hypothesis, the ΔPP_4108 strain was subjected to metabolomics analysis to monitor the accumulation of its expected substrate, 2AA, when grown on glucose and D-lysine. After 12 hours of growth on minimal media supplemented with 10 mM each of glucose and D-lysine, the PP_4108 deletion strain showed a 6.3-fold increase ($p = 0.00016$) in normalized intracellular 2AA concentration compared to WT (Figure 2-4b). Next PP_4108 was expressed and purified from *E. coli* for biochemical characterization. After purified enzyme incubation with DL-2AA, 2KG, and pyridoxal phosphate (PLP) for 16 hours, the reaction mixture was analyzed with LC-TOF. The expected product, 2OA, was detected in the enzymatic reaction but not in a boiled enzyme control, confirming PP_4108 as a transaminase that converts 2AA to 2OA (Figure 2-4c). As many transaminases have broad substrate specificity [18], we also probed the substrate range of PP_4108 using a colorimetric assay for glutamate, a stoichiometric product of the transamination reaction (Figure 2-4d). The enzyme was most active on L-2AA, and only showed 2.8% relative activity ($p = 0.0057$) on its enantiomer, D-2AA. This specificity for the L-2AA isomer may explain why only 50% of the DL-2AA was transformed in the previous experiment (Figure 2-4c). No activity was observed on either lysine isomer; however, the enzyme had slight activity towards 4-aminobutyrate/γ-aminobutyrate (GABA) (2.8% relative activity, $p = 0.0057$) and moderate activity on 5AVA (30.5% relative activity, $p = 0.0139$). Over shorter time scales PP_4108 had no activity on any substrate expect L-2AA (Figure 2-5c). These results suggest *P. putida* KT2440 metabolizes D-lysine to L-2AA, which is then converted to 2OA by the transaminase PP_4108.

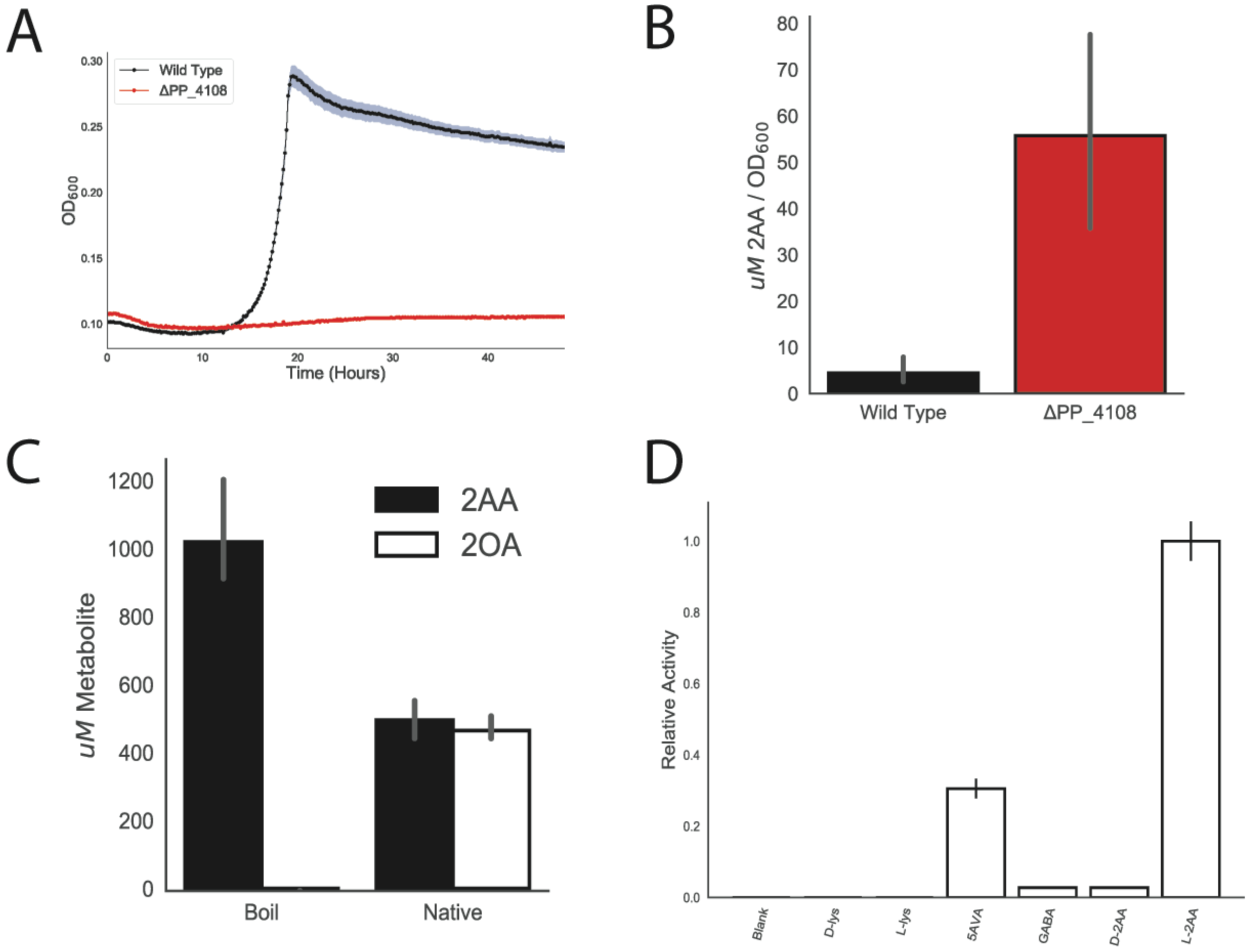


Figure 2-4 Identification of PP_4108 as an L-2AA aminotransferase. A) Growth of wild-type KT2440 and PP_4108 mutants on 2AA as a sole carbon source. Shaded area represents 95% confidence interval (CI), n=3. B) *In vivo* accumulation of 2AA in wild-type KT2440 and a PP_4108 mutant after 12 hours of growth on minimal medium supplemented with 10 mM glucose and 10 mM D-lysine. Bars represent $\log_{10}$ transformed spectral counts, error bars show 95% CI, n=3. C) *In vitro* transamination reactions of PP_4108 with 2KG as an amino acceptor. Bars represent μM metabolite concentration of either 2OA (black) or 2AA (white) in either boiled or native protein reactions. Error bars show 95% CI, n=3. D) *In vitro* transaminations of PP_4108 incubated with different possible amino donors and 2KG as acceptor. Bars represent relative activity of enzyme standardized to L-2AA after 16-hour incubation. Error bars show standard deviation, n=2.

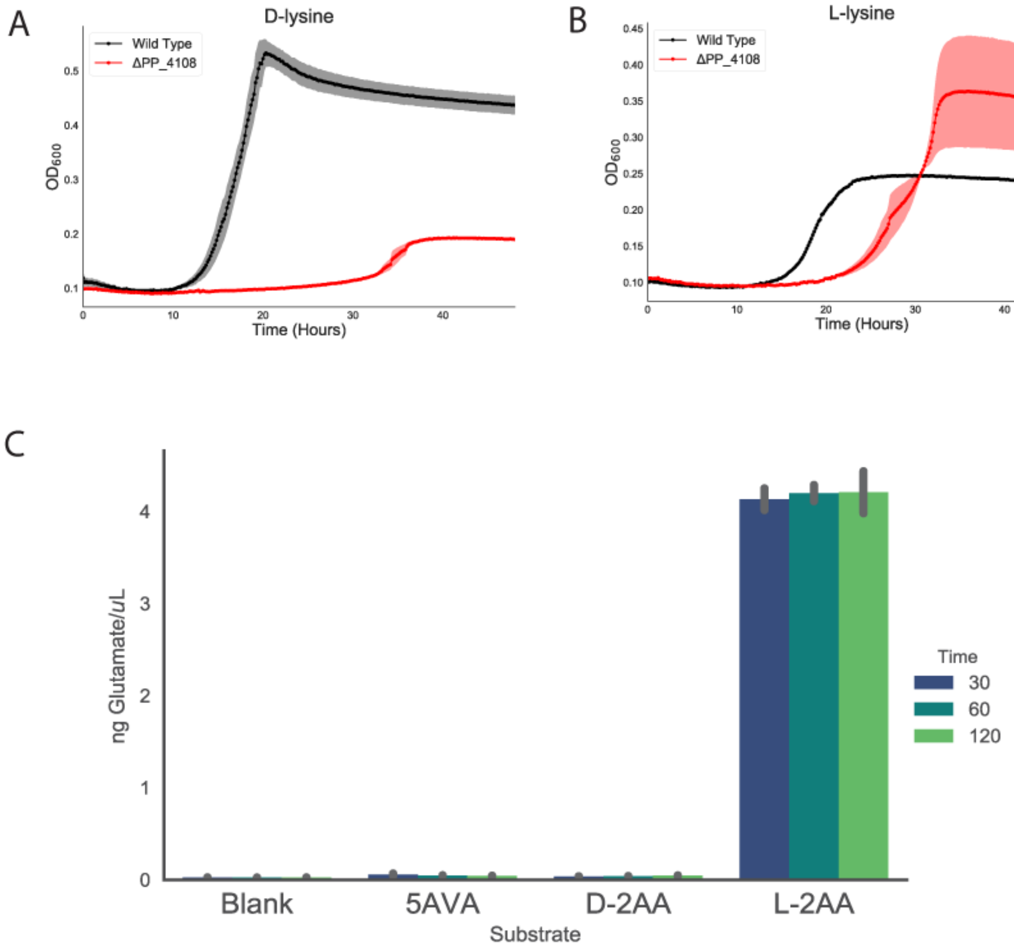


Figure 2-5 Growth of PP_4108 mutants on lysine. A) Growth of wild-type KT2440 and PP_4108 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. B) Growth of wild-type KT2440 and PP_4108 mutant on L-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. C) Colorimetric glutamate formation time course of PP_4108. Error bars are standard deviation of n=2.

2.4.3. PP_5260 is a novel DUF1338 family enzyme that catalyzes the conversion of 2OA to 2HG

Early work proposed 2OA is converted to 2KG via a 2HG intermediate [13,19], while later results suggested a direct conversion of 2OA to glutarate [11]. Either route likely requires a decarboxylation of 2OA, so we initially searched for decarboxylases within our dataset. Our fitness data on either lysine isomer revealed no obvious decarboxylases or enzymes likely to contain a thiamine pyrophosphate (TPP) cofactor, which are commonly employed by decarboxylases. However, a gene near other D-lysine catabolic genes in the *P. putida* genome, PP_5260, showed a significant fitness defect. A ΔPP_5260 strain was unable to grow on either isomer of lysine verifying its importance in lysine degradation (Figure 2-6a).

PP_5260 belongs to the DUF1338 protein family (http://pfam.xfam.org/family/PF07063). Although several unpublished crystal structures of DUF1338 domain containing proteins have been deposited into the Protein Data Bank, their biological function remains elusive. However, these structures combined with protein sequence alignments suggest a putative metal binding site is conserved throughout the DUF1338 family. As we hypothesized PP_5260 serves as the missing decarboxylase in D-lysine metabolism, we purified the enzyme for biochemical analysis. Enzymatic activity on 2OA was probed and analyzed via LC-TOF. After incubation of 2OA with PP_5260, we observed a ~92% (*p*=0.00034) reduction in the abundance of 2OA, whereas no 2OA was consumed in a boiled enzyme control or enzyme treated with EDTA confirming it to be a metalloenzyme (Figure 2-7a). Initially we believed the product would be either glutarate or glutarate semialdehyde, however neither of these was detected in the reaction. Early biochemical work suggested 2HG as a potential intermediate in pipecolate metabolism [19], and when the enzymatic product was compared to a racemic 2HG standard they shared the same mass, retention time and mass-to-charge ratio (Figure 2-7b), as well identical isotopic distributions of [M-H] peaks in the mass spectra (Figure 2-7c).

To identify the metal cofactor, the enzyme was dialyzed against EDTA to remove metals, and individual divalent metals were added back. Only the addition of Fe(II) restored enzymatic activity as measured by HPLC (Figure 2-6b). Subsequent reactions quenched after 5 minutes showed 200 uM of 2HG formed and 800 uM 2OA remaining, demonstrating 1:1 2OA to 2HG reaction stoichiometry (Figure 2-6c). Whether the product of the PP_5260 reaction is either L-2HG or D-2HG is critical to understanding the eventual fate of D-lysine, as *lghO* is specific for L-2HG (Figure 2-1a). An enzyme coupled assay specific for the detection of D-2HG was used to assess the stereochemistry of the PP_5260 product. Standard curves of D-2HG and L-2HG showed that the assay was only responsive to D-2HG (Figure 2-7d). The concentration of *in vitro* reactions of PP_5260 were then measured by both LC-TOF as well as the enzyme coupled assay, revealing all 2HG as the D-isomer (Figure 2-6d).

Kinetic parameters of PP_5260 were determined using an enzyme-coupled assay to spectrophotometrically measure CO_2 evolution via NADH oxidation [20]. PP_5260 displayed a V_{max} of 0.33 mM/min (+/- 0.08 mM), a K_m of 0.06 mM (+/- 0.03 mM), and a K_{cat} of 330 m^{-1} using 2OA as a substrate. Taken together these results reveal that PP_5260 is novel Fe(II) dependent decarboxylase that converts 2OA to D-2HG (Figure 2-6f), a chemical reaction not previously observed in nature.

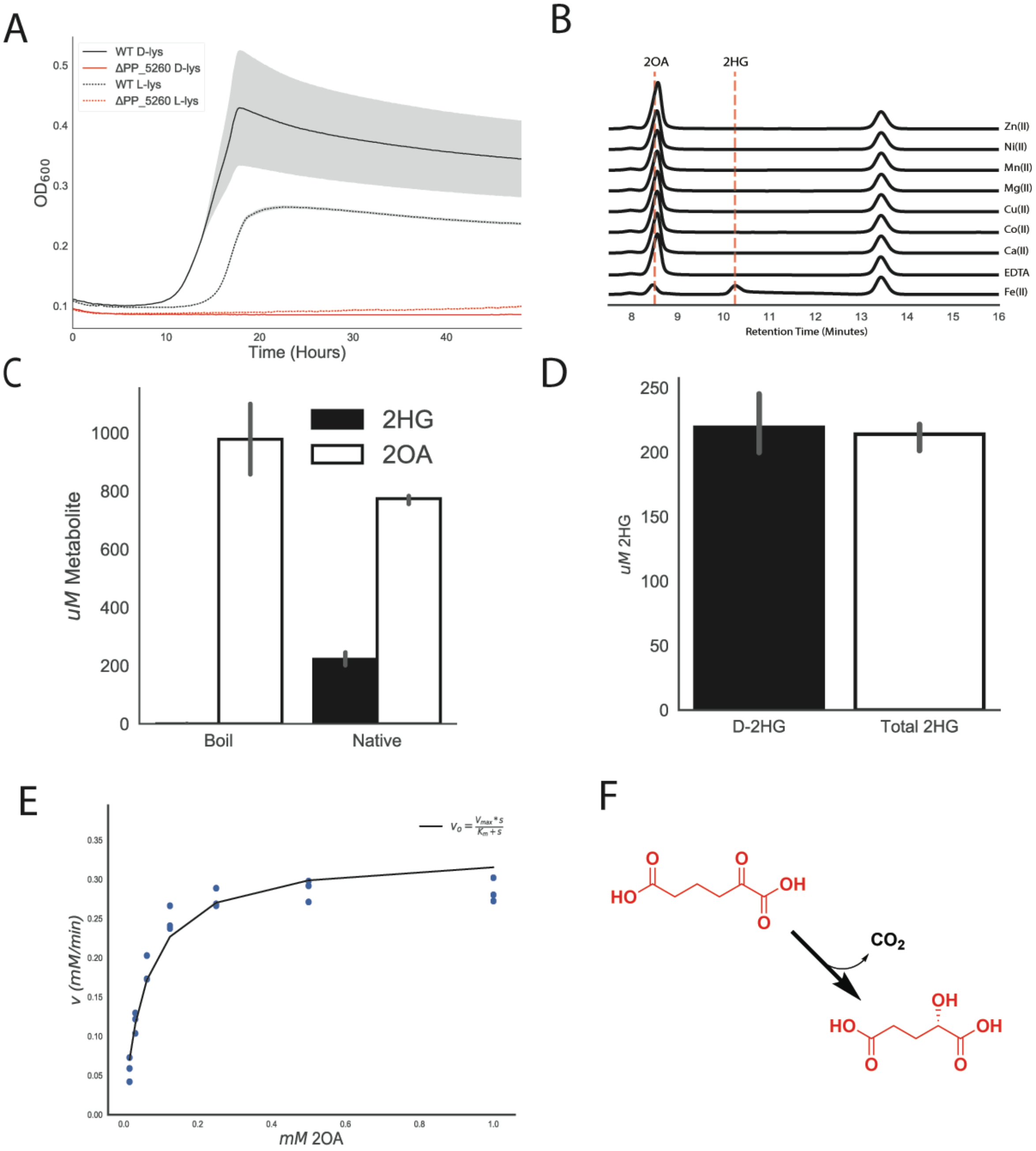


Figure 2-6 Identification of ydcJ (PP_5260) as a 2OA decarboxylase. Growth of wild-type (black) and PP_5260 mutant (red) on D-lysine (line) or L-lysine (dashed line) as a sole carbon source. Shaded area represents 95% CI, n=3. B) HPLC traces of *in vitro* reactions run with apo PP_5260 with exogenous metals added at 50 μM. Retention times for 2OA and 2HG are shown by vertical dashed lines. Metal or EDTA control is indicated to the

right of traces. C) *In vitro* assay of 2OA conversion to 2HG by purified PP_5260 protein analyzed via LC-TOF. 2OG in white, 2HG in black. D) *In vitro* assay of purified PP_5260 protein with 2OA as substrate. Black bar represents concentration of D-2HG measured by enzyme coupled assay. White bar represents total 2HG concentration as measured by LC-TOF. Error bars represent 95% CI, n=3. E) Initial velocity of reaction catalyzed by PP_5260 as a function of 2OA concentration. Blue dots represent individual measurements, while the black fit line shows a Michaelis-Menten fit. F) Chemical reaction catalyzed by PP_5260, 2OA is decarboxylated to D-2HG.

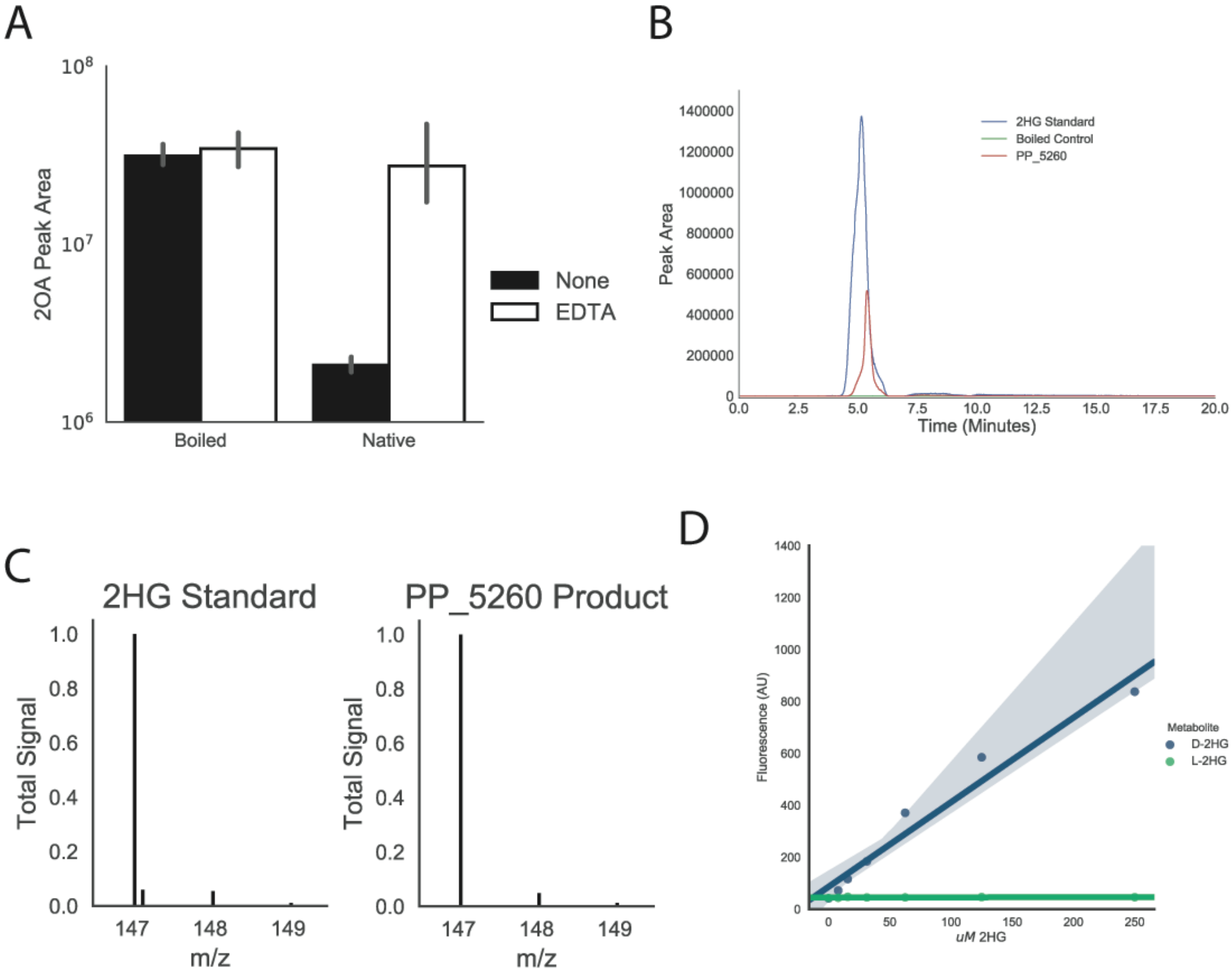


Figure 2-7 Characterization of PP_5260. A) In vitro reactions of boiled or native PP_5260 incubated with 2OA with 50 μM EDTA (white), or without EDTA (black). Bars represent 1og$_{10}$ transformed spectral counts, errors bars represent 95% CI, n=3 B) E LC-TOF analysis of a 2HG standard, and products of PP_5260 incubated with 2OA, and boiled control. C) Mass spectra of 2HG standard and product of PP_5260 in vitro reaction. D) Standard curves of D-2HG and L-2HG using a D-2HG specific enzymatic detection assay. Shaded areas represent 95% CI, n=3.

2.4.4. DUF1338 proteins are a widely distributed enzyme family with a putative conserved role in amino acid catabolism

After functional characterization of PP_5260, we use phylogenomics to propagate the annotation and further explore the biological role of DUF1338 proteins found in other organisms. We found that DUF1338 proteins are widely distributed across the tree of life, with homologs of PP_5260 found in plants and green algae (22), fungi, and bacteria, though they were not found in animals or archaea (Figure 2-8a). Homologs are widely distributed amongst bacteria, with the Firmicutes being a notable exception. PP_5260 homologs within the plant group Streptophyta, as well as bacterial groups Actinobacteria, Cyanobacteria, and Bacteroidetes formed monophyletic clades, while homologs from other taxonomic groups were not monophyletic (Figure 2-8a). DUF1338 homologs are found in bacteria important to biotechnology (*Corynebacterium glutamicum*), the environment (*Nostoc puncitforme*), and medicine (*Yersinia pestis*, *Mycobacterium tuberculosis*, *Burkholderia pseudomallei*).

Publicly available fitness data show both *Pseudomonas fluorescens* FW300-N2C3 and *Sinorhizobium meliloti* PP_5260 homologs have L-lysine specific defects when interrupted [15]. Genomic contexts within other bacteria suggest many DUF1338-containing enzymes may be involved in lysine or other amino acid metabolism (Figure 2-8b). Within the Actinobacteria DUF1338 homologs are often found adjacent to sarcosine oxidases, aldehyde dehydrogenases, and transaminases implying an additional catabolic amino acid function. In both the oleaginous bacterium *Rhodococcus opacus* B4 and *M. tuberculosis*, DUF1338 homologs are found next to predicted L-lysine aminotransferases, suggesting an ancestral homolog functioned in lysine catabolism. Interestingly, the *R. opacus* B4 genome has three DUF1338 homologs, only one of which contains genes predicted to be specific to lysine catabolism. The other two gene neighborhoods are similar in their functional content, mainly differing by containing an oxidoreductase or glycolate dehydrogenase, either of which may perform the same biochemical function. In Alphaproteobacteria, Betaproteobacteria, and Cyanobacteria, the presence of aldehyde dehydrogenases, oxidoreductases, glycolate dehydrogenases, and aminotransferases implies a metabolic function similar to PP_5260.

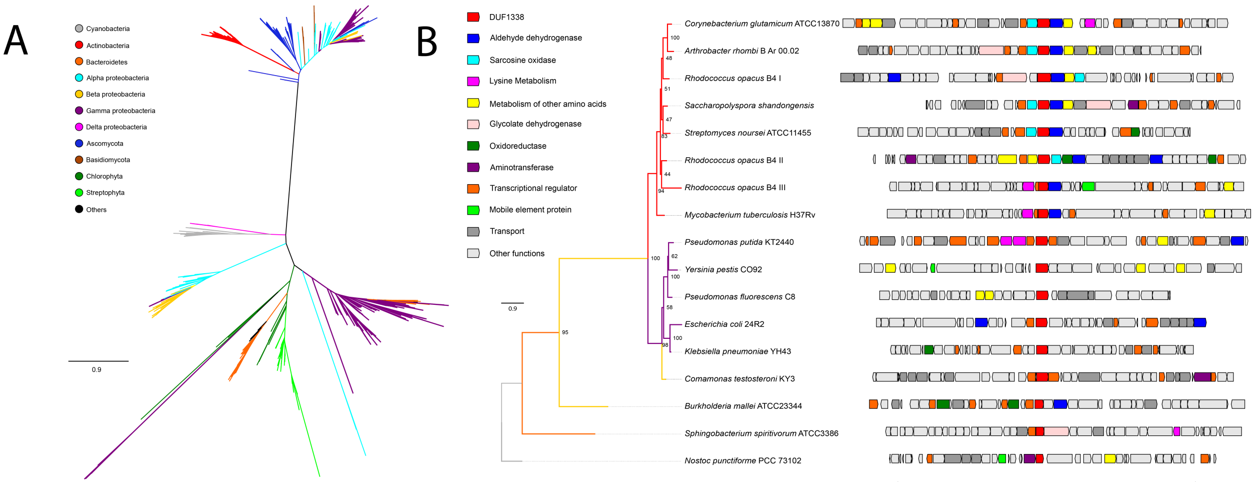


Figure 2-8 Phylogenomics of DUF1338 enzyme family. A) Phylogenetic relationships among DUF1338 homologs and their distribution among major phyla. Branches in the tree are colored by phylum. DUF1338 is found in most bacterial phyla as well as in plants and fungi. Non-monophyletic clades suggest pervasive horizontal gene transfer events in the family. B) Phylogenomics of selected DUF1338 homologs in bacteria. The phylogeny in the left shows the phylogenetic relationships between selected homologs, the branches have been colored according to their adscription to a given phylum and the support values are shown at the nodes. The boxes in the right represent the gene neighborhood for each homolog. The genes have been colored to represent their annotated functions.

.

2.4.5. PP_4493 putatively oxidizes D-2HG to 2KG and connects D-lysine to central metabolism

In the CoA independent route of glutarate metabolism, LghO oxidizes L-2HG to 2KG, however this enzyme is highly specific towards the L-2HG isomer and showed no fitness defect on D-lysine in our RB-TnSeq data (Figure 2-1a). A putative FAD-dependent and 4Fe-4S cluster-containing glycolate dehydrogenase, PP_4493, did show fitness defects on both D-lysine and L-lysine (fitness scores of -5.4 and -2.7 respectively) (Figure 2-2a). Glycolate dehydrogenases are members of a larger family of enzymes that oxidize the alcohol group of an alpha-hydroxyacid to their corresponding alpha-ketoacid (Figure 2-9a). Therefore, we hypothesized PP_4493 could potentially oxidize a similar 2-hydroxyacid, 2HG, to the corresponding alpha-ketoacid, 2KG. Moreover, many PP_5260 homologs were located next to or near putatively annotated glycolate dehydrogenases in other bacteria, underscoring their potential metabolic link (Figure 2-8b). To confirm these hypotheses, we again constructed a deletion strain, *P. putida* ΔPP_4493, which could not grow on D-lysine as a sole carbon source (Figure 2-9b), and showed attenuated growth on L-lysine (Figure 2-10). Furthermore, when grown on 10 mM glucose and 10 mM D-lysine the mutant accumulated ~500 uM 2HG normalized to optical density, whereas wild type *P. putida* did not accumulate any detectable 2HG (Figure 2-9c). Subsequent analysis of accumulated 2HG via a D-2HG specific detection kit revealed that this accumulated 2HG was indeed D-2HG (Figure 2-9c). These data and the conserved function and genomic context of glycolate dehydrogenases strongly suggest PP_4493 catalyzes the last step of L-2AA metabolism, oxidizing D-2HG to 2KG (Figure 2-1b).

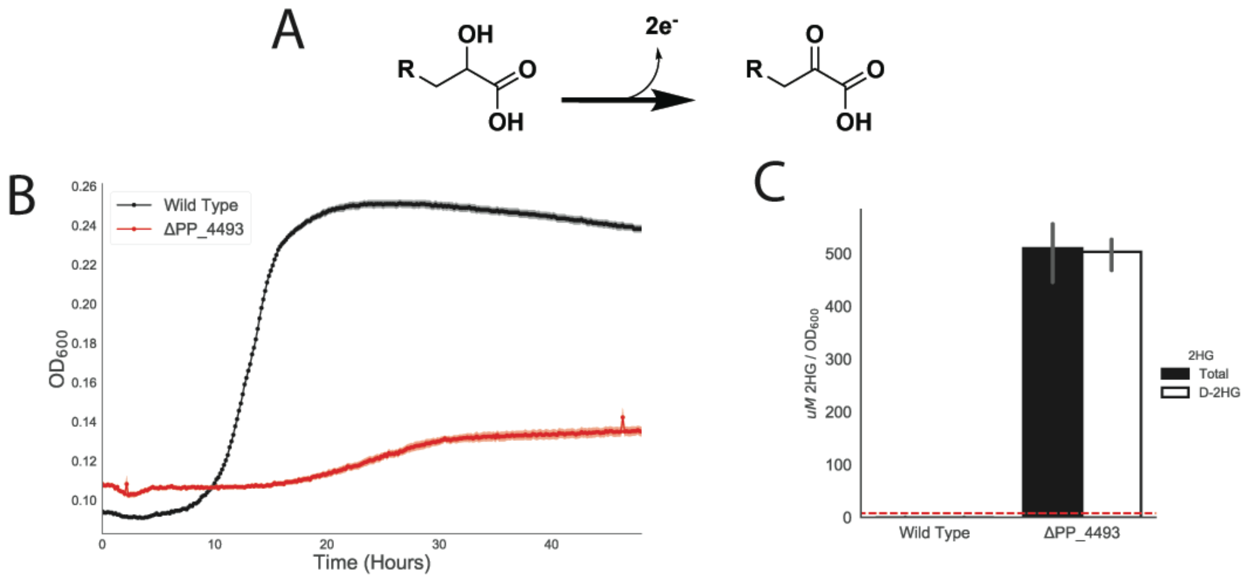


Figure 2-9 Identification of PP_4493 as a putative D-2HG dehydrogenase. General chemical reaction of a dehydrogenase converting a 2-hydroxyacid to a 2-ketoacid B) Growth of *P. putida* KT2440 and PP_4108 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. C) *In vivo* accumulation of 2HG in wild-type KT2440 and a PP_4108 mutant after 12 hours of growth on minimal medium supplemented with 10 mM glucose and 10 mM D-lysine. White bar represents concentration of D-2HG measured by enzyme coupled assay. Black bar represents total 2HG concentration as measured by LC-TOF. Red line represents limit of detection of enzyme coupled assay for D-2HG. Bars represent 1og$_{10}$ transformed spectral counts, error bars show 95% CI, n=3.

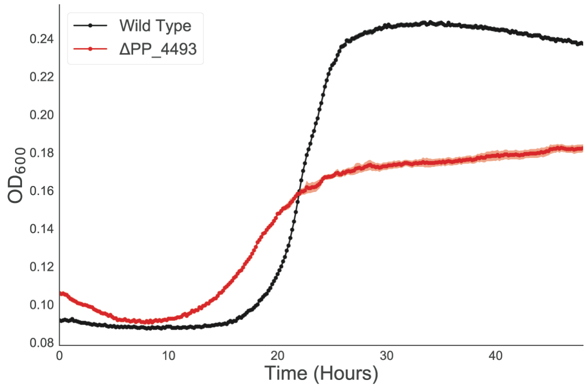


Figure 2-10 Growth of PP_4493 mutant on L-lysine. Growth of wild-type KT2440 and PP_4493 mutant on L-lysine as a sole carbon source. Shaded area represents 95% CI, n=3 .

2.4.6. CsiD is highly specific for glutarate hydroxylation but promiscuous in 2-oxoacid selectivity

During the initial preparation of this manuscript, Zhang et al. discovered a novel pathway of glutarate metabolism in *P. putida* [9]. They describe a cyclic reaction cascade wherein a novel 2KG-dependent non-heme Fe(II) oxygenase, PP_2909 (CsiD), hydroxylates glutarate to form 2HG and succinate using 2KG as a cosubstrate. PP_2910 (LghO), a putative L-2HG oxidase, then subsequently converts L-2HG to 2KG, regenerating the 2KG consumed in the initial reaction. These reactions result in the net incorporation of succinate into central metabolism (Figure S1). Our fitness results of the library grown on 5AVA also identified both *csiD* and *lghO*, in addition to the two enzymes from the CoA-dependent glutarate pathway, glutaryl-CoA ligase (*gcdG*) and glutaryl-CoA dehydrogenase (*gcdH*), mutants of which showed mild fitness defects when grown on 5AVA (Figure 2-11a). We also purified *csiD* and confirmed it hydroxylates glutarate in a 2KG-dependent manner (Figure 2-12a). HPLC analysis demonstrated that as glutarate was consumed, equimolar quantities of succinate and L-2HG were produced (Figure 2-12b). Additionally, a *csiD* deletion mutant showed increased lag time when grown on either L-lysine or 5AVA. By deleting the glutaryl-CoA ligase *gcdG*, and disrupting the CoA-dependent glutarate pathway, we completely prevented growth on 5AVA or L-lysine (Figure 2-12c). These results are in agreement with those found with Zhang et al [9].

Because non-heme Fe(II) oxidases can be promiscuous with respect to the 2-oxoacid cosubstrate [21,22], we evaluated the 2-oxoacid specificity of CsiD. First, we evaluated the hydroxyl acceptor substrate specificity of CsiD family proteins by purifying two additional homologs from *E. coli* and a halophilic bacterium, *Halobacillus* sp. BAB-2008 (Figure 2-11b). We probed the activity of the homologs against a panel of 3 to 6 carbon fatty acids and diacids in the presence of 2KG, and found only glutarate served as a hydroxylation substrate (Figure 2-11c). These results are consistent with the work recently reported by Zhang *et al* [9] and further suggests the specificity of CsiD homologs is conserved across phyla. Although extremely specific for the hydroxylation substrate, all three CsiD homologs could utilize both 2OA and 2KG, but not oxaloacetate, as a cosubstrate for L-2HG formation (Figure 2-11d). The coproduct of the reaction using 2OA as a 2-oxoacid would be glutarate, rather than succinate. This result is particularly interesting as it provides a possible mechanism for the previously observed metabolic link between D-lysine and L-lysine catabolism. Growth defects observed in a ΔPP_2909 ΔPP_0158 double mutant grown on D-lysine also support this hypothesis (Figure 2-13a).

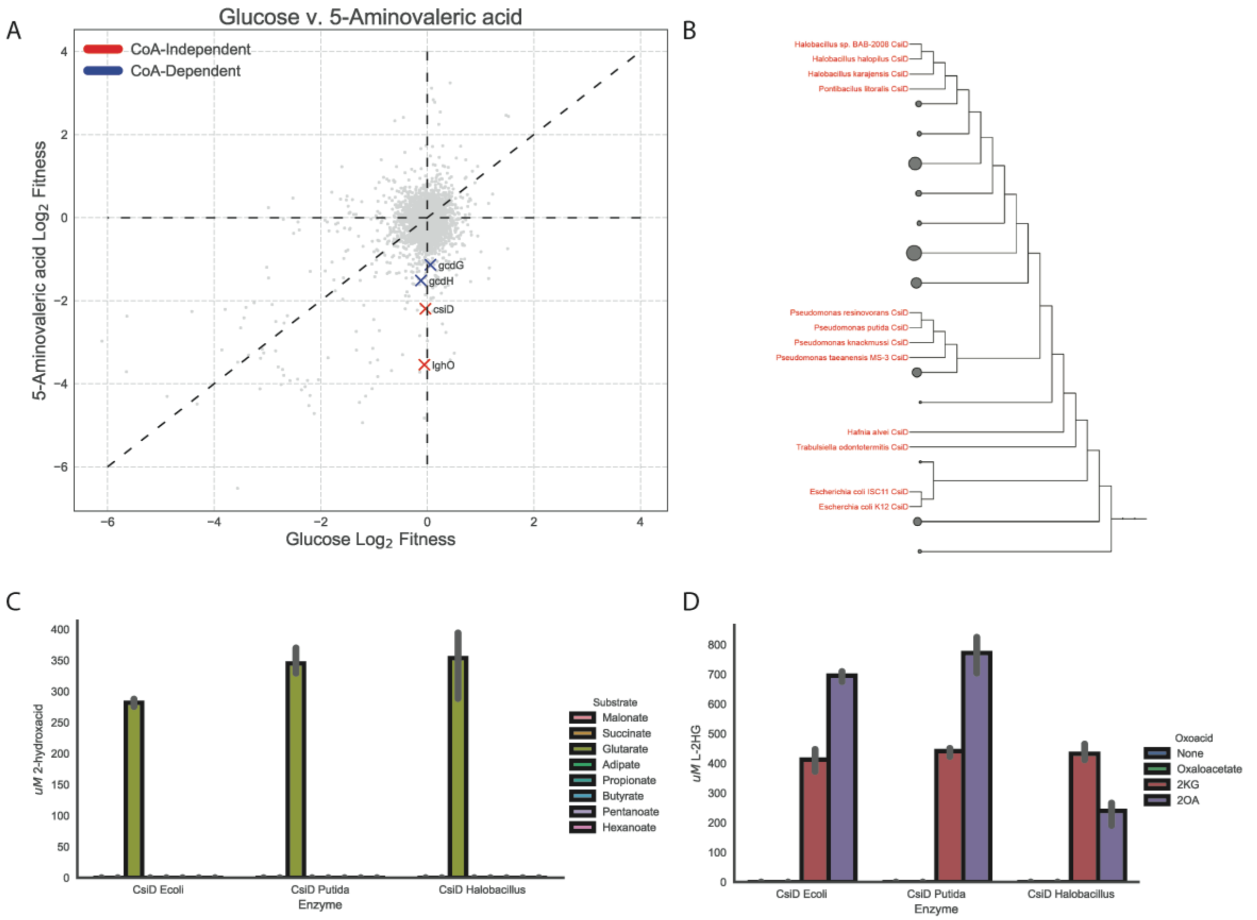


Figure 2-11 Role of CsiD in *P. putida lysine metabolism.* A) Plot of genome wide fitness values of libraries grown on either 5AVA or glucose. CoA-dependent glutarate degradation genes are shown in red, while those known to be involved succinate producing metabolism are shown in blue. B) Phylogenetic tree of bacterial CsiD homologs. Homologs used in *in vitro* assays are highlighted in red. C) *in vitro* reactions of CsiD with different substrates using 2KG as a 2-oxoacid. Bars show peak area of 2-hydroxyacid, error bars show 95% CI, n=3. D) *In vitro* reactions of CsiD homologs with different 2-oxoacids. Bars represent spectral counts of L-2HG. Error bars show 95% CI, n=3.

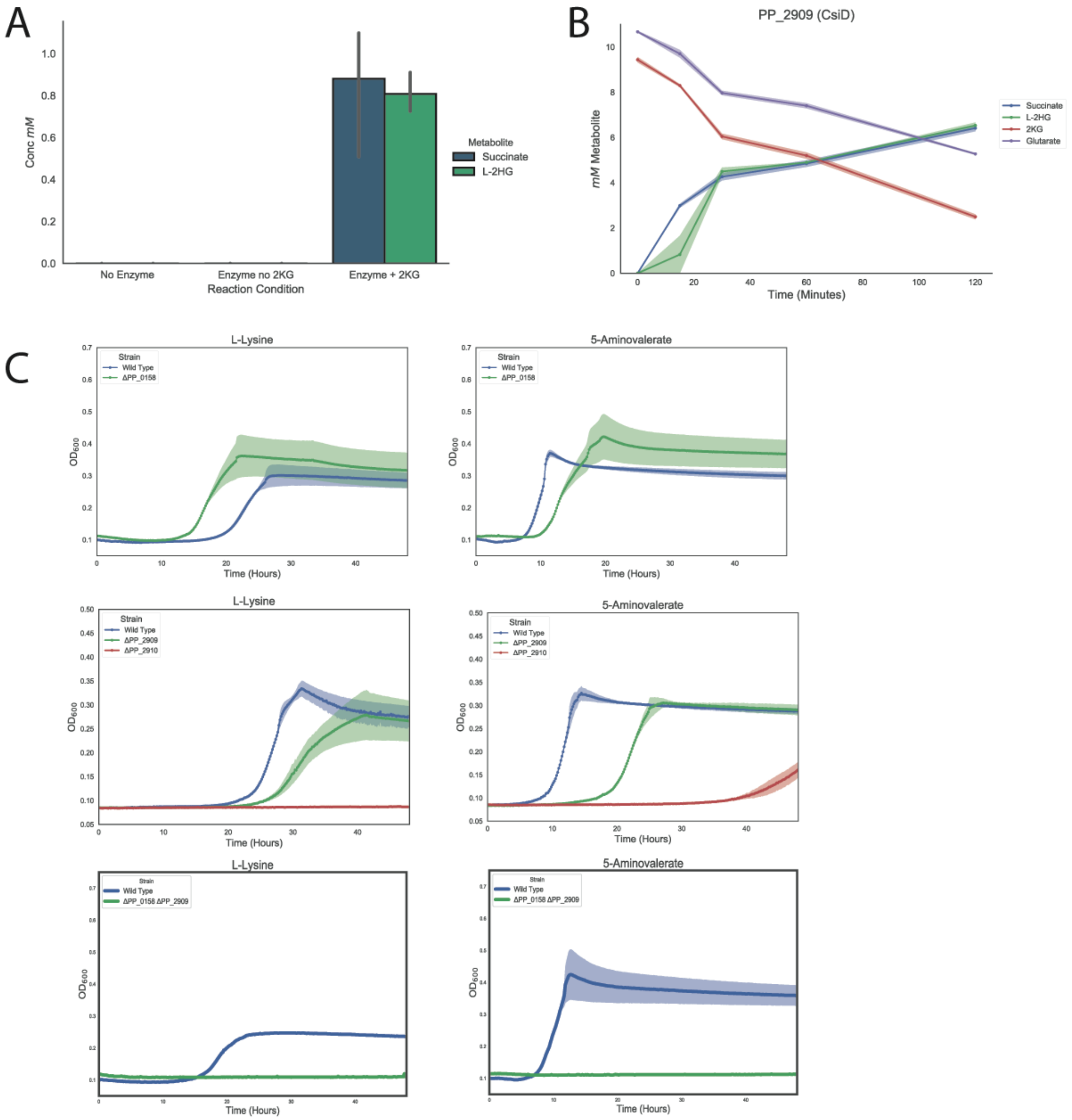


Figure 2-12 Characterization of PP_2909 (CsiD) in *P. putida*.In vitro reactions of PP_2909. Bars show mM of either succinate (blue), or 2HG (green) formed by boiled enzyme control, no 2KG control, or native enzyme with 2KG added. Errors bars show 95% CI, n=3. B) Time course in vitro reaction of PP_2909. Plot shows 2HG, 2KG, succinate, and glutarate overtime. Shaded region shows 95% CI, n=3. C) Growth curves of wild-type KT2440, PP_0158, PP_2909, PP_2910, or PP_2909/PP_0158 double mutants grown on either L-lysine (left column), or 5AVA (right column). Shaded region shows 95% CI, n=3.

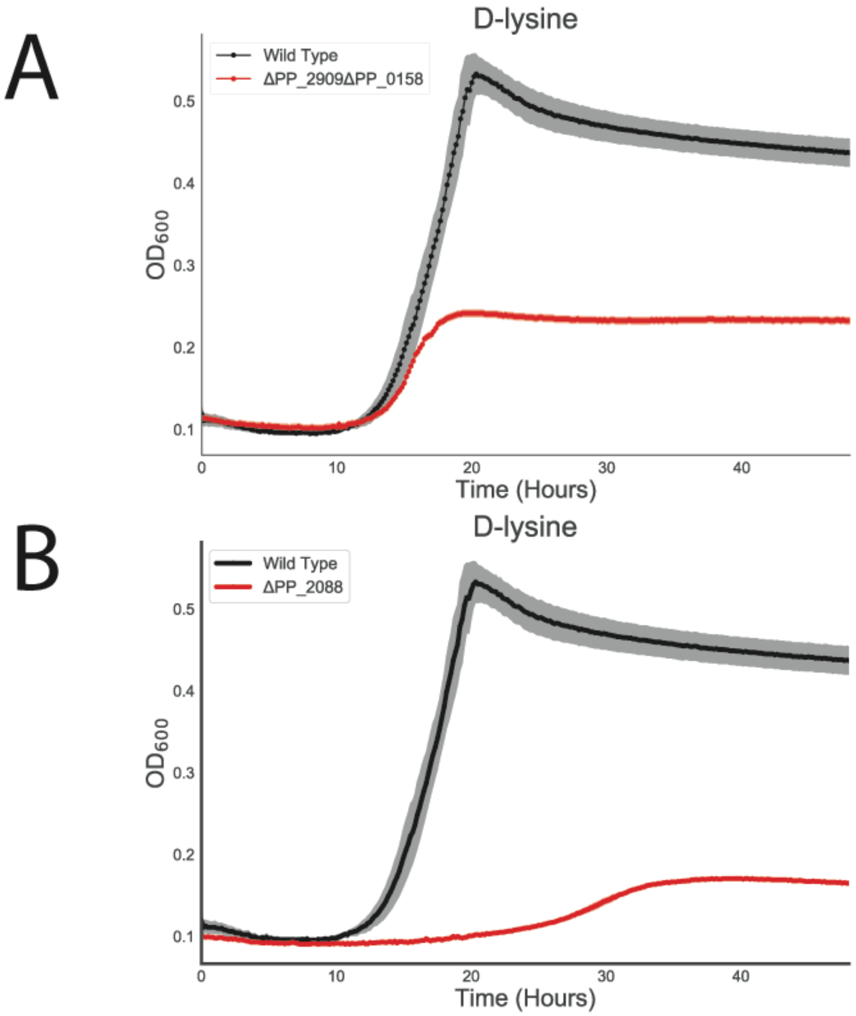


Figure 2-13 Growth of PP_2909/PP_0158 and PP_2088 mutants on D-lysine. A) Growth of wild-type KT2440 and PP_2909/PP_0158 mutants on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3. A) Growth of wild-type KT2440 and PP_2088 mutant on D-lysine as a sole carbon source. Shaded area represents 95% CI, n=3.

2.4.7. Expression of lysine metabolic proteins is responsive to pathway metabolites

Multiple studies have demonstrated the expression of lysine catabolic genes is upregulated in the presence of pathway metabolites [9,12,23]. To investigate the regulation of the newly-discovered lysine catabolic enzymes from this study, wild-type *P. putida* KT2440 was grown in minimal media on glucose or a single lysine metabolite (e.g. D-lysine, L-lysine, 5AVA, 2AA, or glutarate) as a sole carbon source until cultures reached an OD_{600} of 1.0. We then quantified the relative abundance of D- and L-lysine catabolic proteins via targeted proteomics (Figure 2-14). For each protein, all pairwise statistical comparisons of different carbon sources can be found in Supplemental Table 2. All five D-lysine pathway proteins measured (AmaA (PP_5257), AmaB (PP_5258), PP_4108, YdcJ (PP_5260), and YdiJ (PP_4493)) were upregulated when grown on L-lysine, D-lysine or 2AA compared to the glucose control. Neither 5AVA nor glutarate significantly induced expression of any measured D-lysine proteins. Of all the targeted proteins, the three identified in this study that directly degrade 2AA were most strongly induced by 2AA. Somewhat surprisingly, we also found the two enzymes involved in 2AA formation, AmaA and AmaB, were also more highly expressed in the presence of 2AA suggesting the possible involvement of a global regulator. An interesting finding from our initial screen showed sigma factor RpoX (PP_2088) to be required for fitness on D-lysine (Figure 2-2a). Deletion mutants of *rpoX* were severely attenuated in their ability to grown on D-lysine as a sole carbon source (Figure 2-13b). Further work will be necessary to examine complex regulatory network that controls D-lysine metabolism.

The initial two enzymes from L-lysine metabolism, DavA and DavB, were most highly expressed in the presence of L-lysine, but also significantly with D-lysine. As previously observed, DavT and DavD were most strongly upregulated on 5AVA, moderately upregulated on L-lysine, and to a lesser extent D-lysine. The induction of LhgO and CsiD was highest when grown on glutarate, although these proteins were also moderately upregulated by 5AVA and L-lysine. By comparison, PP_0159 (GcdG) expression in the presence of glutarate was stimulated to a lesser extent than LhgO and CsiD expression; in addition, GcdG was slightly upregulated on 5AVA and L-lysine.

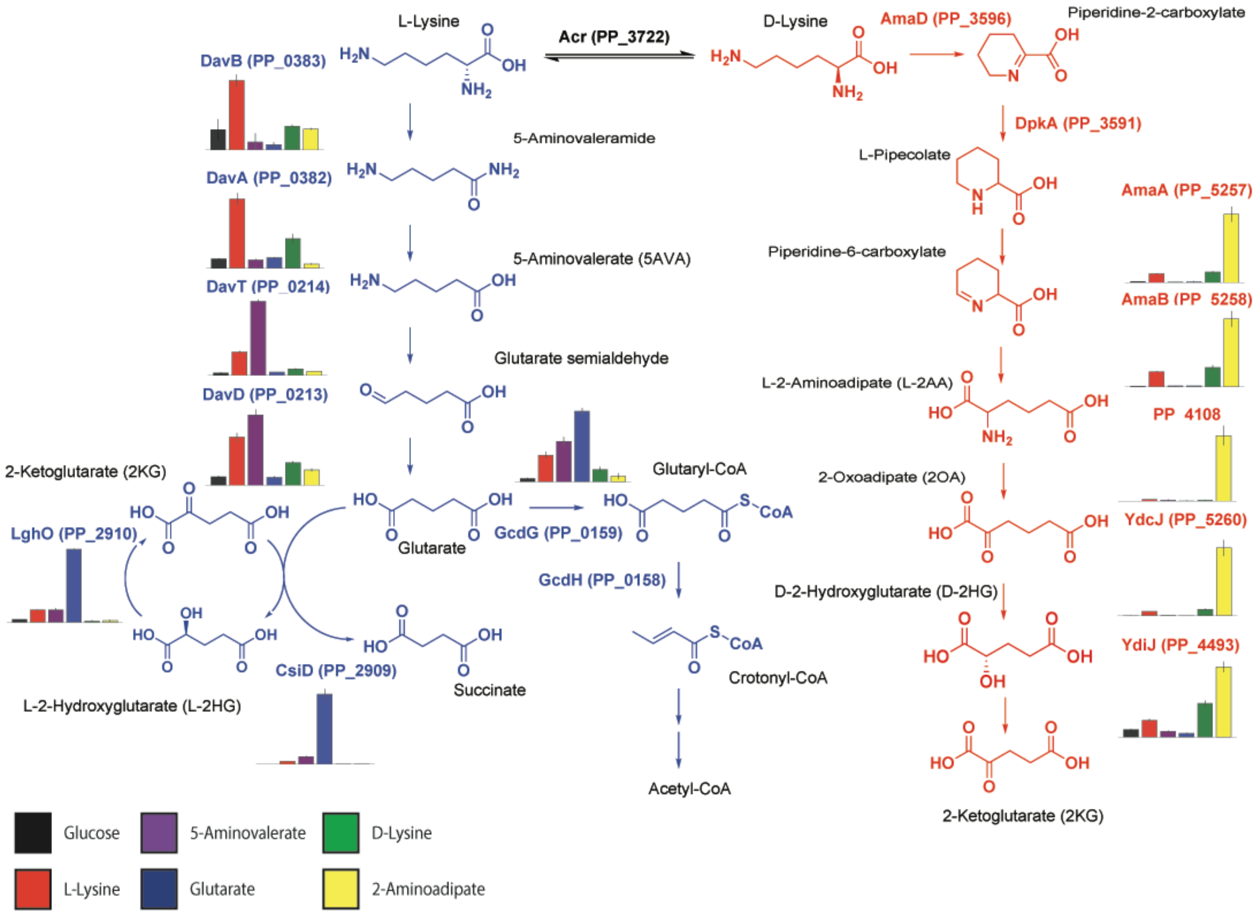


Figure 2-14 Expression of lysine degradation pathways in response to different lysine metabolites. Relative abundance of selected lysine degradation enzymes expressed in wild-type KT2440 in response to different carbon sources. Bars show spectral counts of

proteins after 36 hours of growth on 10 mM glucose (black), 5AVA (purple), D-lysine (green), L-lysine (red), glutarate (blue), or 2AA (yellow). Error bars show 95% CI, n=3.

2.5. Discussion

Despite intensive study, a complete biochemical and genetic understanding of D-lysine catabolism in *P. putida* has remained elusive. A 2OA degradation pathway has been extensively characterized in mammals, because of its implications in human disease [24]. In the mammalian pathway, L-lysine is metabolized to 2OA and eventually converted to acetyl-CoA via a glutaryl-CoA intermediate [24]. However, this pathway has not been observed in bacteria. Previous work suggested 2OA is either converted via decarboxylation to glutarate or through several enzymatic steps to 2HG, yet none of these studies conclusively demonstrated a genetic and biochemical basis for these hypotheses [11,19,25]. In this work we demonstrate plausible biochemical routes to account for both of these previously hypothesized pathways.

The first route, catalyzed by the DUF1338-containing metalloenzyme PP_5260, involves the direct conversion of 2OA to D-2HG. The formation of the D-2HG isomer by PP_5260 maintains stereochemical separation from the L-2HG formed by L-lysine degradation, thus requiring the dehydrogenase PP_4493 rather than the L-2HG specific oxidase *lghO*. This transformation seemingly involves two separate reactions: a decarboxylation and a hydroxylation. Hydroxymandelate synthase has been shown to catalyze a similar enzymatic reaction, via an intramolecular oxidative decarboxylation similar to 2KG dependent Fe(II) oxidases [26]. PP_5260 is also a Fe(II) dependent decarboxylase, and the two share similar Kcat values for their given substrates (330 m^{-1} for PP_5260 and 270 m^{-1} for hydroxymandelate synthase) [27]. Though PP_5260 and hydroxymandelate synthase share little sequence homology, this enzyme may give us insight into the molecular mechanism of DUF1338 enzymes. We have given PP_5260 the tentative title of 2-hydroxyglutarate synthase (*hglS*) until further mechanistic studies (currently underway in our group) are completed and a more accurate enzyme name can be assigned.

In bacteria, homologs of PP_5260 appear widely distributed with their genomic contexts suggesting functions both within and beyond lysine metabolism. Genomic contexts in other bacteria, particularly *Actinobacteria*, suggest these homologs may be involved in other amino acid catabolic pathways. Unfortunately, there is scant evidence for homologous function in model organisms. For example, although DUF1338 proteins are present in other *Ascomycota*, there is no homolog in *Saccharomyces cerevisiae*. Interestingly, the *E. coli* homolog of PP_5260 is located next to a potential glucans biosynthesis gene: Glucans biosynthesis protein D [28]. Another DUF1338-containing protein from rice has been characterized and was implicated in starch granule formation [29]. These results suggest DUF1338 proteins could play a role in sugar metabolism.

Recently, Zhang *et al*. thoroughly characterized a CoA independent glutarate catabolism route ending at succinate involving the Fe(II) dependent oxygenase CsiD [9]. Our RB-TnSeq screening convergently uncovered this pathway, and our biochemical and physiological results fully corroborate their findings. While both works show multiple CsiD homologs from divergent bacteria are highly specific towards glutarate as a hydroxyl acceptor, all three homologs we tested showed promiscuous activity toward 2-oxoacid cosubstrates. The ability of the *P. putida* CsiD to utilize 2OA as a cosubstrate is particularly interesting as it may directly connect L- and D-lysine metabolism. The promiscuity of CsiD may explain earlier studies which reported glutarate formation from D-lysine [11]. Further studies involving labelled substrates may help elucidate the potential link between the two pathways. While CsiD plays a clear role in L-lysine metabolism in *P. putida*, its role in other organisms remains a mystery. In *E. coli*, RpoS controls the expression of CsiD, but *rpoS* mutants showed fitness benefits on all three lysine metabolites tested in our RB-TnSeq data [30]. Recent work has shown that *E. coli* also uses CsiD to metabolize lysine, suggesting a possible conserved role for this pathway across bacteria [31].

Work presented here and previous reports have shown expression of both lysine catabolism pathways are highly responsive to their respective metabolites. While this metabolism appears highly coordinated, the genes themselves are dispersed across the genome, with both PP_4018, and PP_4493 found outside of operons, and relatively distant from other lysine catabolic genes. At least two global regulators appeared to be important to lysine metabolism based on our Rb-TnSeq data, *cbrA* (PP_4695) and *rpoX* (PP_2088). The two-component system CbrAB has been implicated in catabolite repression and C/N balance in *P. aeruginosa*, with mutants unable to grown on multiple amino acids [32]. Further work in *P. putida* KT2440 showed the CbrAB system behaved similarly to that in *P. aeruginosa* [33]. It would be unsurprising if this regulator controls the expression of various genes within lysine catabolism; more work into uncovering the regulon is warranted. RpoX on the other hand has been implicated in osmotic tolerance in *P. aeruginosa* [34,35]. This is interesting as lysine metabolism, and specifically pipecolate metabolism, has been associated with osmotic tolerance across multiple bacteria [36]. As a *rpoX* deletion mutant was unable to grow on D-lysine, these results suggest D-lysine metabolism of *P. putida* may be involved in adaptation to saline or other osmotically stressful environments.

An interesting question remains as to why *P. putida* maintains separate metabolic pathways for D- and L-lysine, and why L-lysine metabolism seems dependent on the presence of an intact D-lysine metabolism. Previously work has proposed that the D-lysine pathway may provide a way of resolving a C/N imbalance that may occur when L-lysine is metabolized. However we believe this is unlikely as both lysine degradation pathways contain one deamination and one transamination reaction [11]. Our fitness results indicate D-lysine metabolism is dispensable for growing on 5AVA. This would suggest only the initial two steps of L-lysine metabolism, the oxidation of lysine to 5-aminopentanamide by DavB and its subsequent deamination to 5AVA by DavA are dependent on D-lysine catabolism. We propose the adjacent AsnC family regulator PP_0384 likely responds to L-lysine as many proteins within this family respond to amino acids including lysine [37,38] and expression of these two enzymes is most responsive to L-lysine. To our knowledge there has been no rigorous characterization of the regulation of the *davAB* operon, nor of the biochemical activities of these two enzymes *in vitro*. Future studies to uncover the mechanistic regulation at the transcriptional and post-translational levels at these two steps may uncover the necessity of D-lysine dependency of the L-lysine catabolic pathway. Overall our work highlights the utility of global fitness profiling to discover novel, complex, metabolic pathways in even well-characterized bacteria.

2.6. Materials and methods

2.6.1. Chemical sources and media

Routine bacterial cultures were grown in in Luria-Bertani (LB) Miller medium (BD Biosciences, USA). *E. coli* was grown at 37 °C, while *P. putida* was grown at 30 °C unless otherwise noted. When indicated, *P. putida* was grown on modified MOPS minimal medium [39]. Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), gentamicin (30 mg/L, Fisher Scientific, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA), when indicated. D-2AA was purchased from Takara Bioscience (USA), all other compounds were purchased through Sigma Aldrich.

2.6.2. Strains and plasmids

All bacterial strains and plasmids used in this work are listed in Table 2-1. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/391). All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software [40–42]. Synthetic DNA coding for the *Halobacillus* sp. BAB-2008 *csiD* homolog was purchased from Integrated DNA Technologies (IDT, Coralville, IA). Plasmids were assembled via Gibson Assembly using standard protocols [43], or Golden Gate Assembly using standard protocols [44]. Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA).

Table 2-1 Plasmid and Strains.

Strain	**JBEI Part ID**	Reference	Genotype
E. coli DH10B			
E. coli S17			
E. coli BL21(DE3)			
P. putida KT2440			
P. putida ΔPP_0158	JPUB_010967	This work	
P. putida ΔPP_2909	JPUB_010968	This work	
P. putida ΔPP_2910	JPUB_010969	This work	
P. putida ΔPP_0158 ΔPP_2909	JPUB_010970	This work	
P. putida ΔPP_4108	JPUB_010971	This work	
P. putida ΔPP_4493	JPUB_010972	This work	
P. putida ΔPP_5260	JPUB_010973	This work	
Plasmids			
pMQ30		Shanks et al. 2009	Gm, SacB
pET28			Kan

pET21b			Amp
pMQ30-PP_0158	JPUB_010989	This work	Gm, SacB
pMQ30-PP_2909	JPUB_010991	This work	Gm, SacB
pMQ30-PP_2910	JPUB_010995	This work	Gm, SacB
pMQ30-PP_4108	JPUB_010981	This work	Gm, SacB
pMQ30-PP_4493	JPUB_010979	This work	Gm, SacB
pMQ30-PP_5260	JPUB_010977	This work	Gm, SacB
pET28-CsiD_Halo	JPUB_010987	This work	Kan
pET28-CsiD_Ecoli	JPUB_010993	This work	Kan
pET28-CsiD_Pput	JPUB_010975	This work	Kan
pET21b-PP_4108	JPUB_010983	This work	Amp
pET21b-PP_5260	JPUB_010985	This work	Amp

2.6.3. Random barcode RB-TnSeq experiments

P. putida RB-TnSeq library JBEI-1 was created by diluting a 1 mL aliquot of the previously described *P. putida* RB-TnSeq library [16] in 500 mL of LB media supplemented with kanamycin which was then grown to an OD_{600} of 0.5 and frozen as 1 mL aliquots after adding glycerol to a final concentration of 20% v/v. Libraries were stored at −80 °C until used. A 1 mL aliquot of *P. putida* RB-TnSeq library JBEI-1 was thawed on ice and diluted in 25 mL of LB supplemented with kanamycin. The culture was grown until it reached an OD_{600} of 0.5, at which point 3 1-mL aliquots were taken, pelleted, decanted, and then stored at −80 °C to use as a time zero control. The library was then washed once in MOPS minimal medium without any carbon source, and then diluted 1:50 into 10 mL MOPS minimal medium supplemented with either 10 mM glucose, 5AVA, D-lysine, or L-lysine. Cells were grown in 50 mL culture tubes for 48 hours at 30 °C shaking at 200 rpm. After growth 2 ml aliquots from the culture tubes were pelleted, decanted and frozen at −80 °C for barcode sequencing. We performed DNA barcode sequencing (BarSeq) as previously described [14,16]. The fitness of a strain is the normalized log2 ratio of barcode reads in the experimental sample to barcode reads in the time zero sample. The fitness of a gene is the weighted average of the strain fitness for insertions in the central 10–90% of the gene. The gene fitness values are normalized so the typical gene has a fitness of zero. The primary statistic *t*-value is of the form of fitness divided by the estimated variance across different mutants of the same gene. Statistic *t*-values of $>|4|$ were considered significant. All experiments described herein pass the quality metrics described previously unless noted otherwise. All fitness data in this work is publicly available at http://fit.genomics.lbl.gov.

2.6.4. Construction of deletion mutants

Deletion mutants in *P. putida* were constructed by homologous recombination and *sacB* counterselection using the allelic exchange vector pMQ30 [45]. Briefly, homology fragments ranging from 1 kbp to 2 kbp up- and downstream of the target gene, including the start and stop codons respectively, were cloned into pMQ30. An exception to these design parameters was plasmid pMQ-PP_5260 which maintained an additional 21 nt at the 5′ end of the gene in addition to the stop codon. Plasmids were then transformed via electroporation into *E. coli* S17 and then mated into *P. putida* via conjugation. Transconjugants were selected for on LB agar plates supplemented with gentamicin 30 mg/mL, and chloramphenicol 30 mg/mL. Transconjugants were then grown overnight on LB media also supplemented with 30 mg/mL gentamicin, and 30 mg/mL chloramphenicol, and then plated on LB agar with no NaCl supplemented with 10% w/v sucrose. Putative deletions were restreaked on LB agar with no NaCl supplemented with 10% w/v sucrose, and then were screened via PCR with primers flanking the target gene to confirm gene deletion.

2.6.5. Plate based growth assays

Growth studies of bacterial strains were conducted a 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 media 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 was 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 and all OD_{600} measurements are reported without pathlength corrections.

2.6.6. Expression and purification of proteins

A 5 mL overnight culture of *E. coli* BL21 (DE3) containing the expression plasmid was used to inoculate a 500 mL culture of LB. Cells were grown at 37 °C to an OD_{600} of 0.6 then induced with isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The temperature was lowered to 30 °C and cells were allowed to express for 18 hours before being harvested via centrifugation. Cell pellets were stored at -80 °C until purification. For purification, cell pellets were 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 40,000xg). 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 step-wise gradient of lysis buffer containing increasing imidazole concentrations (100 mM, 200 mM, and 400 mM). Fractions were collected and analyzed via SDS-PAGE. Purified protein was dialyzed overnight at 4 ˚C against 50 mM HEPES pH 7.5, 5% glycerol.

2.6.7. CsiD *in vitro* assays

The activity of purified CsiD homologs was analyzed in 100 μL reaction mixtures containing 50 mM HEPES (pH 7), 5 mM glutarate, 5 mM 2KG, 25 μM $FeSO_4$, 0.1 mM ascorbate, and 0.5 mM dithiothreitol. For negative control reactions, each respective reaction component was omitted. To initiate reactions, CsiD was added to a final concentration of 7 μM. For the no enzyme control, CsiD was denatured at 98 °C for 10 minutes prior to addition to the reaction mix. Reactions were allowed to proceed at 22 °C for 3 hours. Products were analyzed via LC-TOF method 3 after quenching via the addition of acetonitrile and methanol for a final ACN:H_2O:MeOH ratio of 6:3:1 To analyze products from substrate range as well 2-oxoacid specificity experiments, reactions were measured via LC-TOF method 1.

2.6.8. Transamination assays

To determine product formation via PP_4108, assays were conducted in 50 mM HEPES (pH 7.5), with 5 mM 2KG, 0.1 mM PLP, and 5 mM of substrate, and 10 μM of purified enzyme or boiled enzyme control in 100μL volumes. Reactions were incubated at 30 °C for 16 hours and quenched via the addition of ACN and MeOH for a final ACN:H_2O:MeOH ratio of 6:3:1 for LC-TOF method 3. To determine substrate specificity reactions were set up at 75 μL scale and carried out at 30°C for up to 16 hours before freezing. For analysis, reactions were diluted 15-fold in water and assessed by a colorimetric assay for glutamate (Sigma MAK004) in 96-well format via a SpectraMax M4 plate reader (Molecular Devices, USA).

2.6.9. PP_5260 *in vitro* assays

The activity of PP_5260 was initially assessed in 50 mM HEPES, with 5 mM 2OA as substrate and 10 μM purified enzyme or boiled enzyme control. Reactions were incubated for 16 hours at 30 °C. To test the necessity of metal cofactors EDTA was added to a final concentration of 50 μM. Reactions and quenched via the addition of ACN and methanol MeOH for a final ACN:H_2O:MeOH ratio of 6:3:1 for LC-TOF analysis method 3, or with an equal volume of ice-cold methanol for HPLC analysis and LC-TOF method 2.

To determine the metal cofactor, after purification over Ni-NTA resin the protein was concentrated and dialyzed overnight against 50mM HEPES, 100mM NaCl, pH 7.5. To generate apo-enzyme, the protein was then dialyzed four times at a protein:dialysis buffer ratio of 1:300 against the same buffer containing 5mM EDTA in order to remove any bound metal. The enzyme was dialyzed once more against buffer without EDTA overnight in order to remove any remaining chelating reagent. The apo-enzyme was then assayed in the presence of 50 μM of a variety of potential metal cofactors in 50 mM HEPES with 10 mM 2OA as substrate and 10 μM purified enzyme. Reactions were incubated for 30 minutes at 30 °C and activity was assessed via HPLC.

Determination of enzyme stoichiometry was assessed in 50 mM HEPES, 50 μM $FeCl_2$, with 1 mM 2OA as substrate and 0.1 μM purified enzyme or boiled enzyme control. Reactions were incubated for 5 minutes at 30 °C and then quenched with an equal volume of ice-cold methanol then quantified with LC-TOF method 2.
Enzyme coupled decarboxylation assays were carried out as previously described [20]. Reaction mixtures contained 100 mM Tris-HCl (pH 7), 10 mM $MgCl_2$, 0.4 mM NADH, 4 mM, 50 μM $FeCl_2$, phosphoenol pyruvate (PEP), 100U/mL pig heart malate dehydrogenase(Roche), 2U/mL microbial PEP carboxylase (Sigma), and 10 mM 2OA. Reactions were initiated by the addition of purified PP_5260 or boiled enzyme controls, and absorbance at 340 nm was measured via a SpectraMax M4 plate reader (Molecular

Devices, USA). Michaelis-Menten behavior was formulated as Vo= Vmax*[S]*(1+e)Km+[S]where *e* is a multiplicative error term (assumes scaling error with magnitude of reaction velocities) [46]. Least-squares minimization was used to derive K_m and K_{cat}. Determination of D-2HG concentration was assayed with a D-2-Hydroxyglutarate (D-2HG) Assay Kit (Sigma MAK320).

2.6.10. HPLC analysis

HPLC analysis was performed on an Agilent Technologies 1200 series liquid chromatography instrument coupled to a refractive index detector (35°C, Agilent Technologies, Santa Clara, CA). Samples were injected onto an Aminex HPX-87H Ion Exclusion Column (300 x 7.8 mm, 60°C, Bio-Rad, Hercules, CA) and eluted isocratically with 4 mM H_2SO_4 at 600 uL/min for 20 minutes. Compounds were quantified via comparison to a calibration curve prepared with authentic standards and normalized to injection volume.

2.6.11. Proteomics analysis

P. putida KT2440 wild type was grown on MOPS minimal media with 10 mM of either glucose, L-lysine, D-lysine, 5AVA, 2AA, or glutarate. Cells were harvested when cultures reached an OD_{600} of 1.0 with a 1 cm pathlength. Cell lysis and protein precipitation were achieved by using a chloroform-methanol extraction as previously described [47]. Thawed pellets were loosened from 14 mL falcon tubes and transferred to PCR 8-well tube strip, followed by the addition of 80 μL of methanol, 20 μL of chloroform, and 60 μL of water, with vortexing. The samples were centrifuged at ~20,000 x g for 1 minute for phase separation. The methanol and water (top) layer was removed, then 100 μL of methanol was added and the sample was vortexed briefly. The samples were centrifuged at ~20,000 x g for 1 minute to isolate the protein pellet. The protein pellet was air-dried for 10 minutes and resuspended in 100 mM ammonium bicarbonate with 20% methanol. The protein concentration was measured using the DC Protein Assay Kit (Bio-Rad, Hercules, CA) with bovine serum albumin for the standard curve. A total of 100 μg of protein from each sample was digested with trypsin for targeted proteomic analysis. The protein was reduced by adding tris 2-(carboxyethyl) phosphine (TCEP) at a final concentration of 5 mM, alkylated by adding iodoacetamide at a final concentration of 10 mM, and digested overnight at 37 °C with trypsin at a ratio of 1:50 (w/w) trypsin:total protein. As previously described [48], peptides were analyzed using an Agilent 1290 liquid chromatography system coupled to an Agilent 6460QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA). Peptide samples (10 μg) were separated on an Ascentis Express Peptide ES-C18 column (2.7 μm particle size, 160 Å pore size, 50 mm length x 2.1 mm i.d., 60 °C; Sigma-Aldrich, St. Louis, MO) by using a chromatographic gradient (400 μL/min flow rate) with an initial condition of 95% buffer A (99.9% water, 0.1% formic acid) and 5% buffer B (99.9% acetonitrile, 0.1% formic acid) then increasing linearly to 65% buffer A/35% buffer B over 5.5 minutes. Buffer B was then increased to 80% over 0.3 minutes and held at 80% for two minutes followed by ramping back down to 5% buffer B over 0.5 minutes where it was held for 1.5 minutes to re-equilibrate the column for the next sample. The peptides were ionized by an Agilent Jet Stream ESI source operating in positive-ion mode with the following source parameters: gas Temperature: 250 °C, gas Flow: 13 L/min, nebulizer pressure: 35

psi, sheath gas temperature: 250 °C, sheath gas flow: 11 L/min, nozzle voltage: 0 V, chamber voltage: 3,500 V. The data were acquired using Agilent MassHunter, version B.08.02, processed using Skyline [49] version 4.1, and peak quantification was refined with mProphet [50] in Skyline. Data are available at Panorama Public via this link: https://panoramaweb.org/massive_fitness_profiling_Pseudomonas_putida.url. All pairwise combinations of spectral counts from carbon sources for each protein were compared via Student's t-test followed by a Bonferroni correction.

2.6.12. Detection of metabolites

Sampling of intracellular metabolites was conducted as described previously [51]. Multiple methods were used to detect compounds in this work. Method (1) HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using positive mode and an Atlantis HILIC Silica 5 μm column (150 x 4.6 mm) with a linear of 95 to 50% acetonitrile (v/v) over 8 minutes in water with 40 mM ammonium formate, pH 4.5, at a flow rate of 1 mL minute^{-1} . Method (2) HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using negative mode and an Atlantis HILIC Silica 5 μm column (150 x 4.6 mm) with an isocratic mobile phase (80% acetonitrile (v/v) with 40 mM ammonium formate, pH 4.5) for 20 minute at a flow rate of 1 mL minute^{-1} . Method (3) is described in George et al [51]. Briefly, samples were separated via a SeQuantZIC-pHILIC guard column (20-mm length, 2.1-mm internal diameter, and 5-µm particle size; from EMD Millipore, Billerica, MA, USA), then with a short SeQuantZIC-pHILIC column (50-mm length, 2.1-mm internal diameter, and 5-µm particle size) followed by a long SeQuantZIC-pHILIC column (150-mm length, 2.1-mm internal diameter, and 5-µm particle size) using an Agilent Technologies 1200 Series Rapid Resolution HPLC system (Agilent Technologies, Santa Clara, CA, USA). The mobile phase was composed of 10 mM ammonium carbonate and 118.4 mM ammonium hydroxide in acetonitrile/water (60.2:39.8, v/v). Metabolites were eluted isocratically via a flow rate of 0.18 mL/min from 0 to 5.4 minutes, which was increased to 0.27 mL/min from 5.4 to 5.7 minutes, and held at this flow rate for an additional 5.4 minutes. The HPLC system was coupled to an Agilent Technologies 6210 TOF-MS system in negative mode. Determination of D-2HG concentration was assayed with a D-2-Hydroxyglutarate (D2HG) Assay Kit (Sigma MAK320).

2.6.13. Phylogenomic analyses

Amino acid sequences of CsiD homologs were downloaded from the pFAM database and aligned with MAFFT-linsi [52]. Phylogenetic trees of CsiD alignments were constructed with FastTree 2, and trees were visualized on iTOL [53,54].

Representative DUF1338 sequences were obtained from pFAM (https://pfam.xfam.org/family/PF07063#tabview=tab3). All genomes analyzed were downloaded from the NCBI FTP site and annotated using RAST [55]. Amino acid sequences of DUF1338 proteins from these genomes were retrieved using BlastP with a bit score cutoff of 150 and an E-value of 0.000001. All sequences alignments were performed using Muscle v3.8 [56] and the alignments were manually curated using Jalview V2 [57].

For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder implemented on IQ-tree [58] the phylogenies were obtained using IQ-tree v 1.6.7 [59], with 10,000 bootstrap replicates. The final trees were visualized and annotated using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Genome neighborhoods of DUF1338 were obtained using CORASON-BGC [60] and manually colored and annotated.

2.6.14. Statistical analyses and data presentation

All numerical data were analyzed using custom Python scripts. All graphs were visualized using either Seaborn or Matplotlib. Calculation of 95% confidence intervals, standard deviations, and T-test statistics were conducted via the Scipy library. Bonferroni corrections were calculated using the MNE python library [61].

2.7. Miscellaneous

2.7.1. Acknowledgements

This manuscript is dedicated to the memory of Cornell Professor Dr. Eugene Madsen. The authors would like to thank Morgan Price, Dr. John Hangasky, Dr. Jamie Meadows, Dr. Robert Haushalter, Dr. Bo Pang, Dr. Nick Weathersby, Mary Thompson, and Catharine Adams for their helpful discussions in preparing this manuscript. We would also 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 protein purification and homology modelling components were 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. 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.

2.7.2. Author contributions

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

2.7.3. Competing interests

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

.

Chapter 3. A Clash of Glutarate: Robust characterization of two distinct glutarate sensing transcription factors of *Pseudomonas putida* L-lysine metabolism

Including material from published work from **Mitchell G Thompson**, Zak Costello, Niklas FC Hummel, Pablo Cruz-Morales, Jacquelyn M Blake-Hedges, Rohith N Krishna, Will Skyrud, Matthew R Incha, Patrick M Shih, Hector Garcia-Martin, Jay D Keasling, "Robust characterization of two distinct glutarate sensing transcription factors of *Pseudomonas putida* L-lysine metabolism". BioRxiv, 2019

3.1. 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.
.

3.2. 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 [62,63]. Though recent advances in analytical chemistry have dramatically increased sample throughput [64], 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 [62–64], 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 [64]. The evolution of diverse microbial metabolism has provided researchers with a wide array of ligands capable of being sensed, ranging from complex natural products [65,66], to small central metabolites [62,67].

Diacids, polyamine, and lactams are all used to produce various polyester and nylon fibers which are currently derived from petrochemicals [68,69]. In an effort to make production of these chemicals sustainable, many groups have developed engineered microbes to synthesize these precursors [70–73]. The L-lysine metabolism of *Pseudomonas putida* has been leveraged to produce valerolactam [74], as well as the diacid glutarate both in the native host and heterologously [8,9]. 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 [9,75] (Figure 3-1). Recent work has also demonstrated that both glutarate catabolic pathways are highly upregulated in the presence of glutarate [75]. While work in *Pseudomonas aeruginosa* has characterized a homolog of the regulator of the ketogenic pathway (GcdR) [76], 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.

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.

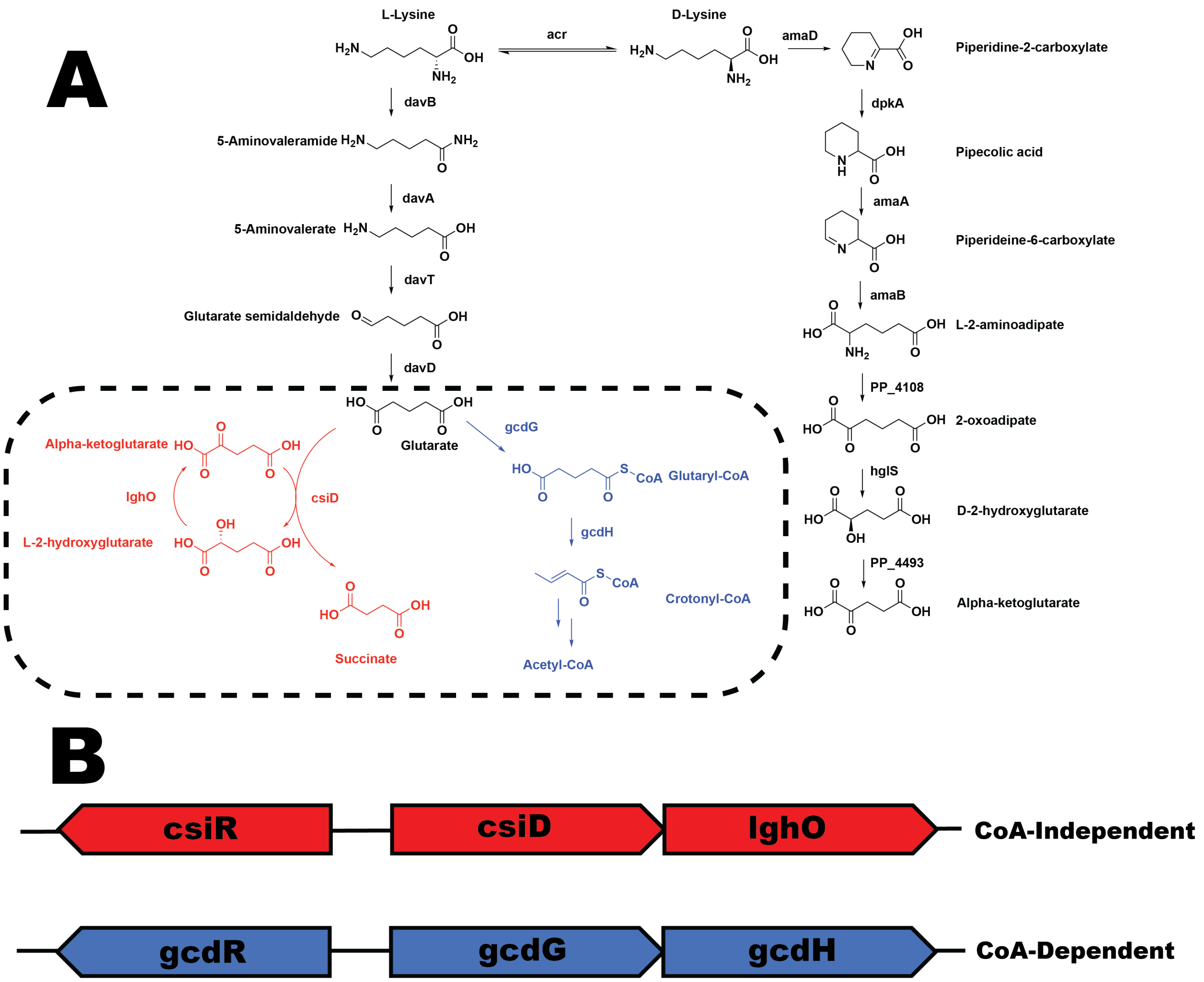


Figure 3-1 Glutarate metabolism of *P. putida*. A) The known lysine catabolism of *P. putida*. Dashed box shows the two known pathways of glutarate catabolism in *P. putida*. Highlighted in red is the CoA-independent route of glutarate catabolism, in blue the CoA-dependent route. B) Operonic structure of the two routes of glutarate metabolism in *P. putida*.

3.3. Results

3.3.1. 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 [76]. 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 [31]. 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 3-2). 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 [77]. 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_5TTTT and ATGTN_5TTTT respectively [31]. 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.

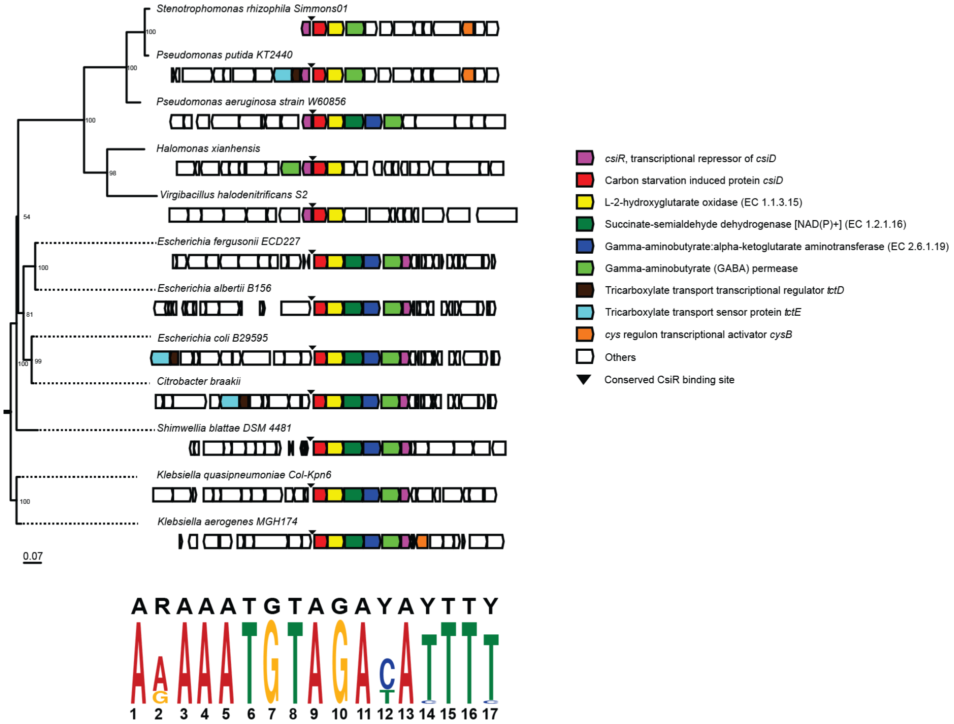


Figure 3-2 Genomic contexts of *csiR* homologs and predicted binding regions. Phylogeny of assorted gammaproteobacteria and the location of their *csiD* operons. MEME analysis of the intergenic regions upstream of the *csiD* operon resulted in the sequence motifs depicted below the tree.

3.3.2. Biochemical characterization of CsiR and GcdR

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 [31]. CsiR had a high affinity for the DNA probe, with a calculated K_d of approximately 30 nM (Figure 3-3a), which is similar to the 10 nM CsiR/DNA K_d of the *E. coli* CsiR homolog [31]. GcdR also bound to its cognate probe, with a single distinct shift, with an estimated K_d of approximately 62.5 nM (Figure 3-3a) [76]. As CsiR is a GntR family transcriptional regulator, many of which act as repressors [78], 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 3-3b). These results confirm that the *P. putida* CsiR is a glutarate-responsive repressor of the *csiDlhgO* operon [31].

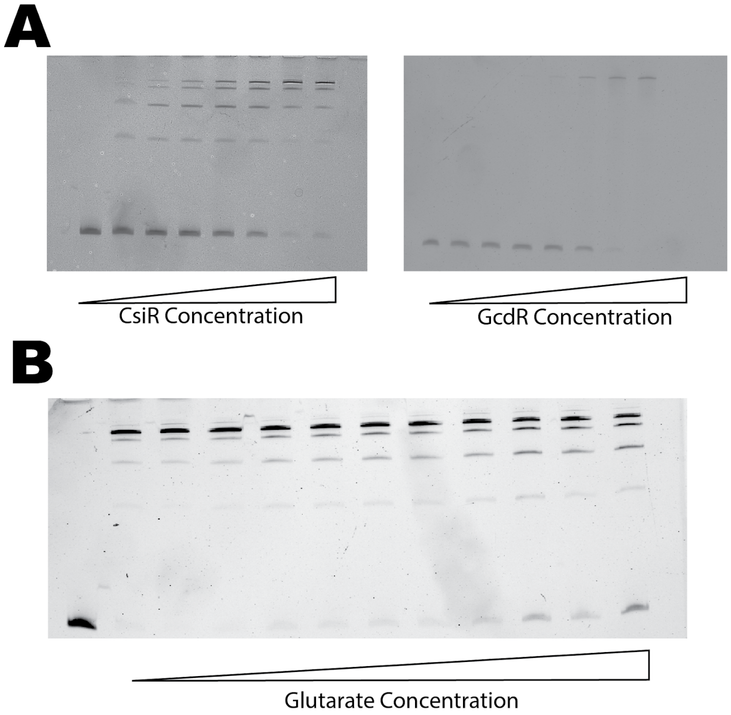


Figure 3-3 Biochemical characterization of CsiR and GcdR. A) EMSA of either purified CsiR (left) or GcdR (right) incubated with fluorescent probe of intergenic region between *csiR* and *csiD*. Concentrations of CsiR protein range from 0 nM to 250 nM. (B) EMSA showing 100 nM CsiR and fluorescent probe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

3.3.3. 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 was 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 3-4A). 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 0.6, a K_d of 1.25 mM, and a maximum normalized RFP expression of 9958, while the CsiR system was shown to have a Hill coefficient of 0.89, a K_d of 0.31 mM, and a maximum normalized RFP expression of 8546 (Figure 3-4A). To test for the ability of CsiR or GcdR to sense other diacids, *E. coli* harboring either vector was 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 3-5).

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 [79]. 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 : 0.21 mM, J23101 : 0.05 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, J23101, showing the lowest max RFP at 15,558 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 3-4B). 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 3-4B).

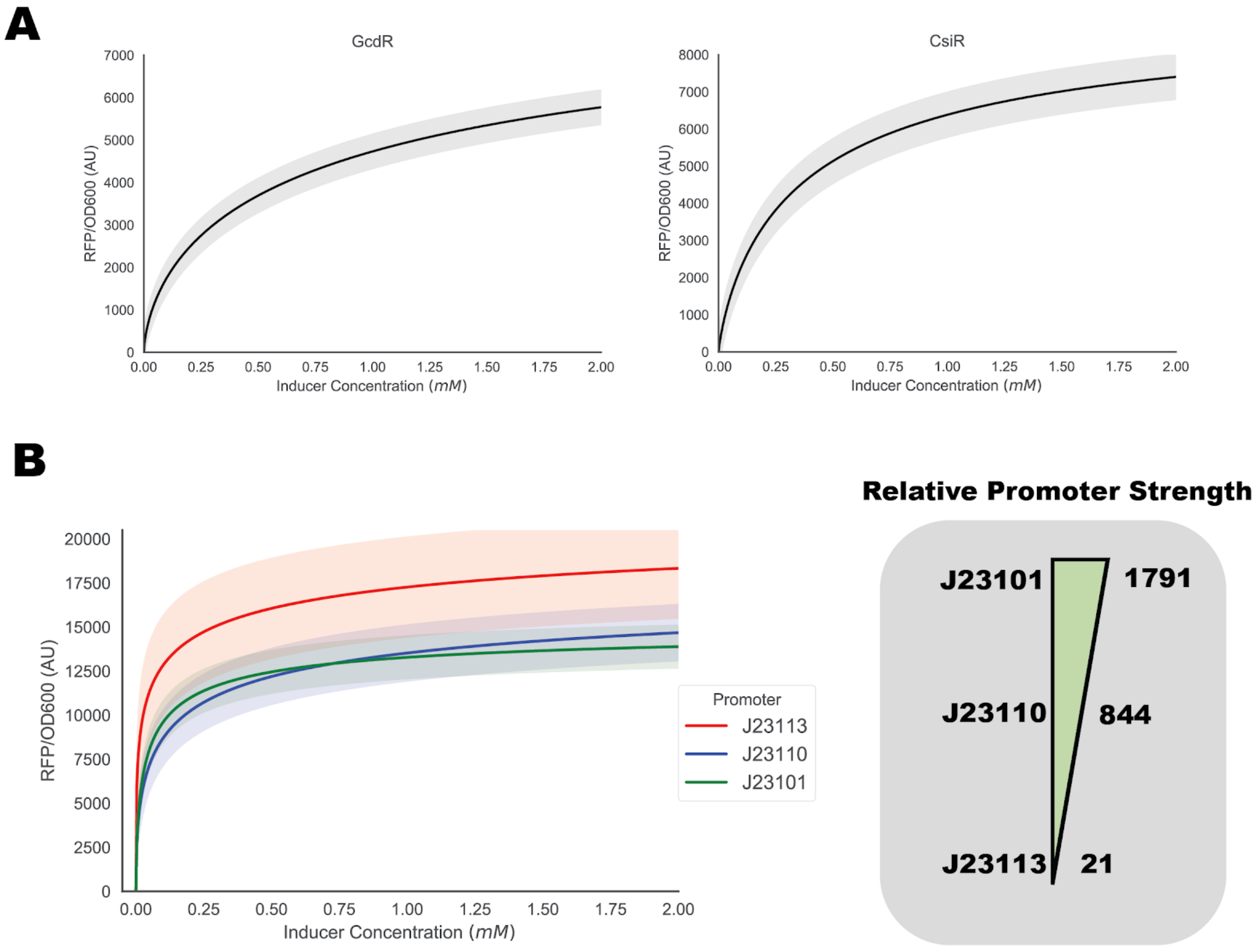


Figure 3-4 Development of glutarate responsive vectors for *E. coli*.Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for native CsiR and GcdR systems. Shaded area represents (+/-) one standard deviation, n=4. (B) Fluorescence data fit to the Hill equation to 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 3-1 Best fits of biosensor performance parameters with standard deviations.

Reporter	Hill Coef.	K_d (mM)	Max	Induction
CsiR-Native	0.89 (+/- 0.04)	0.31 (+/-0.03)	8546 (+/- 245)	2.98

GcdR-Native	0.63 (+/- 0.04)	1.25 (+/- 0.61)	9958 (+/- 1183)	2.07
GcdR-J23113	0.33 (+/- 0.05)	0.07 (+/- 0.006)	24190 (+/- 7454)	1.91
GcdR-J23110	0.46 (+/- 0.04)	0.21 (+/- 0.14)	19848 (+/- 1905)	3.17
GcdR-J23101	0.58 (+/- 0.03)	0.05 (+/- 0.009)	15588 (+/- 561)	2.65

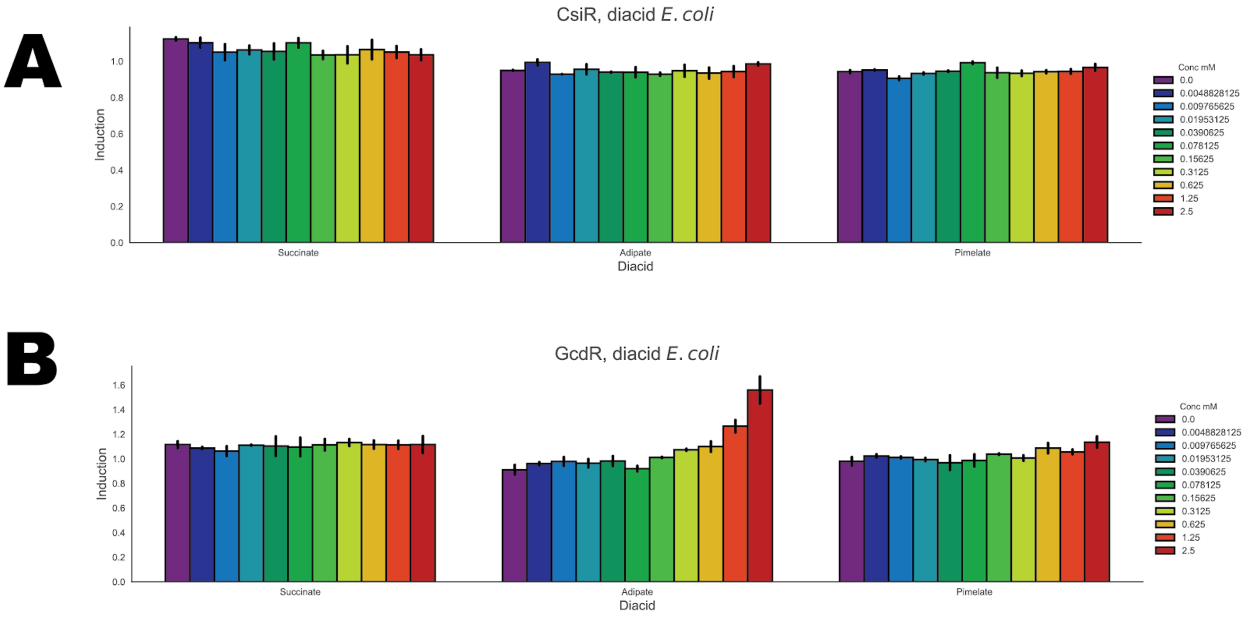


Figure 3-5 Development of glutarate responsive vectors for *E. coli* by alternative diacids. Induction of the CsiR (A), or GcdR (B) biosensor vectors by succinate, adipate, and palmitate. Error bars represent 95% CI, n=4**.**

3.3.4. 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 [62,67,80]. 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 [81]. A detailed methodological description can be found in the supplemental Jupyter notebook, which can also be used to analyze any other biosensor data and can be found at https://github.com/JBEI/biosensor_characterization.

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 3-6A). 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 3-6A). 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 3-6B). 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 3-2). The upper limit of detection is determined by finding the largest inducer concentration that can be statistically distinguished from maximum tested inducer concentration (Table 3-2). 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 3-6C). 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 in Figure 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 3-6D). This “power” analysis may serve as a guide for experimental design when a certain biosensor resolution is required for a given application.

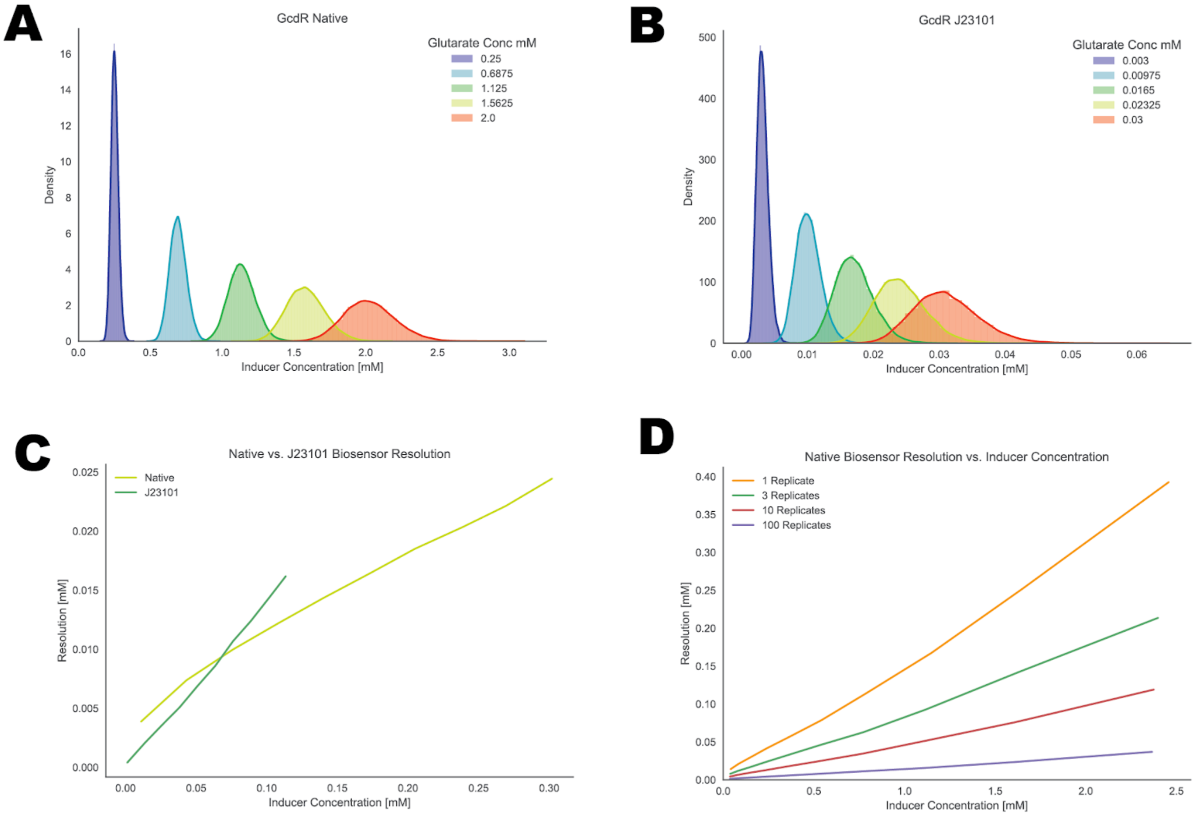


Figure 3-6 Development of analytics for biosensor performance. Probability density functions of fluorescence values compatible with concentrations of glutarate for selected ligand concentrations modelled to the Hill equation of the native GcdR system (A) or the GcdR J23101 system (B). (C) Resolution 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 3-2 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

3.3.5. 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 3-7A). 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 3-7A). 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 semialdehyde (Figure 3-1A) 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 3-7B). 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 3-7C). 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 3-8A and Figure 3-8B). 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 3-8A). Conversely, the strain harboring the CsiR vector only displayed induction of RFP above background when grown on 5-aminovalerate (Figure 3-8B). 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 3-8).

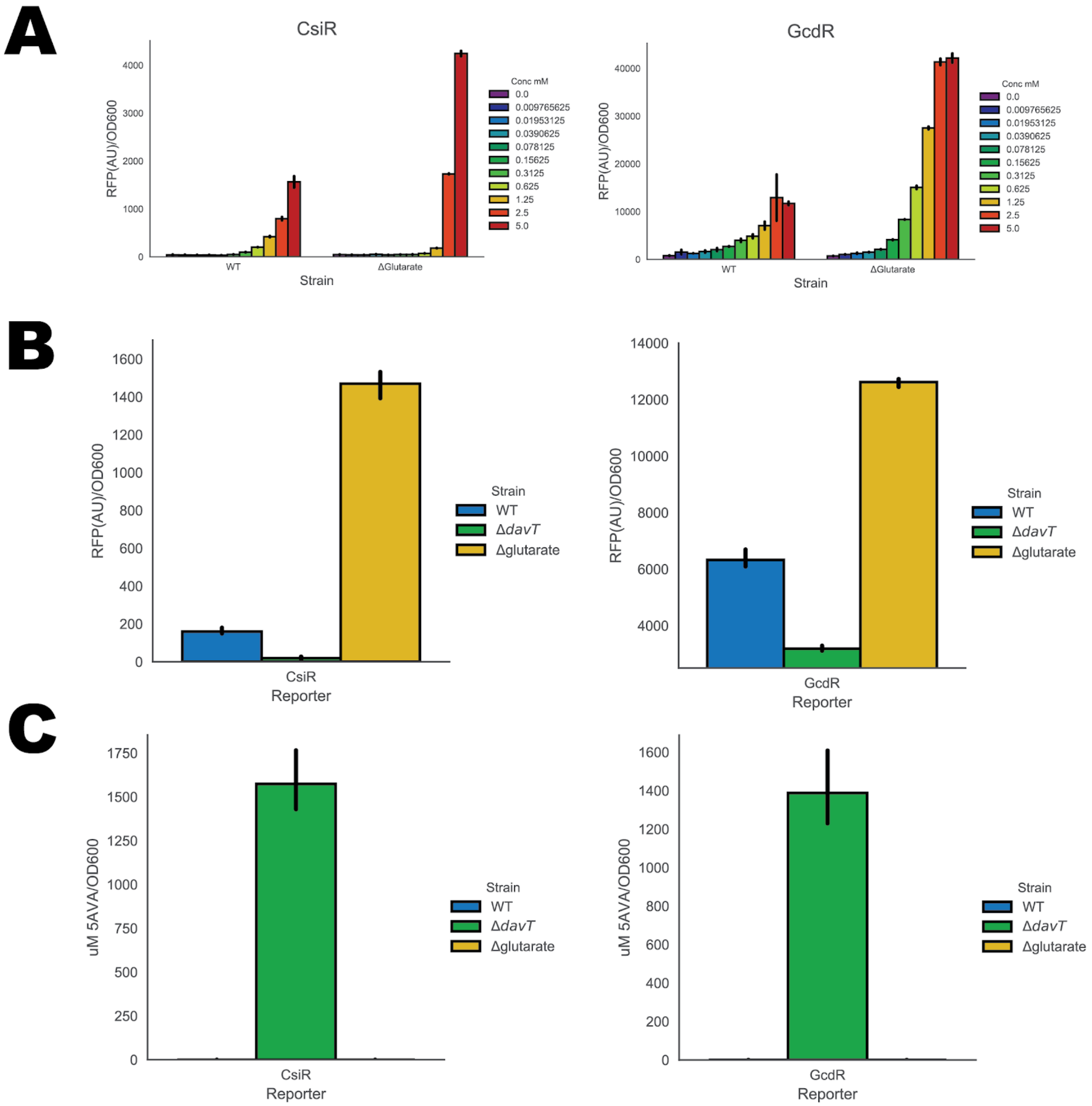


Figure 3-7 Performance of CsiR and GcdR biosensors in *P. putida*. (A) RFP expression of either wild type *P. putida* or *P. putida* with the ability to metabolize glutarate knocked out measured with either the CsiR or GcdR biosensor under different external glutarate concentrations. Error bars represent 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.

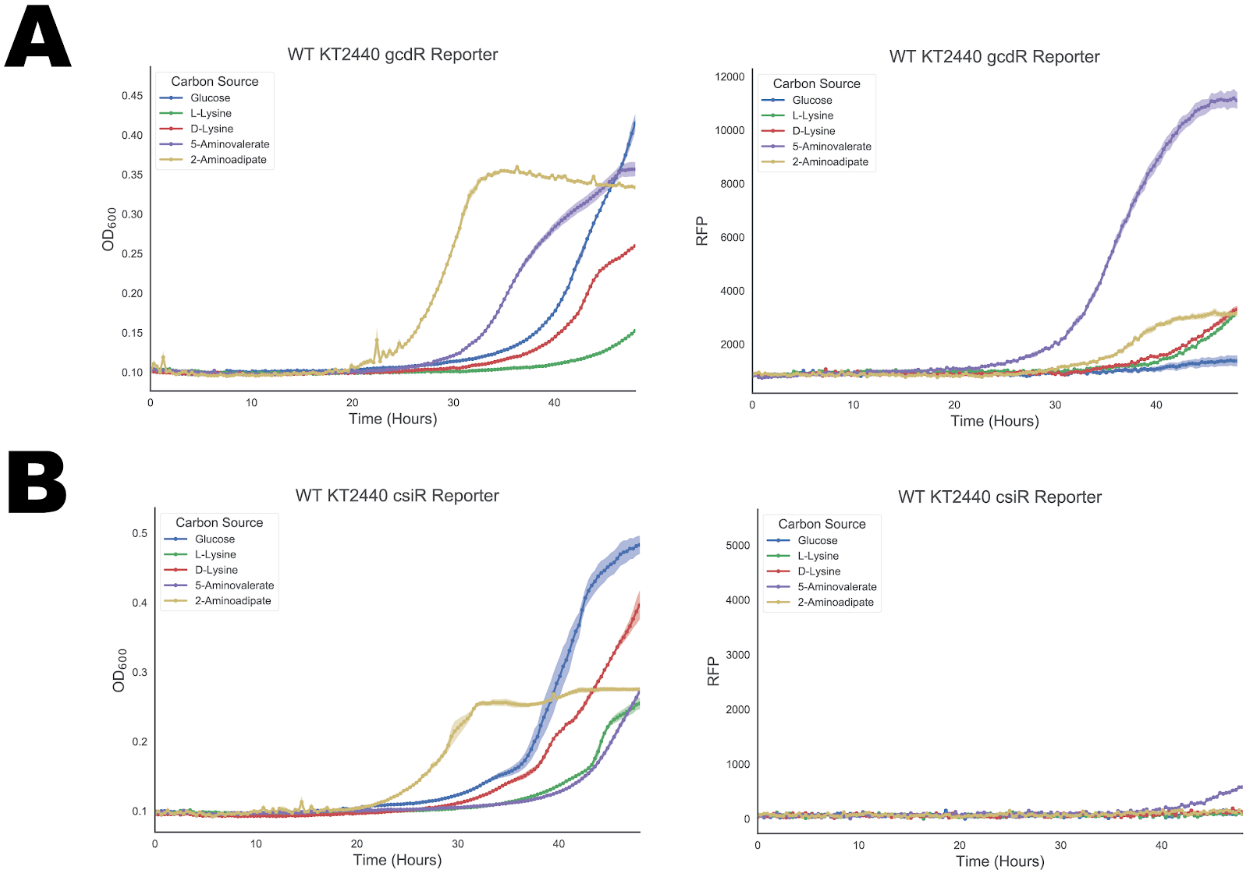


Figure 3-8 Performance of CsiR and GcdR biosensors in *P. putida* grown on lysine metabolites. Growth and fluorescence of *csiR* and *gcdR* vectors in P. putida KT2440 (A) Growth of strains harboring *gcdR* vector (left), RFP fluorescence (right) on MOPS minimal media supplemented with 10mM of glucose, L-lysine, D-lysine, 5-aminovalerate, or 2-aminoadipate as a sole carbon source. Shaded region represents 95% CI, n=3. (B) Growth of strains harboring *csiR* vector (left), RFP fluorescence (right) on MOPS minimal medium supplemented with 10mM of glucose, L-lysine, D-lysine, 5-aminovalerate, or 2-aminoadipate as a sole carbon source. Shaded region represents 95% CI, n=3.

3.4. 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 [15,82]. 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 [76]. 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 [9,31]. 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 [83]. 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 regulators 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 [84]. 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 [85,86]. 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 [75]. While cross-feeding between the pathways has been proposed previously [23], recent work by our group has proposed a molecular mechanism for metabolite exchange between the D- and L- catabolic pathways [75]. 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 [9,87,88]. 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 [89]. 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 [65]. Such methods could be applied to GcdR in order to expand its utility in sensing other industrially important diacids.

3.5. Materials and methods

3.5.1. Chemicals and media

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 [39]. 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).

3.5.2. Strains and plasmids

All bacterial strains and plasmids used in this work are listed in Table 3-3. 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 [40–42]. Plasmids were assembled via Gibson Assembly using standard protocols [43], or Golden Gate Assembly using standard protocols [44]. 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 [75].

Table 3-3 Strains and plasmids used in this study.

Strain	**JBEI Part ID**	**Reference**
E. coli DH10B		[90]
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*		[75]
Plasmids		
pET21b		Novagen
pBADT		[91]
pMQ30		[45]
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

3.5.3. Expression and purification of proteins

Proteins were purified as described previously [92]. Briefly, 500 mL cultures of *E. coli* BL21 (DE3) harboring expression plasmids were grown in LB medium at 37 °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 °C before being harvested via centrifugation. Cell pellets were stored at -80 °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 °C until *in vitro* analysis.

3.5.4. Plate based growth on 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.

3.5.5. 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 [93]. 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.

3.5.6. Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were performed as previously described [66]. 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 °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.

3.5.7. Measurement of 5-aminovalerate

To measure intracellular concentrations of 5-aminovalerate cells were quenched as previously described [51]. 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.

3.5.8. Analysis of biosensor parameters

A model relating inducer concentrations and fluorescence measurements to characterize the performance of a biosensor was generated 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 hθ(.) is the Hill function, θ are the parameters of the hill function, and σ 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 μ±2σ.

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 c^=βmax(C). 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 html.

3.5.9. Bioinformatic analysis

For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder as implemented on IQ-tree [58] 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 [60] and manually colored and annotated. DNA-binding sites were predicted with MEME [77].

3.6. Miscellaneous

3.6.1. 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).

3.6.2. Collaborator 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.

3.6.3. Competing interests

JDK has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

Chapter 4. A Feast on Valerolactam: Host engineering for improved valerolactam production in *Pseudomonas putida*

Including material from published work from **Mitchell G Thompson**, Luis E Valencia, Jacquelyn M Blake-Hedges, Pablo Cruz-Morales, Alexandria E Velasquez, Allison N Pearson, Lauren N Sermeno, William A Sharpless, Veronica T Benites, Yan Chen, Edward EK Baidoo, Christopher J Petzold, Adam M Deutschbauer, Jay D Keasling, “Host engineering for improved valerolactam production in *Pseudomonas putida*”. BioRxiv, 2019

4.1. Abstract

Pseudomonas putida is a promising bacterial chassis for metabolic engineering given its ability to metabolize a wide array of carbon sources, especially aromatic compounds derived from lignin. However, this omnivorous metabolism can also be a hindrance when it can naturally metabolize products produced from engineered pathways. Herein we show that *P. putida* is able to use valerolactam as a sole carbon source, as well as degrade caprolactam. Lactams represent important nylon precursors, and are produced in quantities exceeding one million tons per year[8]. To better understand this metabolism we use a combination of Random Barcode Transposon Sequencing (RB-TnSeq) and shotgun proteomics to identify the *oplBA* locus as the likely responsible amide hydrolase that initiates valerolactam catabolism. Deletion of the *oplBA* genes prevented *P. putida* from growing on valerolactam, prevented the degradation of valerolactam in rich media, and dramatically reduced caprolactam degradation under the same conditions. Deletion of *oplBA*, as well as pathways that compete for precursors L-lysine or 5-aminovalerate, increased the titer of valerolactam from undetectable after 48 hours of production to ~90 mg/L. This work may serve as a template to rapidly eliminate undesirable metabolism in non-model hosts in future metabolic engineering efforts.

4.2. Introduction

Pseudomonas putida has attracted great attention as a potential chassis organism for metabolic engineering due in large part to its ability to metabolize a wide variety of carbon sources, particularly aromatic compounds that can be derived from lignin [3,94]. To more fully realize this vision, much effort has been put forth recently to better characterize the central metabolism of *P. putida* with updated genome-scale models [95], C^{13} flux experiments [96,97], and high-throughput fitness assays, which have all contributed to a more complete understanding of the bacterium [16,75]. Despite these advances, the metabolic capabilities of *P. putida* are not yet fully understood.

One consequence of this omnivorous metabolism is that *P. putida* possesses the ability to degrade or fully catabolize chemicals metabolic engineers seek to produce in the host. An ongoing challenge for *P. putida* host engineering will be to rapidly identify catabolic pathways of important target molecules and eliminate them from the genome. The recent report of a novel pathway for levulinic acid catabolism in *P. putida* KT2440 underscores the catabolic flexibility of the host, and an additional obstacle towards producing high product titer [16]. While challenging, this is not surprising; as a genus, *Pseudomonads* are well known for their ability to degrade a wide range of naturally occurring or xenobiotic chemicals [98,99].

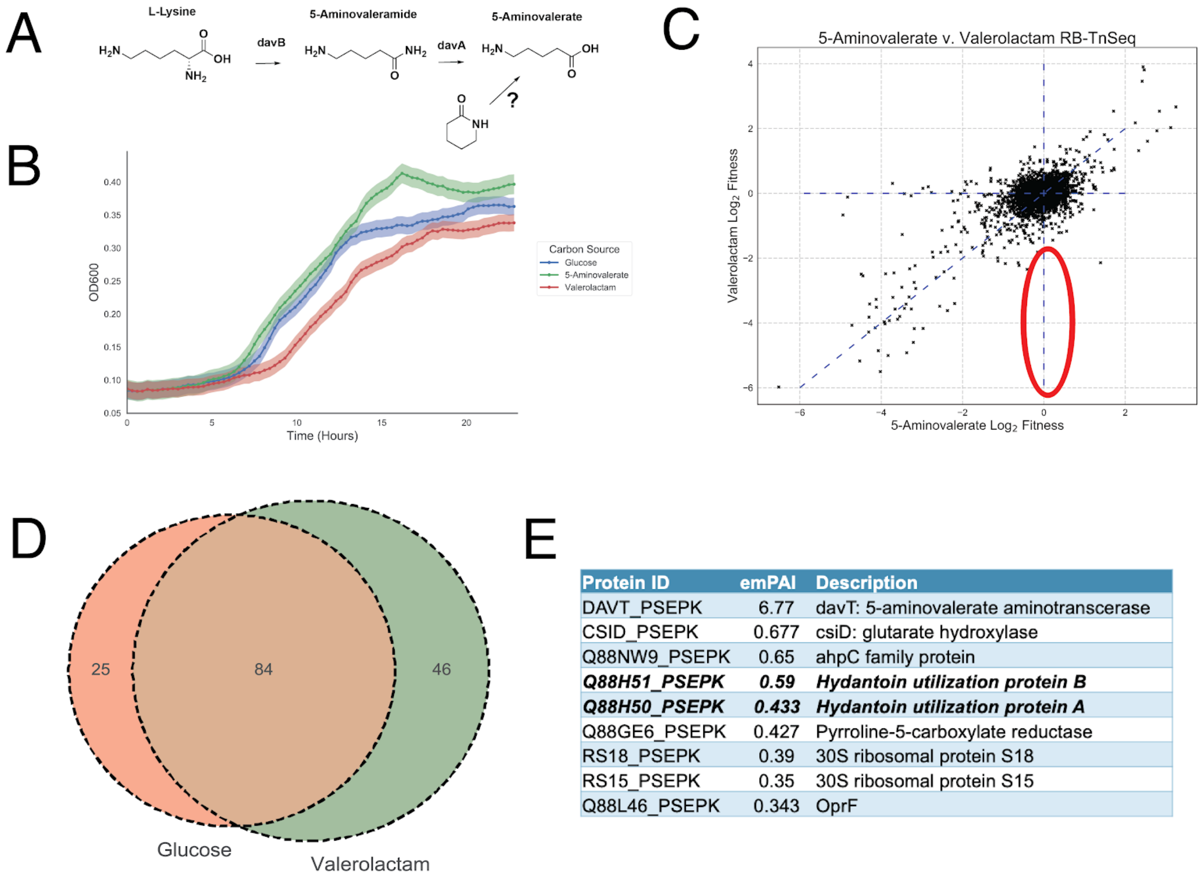
Caprolactam and valerolactam are both important commodity chemicals used in the synthesis of nylon polymers, with global production of caprolactam reaching over four million metric tons [8]. Multiple efforts have been made to produce these chemicals biologically, with the titers of valerolactam approaching 1g/L in *Escherichia coli* [7,8]. The engineered pathway to valerolactam in *E. coli* converts L-lysine to 5-aminovalerate (5AVA) via DavBA, two genes endogenous to *P. putida*, and then cyclizes it via a promiscuous coA-ligase [8]. While the endogenous L-lysine catabolism catabolism of *P. putida* has been leveraged to produce the C5 diacid glutarate, there has yet to be any attempt to produce valerolactam in the bacterium [9]. A recent report that *Pseudomonas jessenii* can catabolize caprolactam suggested that catabolism of lactams could dramatically impact titers of valerolactam in *P. putida* [100].

In this work we demonstrate that *P. putida* can utilize valerolactam as a sole carbon source, as well as degrade caprolactam. Using a combination of Random Barcode Transposon Sequencing (RB-TnSeq) and shotgun proteomics we were able to identify that OplBA, a predicted oxoprolinase, is responsible for this hydrolysis. By knocking out *oplBA* in addition to two pathways that compete for precursors we were able to dramatically increase the titers of valerolactam in *P. putida*.

4.3. Results

4.3.1. Identification of a lactam hydrolase in *P. putida*

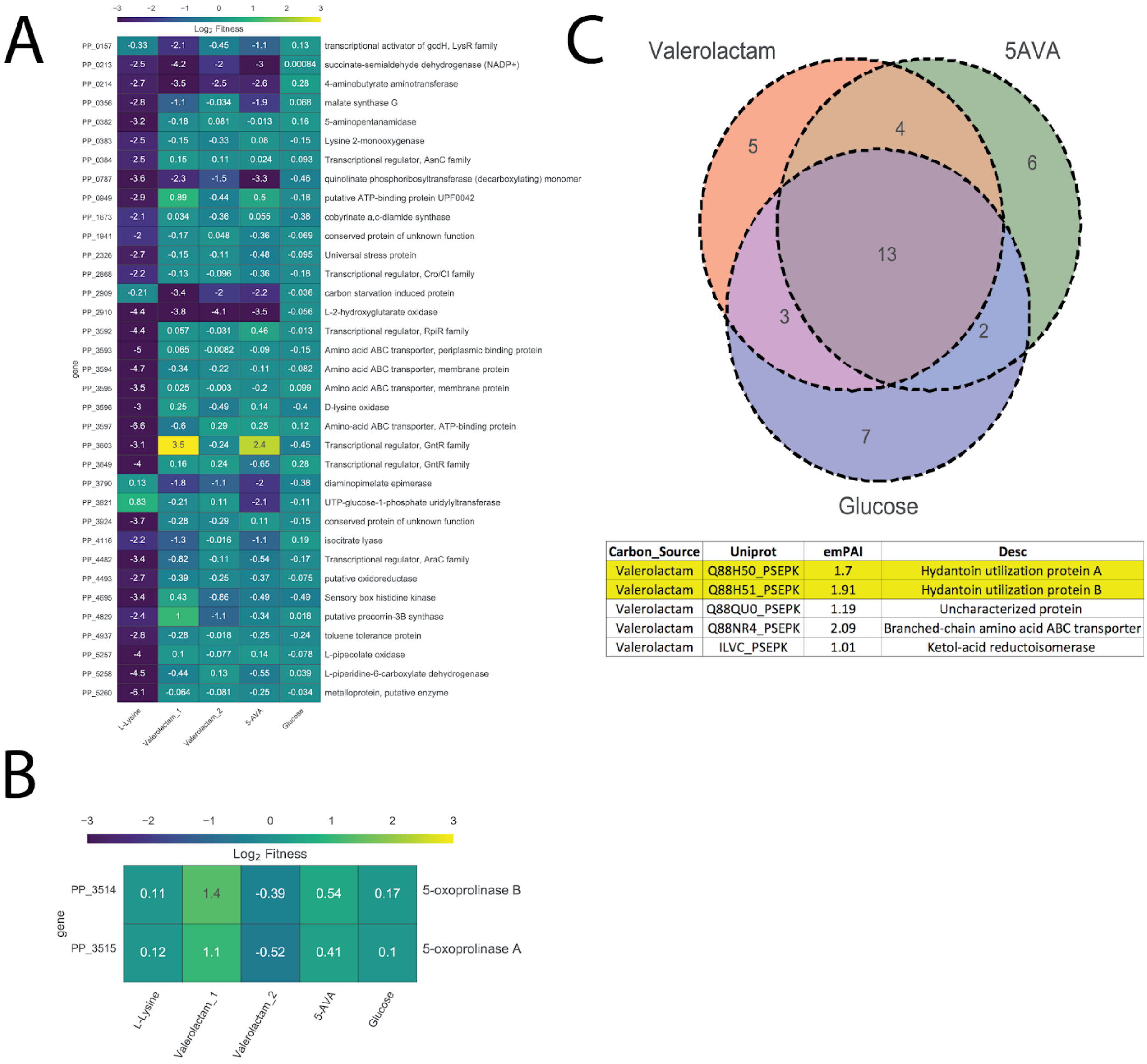
The hydrolysis product of valerolactam is 5AVA, an intermediate in L-lysine metabolism (Figure 4-1A). Growth curves of *P. putida* on valerolactam as a sole carbon source demonstrated that the bacterium was readily able to catabolize valerolactam and produce biomass, with growth similar to that on either 5AVA and glucose (Figure 4-1B). Initially, we attempted to identify the enzyme responsible for valerolactam hydrolysis using RB-TnSeq, a technique that has previously been used to identify novel enzymes in D-lysine metabolism [75]. RB-TnSeq measures the relative fitness of transposon mutant pools to infer gene function through changes in relative abundance of all non-essential genes in a bacterium under a selective condition [14,15]. Mutant pools of *P. putida* were grown on either 5AVA or valerolactam as a sole carbon source in an attempt to identify enzymes solely essential for growth on valerolactam. Results of RB-TnSeq experiments suggested that valerolactam was indeed being catabolized through the same pathway as 5AVA with both conditions showing significant defects in the *davTD* and *csiD-lghO* operons, the known catabolic route of 5AVA to the TCA cycle (Figure S1A). However, there were no genes that showed obvious fitness defects only under the valerolactam growth condition (Figure 4-1C).



Protein ID	emPAI	Description
DAVT_PSEPK	6.77	davT: 5-aminovalerate aminotranscerase
CSID_PSEPK	0.677	csiD: glutarate hydroxylase
Q88NW9_PSEPK	0.65	ahpC family protein
Q88H51_PSEPK	***0.59***	***Hydantoin utilization protein B***
Q88H50_PSEPK	***0.433***	***Hydantoin utilization protein A***
Q88GE6_PSEPK	0.427	Pyrroline-5-carboxylate reductase
RS18_PSEPK	0.39	30S ribosomal protein S18
RS15_PSEPK	0.35	30S ribosomal protein S15
Q88L46_PSEPK	0.343	OprF

Figure 4-1 Identification of the *P. putida* valerolactam hydrolase. (A) Route of valerolactam catabolism through the L-lysine catabolic route of *P. putida* (B) Growth of *P. putida* in minimal medium supplemented with either 10 mM glucose, 5AVA, or valerolactam. Shaded area represents the 95% confidence interval (cI), n=3. (C) RB-TnSeq analysis of genome fitness assays of *P. putida* libraries grown on either 5AVA or valerolactam as a sole carbon source. Red oval shows the predicted fitness result of a valerolactam hydrolase. (D) Results of shotgun proteomics of proteins found in the supernatant of *P. putida* grown on either 10 mM glucose or 10 mM valerolactam as a sole carbon source. Venn diagram shows the number of proteins with an exponentially modified protein abundance index (emPAI) relative abundance above 0.1 shared or unique to each carbon source (E) Table shows the most abundant proteins specific to grown on valerolactam. OplA (Q88H50_PSEPK) and OplB (Q88H51_PSEPK) are in bold.

The inability of the RB-TnSeq approach to identify the hydrolytic enzyme could result from the enzyme being secreted from the cell. In this case, the secreted enzymes from cells containing the intact hydrolase gene produce 5AVA which can freely diffuse into hydrolase mutant cells, eliminating any selective pressure for lactam-based growth. To test whether the enzymes responsible for lactam hydrolysis may be secreted, cultures of *P. putida* were either grown on glucose or valerolactam as a sole carbon source, their supernatants filtered, concentrated, and then subjected to shotgun proteomics. Of the most abundant proteins in the supernatant, there were 25 proteins that were specific to glucose, while 46 were specific to valerolactam, with 86 proteins being shared between the two conditions (Figure 4-11D). Within the top five most abundant proteins in the valerolactam supernatant were OplB (Q88H51_PSEPK) and OplA (Q88H51_PSEPK), which together are annotated as the two subunits of a 5-oxoprolinase, orthologs of which have been suggested to participate in the caprolactam catabolism of *P. jessenii* (Figure 4-1E) [100]. Additional shotgun proteomics of cell pellets grown on either glucose, 5AVA, or valerolactam showed that OplB and OplA were more highly expressed in the presence of the lactam in comparison to the other carbon sources (Figure 4-2C). Interestingly, fitness data from two valerolactam RB-TnSeq experiments in *P. putida* KT2440 show *oplBA* mutants having no significant fitness defects (Figure 4-2B). Orthologs of *oplBA* are widely distributed across many bacteria including other attractive metabolic engineering chassis such as *Rhodococcus opacus*, and are nearly always located adjacent to one another on the genome (Figure 4-3).



gene	L-Lysine	Valerolactam_1	Valerolactam_2	5-AVA	Glucose	
PP_0157	-0.33	-2.1	-0.45	-1.1	0.13	transcriptional activator of gcdH, LysR family
PP_0213	-2.5	-4.2	-2	-3	0.00084	succinate-semialdehyde dehydrogenase (NADP+)
PP_0214	-2.7	-3.5	-2.5	-2.6	0.28	4-aminobutyrate aminotransferase
PP_0356	-2.8	-1.1	-0.034	-1.9	0.068	malate synthase G
PP_0382	-3.2	-0.18	0.081	-0.013	0.16	5-aminopentanamidase
PP_0383	-2.5	-0.15	-0.33	0.08	-0.15	Lysine 2-monooxygenase
PP_0384	-2.5	0.15	-0.11	-0.024	-0.093	Transcriptional regulator, AsnC family
PP_0787	-3.6	-2.3	-1.5	-3.3	-0.46	quinolinate phosphoribosyltransferase (decarboxylating) monomer
PP_0949	-2.9	0.89	-0.44	0.5	-0.18	putative ATP-binding protein UPF0042
PP_1673	-2.1	0.034	-0.36	0.055	-0.38	cobyrinate a,c-diamide synthase
PP_1941	-2	-0.17	0.048	-0.36	-0.069	conserved protein of unknown function
PP_2326	-2.7	-0.15	-0.11	-0.48	-0.095	Universal stress protein
PP_2868	-2.2	-0.13	-0.096	-0.36	-0.18	Transcriptional regulator, Cro/CI family
PP_2909	-0.21	-3.4	-2	-2.2	-0.036	carbon starvation induced protein
PP_2910	-4.4	-3.8	-4.1	-3.5	-0.056	L-2-hydroxyglutarate oxidase
PP_3592	-4.4	0.057	-0.031	0.46	-0.013	Transcriptional regulator, RpiR family
PP_3593	-5	0.065	-0.0082	-0.09	-0.15	Amino acid ABC transporter, periplasmic binding protein
PP_3594	-4.7	-0.34	-0.22	-0.11	-0.082	Amino acid ABC transporter, membrane protein
PP_3595	-3.5	0.025	-0.003	-0.2	0.099	Amino acid ABC transporter, membrane protein
PP_3596	-3	0.25	-0.49	0.14	-0.4	D-lysine oxidase
PP_3597	-6.6	-0.6	0.29	0.25	0.12	Amino-acid ABC transporter, ATP-binding protein
PP_3603	-3.1	3.5	-0.24	2.4	-0.45	Transcriptional regulator, GntR family
PP_3649	-4	0.16	0.24	-0.65	0.28	Transcriptional regulator, GntR family
PP_3790	0.13	-1.8	-1.1	-2	-0.38	diaminopimelate epimerase
PP_3821	0.83	-0.21	0.11	-2.1	-0.11	UTP-glucose-1-phosphate uridylyltransferase
PP_3924	-3.7	-0.28	-0.29	0.11	-0.15	conserved protein of unknown function
PP_4116	-2.2	-1.3	-0.016	-1.1	0.19	isocitrate lyase
PP_4482	-3.4	-0.82	-0.11	-0.54	-0.17	Transcriptional regulator, AraC family
PP_4493	-2.7	-0.39	-0.25	-0.37	-0.075	putative oxidoreductase
PP_4695	-3.4	0.43	-0.86	-0.49	-0.49	Sensory box histidine kinase
PP_4829	-2.4	1	-1.1	-0.34	0.018	putative precorrin-3B synthase
PP_4937	-2.8	-0.28	-0.018	-0.25	-0.24	toluene tolerance protein
PP_5257	-4	0.1	-0.077	0.14	-0.078	L-pipecolate oxidase
PP_5258	-4.5	-0.44	0.13	-0.55	0.039	L-piperidine-6-carboxylate dehydrogenase
PP_5260	-6.1	-0.064	-0.081	-0.25	-0.034	metalloprotein, putative enzyme

gene	L-Lysine	Valerolactam_1	Valerolactam_2	5-AVA	Glucose	
PP_3514	0.11	1.4	-0.39	0.54	0.17	5-oxoprolinase B
PP_3515	0.12	1.1	-0.52	0.41	0.1	5-oxoprolinase A

Carbon_Source	Uniprot	emPAI	Desc
Valerolactam	Q88H50_PSEPK	1.7	Hydantoin utilization protein A
Valerolactam	Q88H51_PSEPK	1.91	Hydantoin utilization protein B
Valerolactam	Q88QU0_PSEPK	1.19	Uncharacterized protein
Valerolactam	Q88NR4_PSEPK	2.09	Branched-chain amino acid ABC transporter
Valerolactam	ILVC_PSEPK	1.01	Ketol-acid reductoisomerase

Figure 4-2 RB-TnSeq and Cellular Shotgun Proteomics Results. (A) Genes that show significant (t < -4) and large (fitness < -2) fitness defects specific to either L-lysine, 5AVA, or valerolactam, but not glucose. All non-valerolactam fitness experiments are from Thompson et. al 2019. (B) Fitness of the *oplA* and *oplB* genes on all carbon sources (C) Venn diagram showing the number of specific proteins found in the 100 most abundant proteins found within *P. putida* grown on either glucose, valerolactam, or 5-aminovalerate, based on emPAI. Below we can see the 5 most abundant proteins specific to growth on valerolactam. OplA and OplB are highlighted.

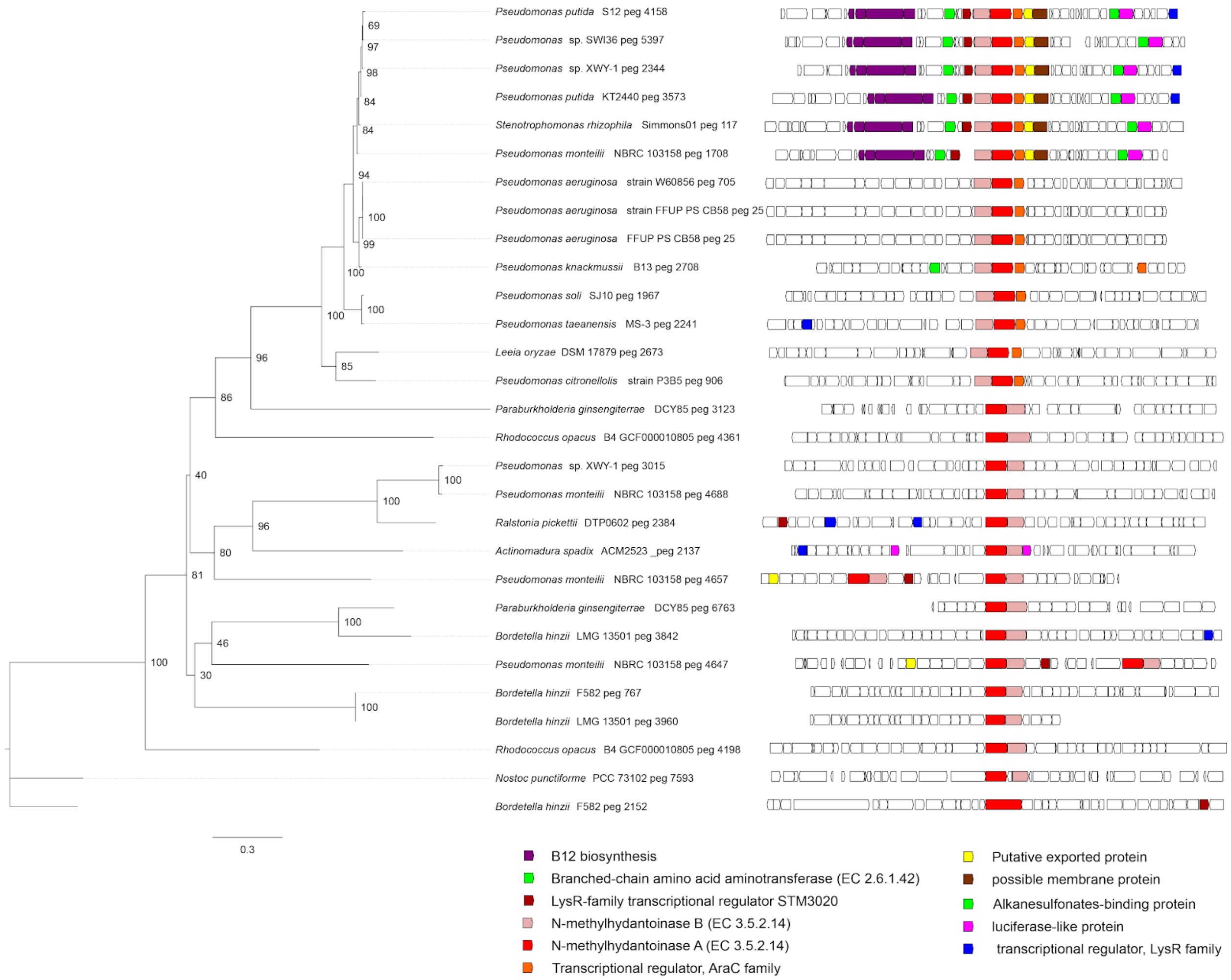


Figure 4-3 Distribution of OplBA orthologs. Phylogenomics of selected OplBA homologs across bacteria. The boxes represent the gene neighborhood for each homolog. The genes have been colored to represent their annotated functions.

4.3.2. Deletion of *oplBA* mitigates consumption of valerolactam and caprolactam

To confirm the role of OplBA in the catabolism of valerolactam, deletions were constructed of the *oplBA* locus in *P. putida* via homologous recombination. Growth of the mutant was compared to the wild type as well as to a deletion mutant of *davT*, which catalyzes the first step in 5AVA catabolism. All strains showed identical growth on glucose as a sole carbon source (Figure 4-4A). In the presence of 5AVA the *oplBA* deletion strain showed no growth defect compared to the wild-type, while the *davT* mutant predictably was unable to grow (Figure 4-4B). However, on valerolactam only the wild type strain grew, while both the *oplBA* and *davT* mutants showed no measurable growth after 40 hours (Figure 4-4C). These results suggest that OplBA is the sole enzyme responsible for the conversion of valerolactam to 5AVA under these conditions.

While *P. putida* KT2440 can utilize valerolactam as a sole carbon source, it is unable to grow on caprolactam (data not shown). In order to determine whether the OplBA of *P. putida* is capable of degrading other lactams, wild type and Δ*oplBA* strains were grown in LB medium supplemented with 10 mM caprolactam, valerolactam, or butyrolactam for 24 hours after which the remaining lactam concentration was compared to an uninoculated medium control. Wild type *P. putida* consumed all detectable valerolactam within 24 hours, and less than 50% of both butyrolactam and caprolactam remained compared to the uninoculated control (Figure 4-4D). There was no significant decrease in amount of valerolactam in the Δ*oplBA* cultures compared to the uninoculated control (t-test of *p*=0.178), though there was a slight but significant decrease of 0.74 mM caprolactam (t-test of *p*=0.033). No butyrolactam remained in the Δ*oplBA* culture after 24 hours (Figure 4-4D). This result was surprising as the annotated substrate of OplBA, 5-oxoproline, has the same ring size as butyrolactam. Previous work has shown that homologs of OplBA are ATP-dependent amidohydrolases that hydrolyze 5-oxoproline [101]. However, when purified OplBA was incubated with valerolactam in addition to ATP and magnesium we observed no hydrolysis relative to boiled enzyme controls (Figure 4-5).

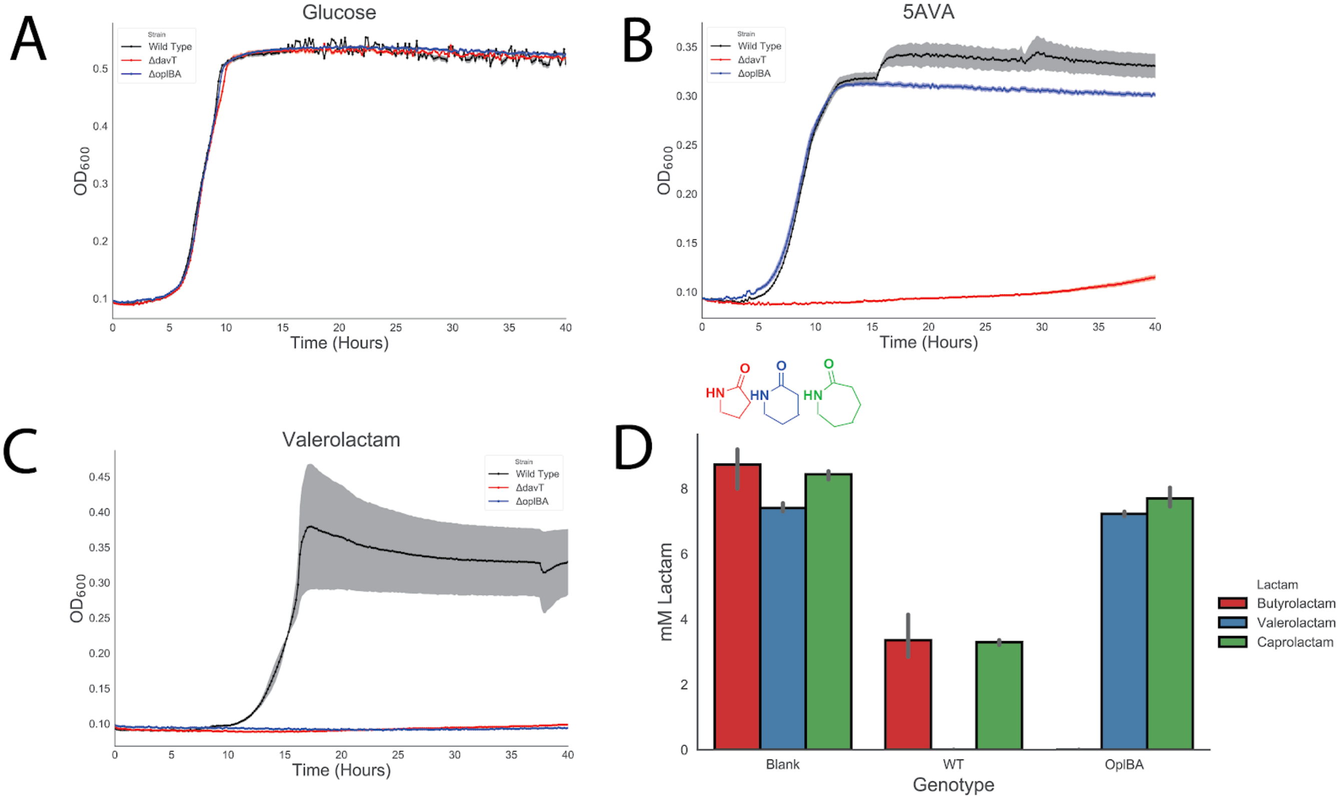


Figure 4-4 OplBA controls valerolactam and caprolactam degradation in *P. putida*. Growth of wild-type, Δ*davT*, or Δ*oplBA* in minimal media supplemented with either 10 mM glucose (A), 5-aminovaleroate (B), or valerolactam (C). (D) Remaining butyrolactam, valerolactam, or caprolactam in LB media after 24-hour incubation with no *P. putida*, wild-type, or a Δ*oplBA* mutant.

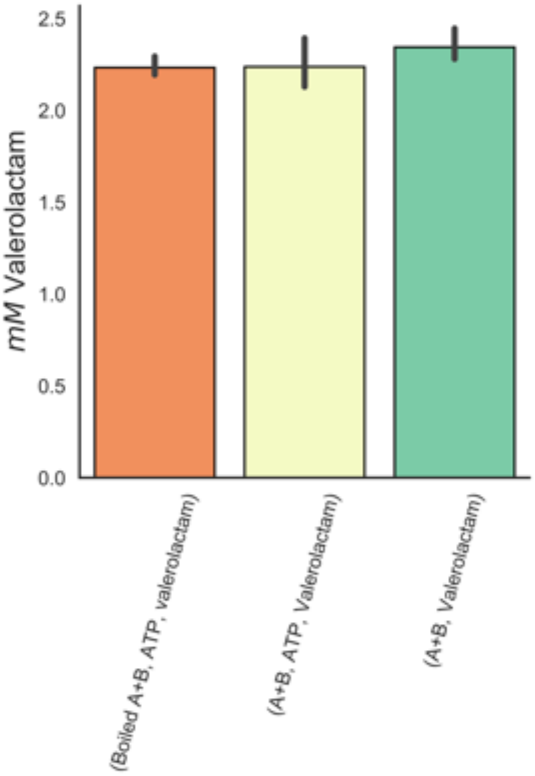


Figure 4-5 Biochemical characterization of OplAB. When purified OplA and OplB were incubated in the presence of valerolactam with or without ATP for 4 hours, there was no significant decrease in lactam concentration compared to boiled enzyme control. Error bars represent 95% cI, n=3.

4.3.3. Host engineering for increased valerolactam production

The published pathway for the production of valerolactam from L-lysine in *E. coli* utilized the formation of 5AVA via the *davBA* pathway native to *P. putida*, followed by cyclization to the lactam via a promiscuous acyl-coA ligase from *Streptomyces aizunensis* (Figure 4-6A) [8]. To produce valerolactam in *P. putida*, not only will the production pathway need to be overexpressed, but the *oplBA* locus and pathways that compete for 5AVA must be eliminated. Loss of flux to valerolactam occurs through two known competing pathways: the Alr-mediated isomerization to D-lysine or catabolism of 5AVA to glutarate via the action of DavTD (Figure 4-6A) [75].

To investigate the relative contributions of pathways that contribute to either lactam hydrolysis or loss of substrate, we expressed the *davBA-ORF26* pathway via a arabinose-inducible broad host range vector pBADT in backgrounds where these pathways had been sequentially deleted. Wild type *P. putida* produced 0.43 mg/L valerolactam after 24 hours, but no valerolactam could be detected after 48 hours, presumably due to host consumption (Figure 4-6B). Simple deletion of the *oplBA* locus resulted in a 10-fold increase of production at 24 hours to 4.47 mg/L and a 48-hour titer of 9.27 mg/L (Figure 4-6B). Additional deletion of *davT* resulted in an increase of titer to 19.29 mg/L and 85.19 mg/L, and by deleting the amino acid racemase *alr* titers increased to 63.66 mg/L and 91.97 mg/L at 24 and 48 hours, respectively (Figure 4-6B).

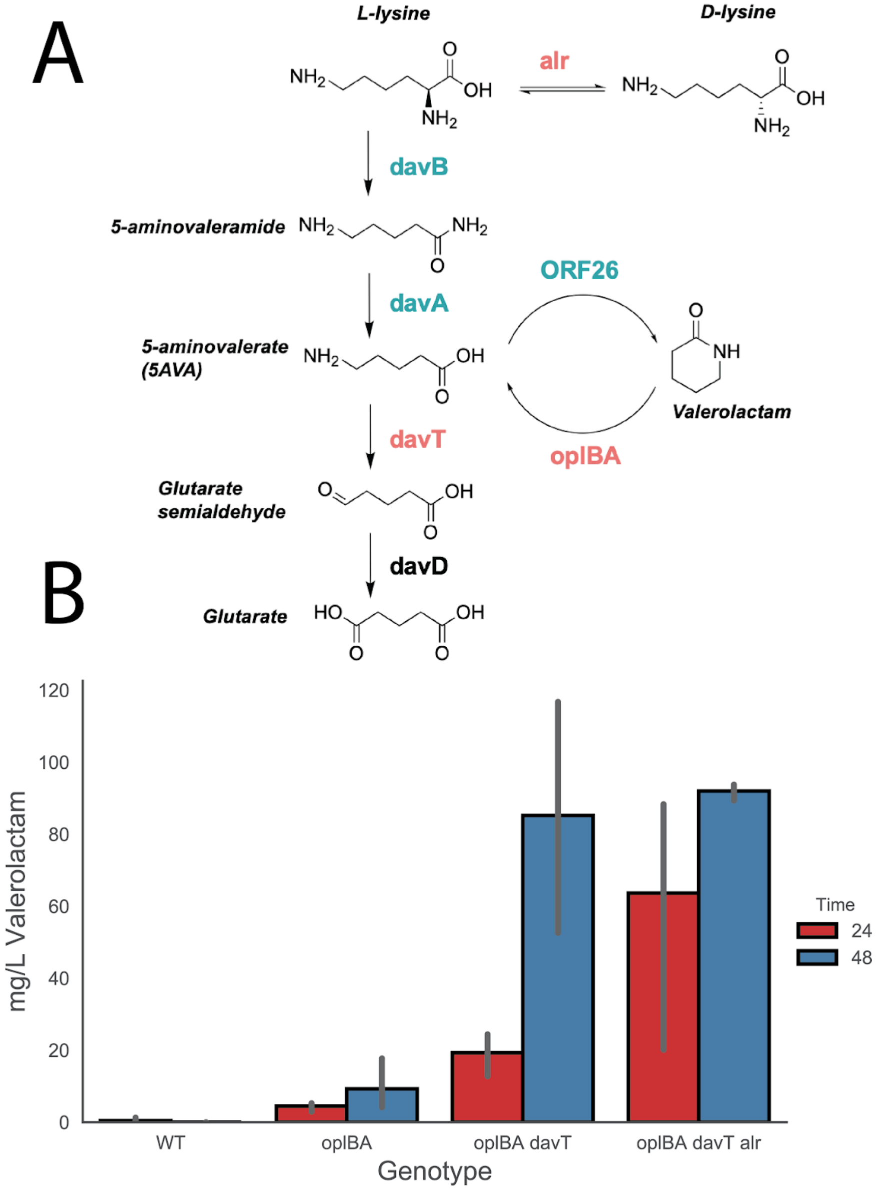


Figure 4-6 Production of valerolactam in *P. putida*. (A) Design of valerolactam overproducing *P. putida*. The biosynthetic pathway genes are shown in green and were overexpressed heterologously from a pBBR ori plasmid using an arabinose inducible promoter. Pathways that catabolize products or divert precursors are shown in red. (B) Valerolactam production from different *P. putida* strains grown in LB medium supplemented with 25 mM L-lysine and 0.2% (w/v) arabinose at 24 and 48 hours post inoculation. Error bars show 95% cI, n=3.

4.4. Discussion

Recent economic analyses highlight the necessity of lignin valorization to create a sustainable bioeconomy [102,103]. With its robust aromatic metabolism combined with novel methods of biomass deconstruction, *P. putid*a has great potential to convert lignocellulosic biomass to value-added products [104,105]. Though the ability of *P. putida* to catabolize many carbon sources is often viewed as an asset, it has well documented metabolic pathways to degrade many compounds that metabolic engineers may wish to produce such as levulinic acid [16], various alcohols [106], and the diacid glutarate [9]. As these catabolic phenotypes are encountered it will be critical to rapidly identify the offending genomic loci.

The recent surge in development of functional genomics techniques has dramatically increased the throughput at which we can identify the genetic basis of unknown metabolism. RB-TnSeq has been used to uncover novel glutarate and levulinic acid metabolism, though was ineffective at identifying the *oplBA* locus [16,75]. Proteomics techniques have also grown more robust, and were used in *P. jessenii* to predict a route of caprolactam catabolism [100,107]. Here, proteomics also proved to be an effective means of identifying the enzyme responsible for the hydrolysis of the lactam. Our proteomics results showed that OplBA was specifically expressed when grown on valerolactam, but not 5AVA. These results suggest that *P. putida* may have lactam-specific transcription factors which could be developed into valuable tools for metabolic engineering if identified.

The inability of RB-TnSeq to identify these genes is curious as single deletion mutants were unable to grow on valerolactam. A possible explanation for this is that OplBA may be secreted, which would create a public pool of 5AVA which *oplBA* mutants could still utilize. However, OplBA orthologs have been shown to be ATP-dependent [101], which would be inconsistent with this extracellular localization. Unfortunately, our attempts to characterize OplBA *in vitro* were unsuccessful, preventing us from identifying the substrate requirements of the enzyme. Though we were unable to reconstitute the activity of OplBA *in vitro*, deletion of *oplBA* did not prevent the degradation of butyrolactam which is a 5-membered lactam ring. The annotated function of the *oplBA* loci is a 5-oxoprolinase, which hydrolyzes the 5-membered lactam ring of 5-oxoproline. These results may suggest that the OplBA may function naturally as something other than a 5-oxoprolinase. More work will be necessary to resolve the results of our RB-TnSeq and feeding experiments to elucidate the biochemical requirements of OplBA as well as to better understand its cellular localization.

Without deleting *oplBA P. putida* is able to metabolize up to 10 mM valerolactam in rich media after 24 hours, and simple deletion of these genes increases production 10-fold after 24 hours of fermentation. Subsequent deletion of the 5AVA transaminase *davT* and the racemase *alr* resulted in 48-hour titers of ~90 mg/L, whereas there was no detectable valerolactam production in wild-type cultures at this time point. To achieve these titers we fed in 25 mM L-lysine (3.65 g/L). Previous work in *E. coli* achieved titers of ~200 mg/L by feeding 1 g/L lysine and ~300 mg/L by feeding 5 g/L after 48 hours [8]. Our results indicate that with significant host engineering, *P. putida* can produce titers approaching those of model organisms. Optimization of pathway expression could narrow this gap even further and should be a focus of future efforts.

While this initial work is encouraging, it still requires the feeding of L-lysine in rich media for conversion to valerolactam. Ideally, engineering *P. putida* would be able to metabolize lignin hydrolysis products directly to L-lysine on the way to the final product. A great deal of work has been conducted to elucidate the complex catabolism of lysine in *P. putida* [9,75], yet relatively little work has been done to increase flux to lysine within the bacterium. While there has been little to divert flux to L-lysine in *P. putida*, there is a wealth of evidence in other bacteria where high titers of intracellular lysine have been achieved [108,109]. The ever-increasing body of research to characterize the sprawling metabolism of *P. putida* will greatly aid in future efforts of efficient production of valerolactam from lignocellulosic feedstocks.

4.5. Materials and methods

4.5.1. Chemicals and media

General *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 [39]. Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), gentamicin (30 mg/L, Fisher Scientific, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA), when indicated. All other compounds were purchased through Sigma Aldrich (Sigma Aldrich, USA) .

4.5.2. Strains and plasmids

All bacterial strains and plasmids used in this work are listed in Table 1. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/456). All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software [40–42]. Plasmids were assembled via Gibson Assembly using standard protocols [43], or Golden Gate Assembly using standard protocols [44]. 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 [75].

Table 4-1 Strains and plasmids used in this study.

Strain	**JBEI Part ID**	**Reference**
E. coli DH10B		[90]
E. coli BL21(DE3)		Novagen

P. putida KT2440		ATCC 47054
P. putida ΔdavT		[110]
P. putida Δ*oplBA*	JBEI-104285	This work
P. putida Δ*oplBA*Δ*davT*	JBEI-104286	This work
P. putida Δ*oplBA*Δ*davT*Δ*alr*	JBEI-104287	This work
Plasmids		
pET28		Novagen
pET28 oplA	JBEI-104336	This work
pET28 oplB	JBEI-104337	This work
pBADT		[91]
pBADT-davBA-ORF26	JBEI-104356	This work
pMQ30		[45]
pMQ30 oplBA	JBEI-104355	This work
pMQ30 alr	JBEI-104354	This work
pMQ30 davT		[110]

4.5.3. Plate based growth assays

Growth studies of bacterial strains were conducted a 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 media without any added carbon and diluted 1:100 into 500 μL 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 an Biotek Synergy 4 plate reader (BioTek, USA) at 30 °C with fast continuous shaking. Optical density was measured at 600 nm.

4.5.4. Production assays and lactam quantification

To assess valerolactam production in strains of *P. putida* overnight cultures of strains harboring pBADT-*davBA-ORF26* were grown in 3 mL of LB supplemented with kanamycin and grown at 30 °C. Production cultures of 10 mL of LB supplemented with kanamycin, 25mM L-lysine, and 0.2% w/v arabinose were then inoculated 1:100 with overnight cultures and then grown at 30 °C shaking at 250 rpm. Samples for valerolactam production were taken at 24- and 48-hours post-inoculation, with 200 μL of culture being quenched with an equal volume of ice cold methanol and then stored at -20 °C until analysis.

For measurement of lactams, liquid chromatographic separation was conducted at 20°C with a Kinetex HILIC column (50-mm length, 4.6-mm internal diameter, 2.6-*μ*m particle size; Phenomenex, Torrance, CA) using a 1260 Series HPLC system (Agilent Technologies, Santa Clara, CA, USA). The injection volume for each measurement was 5 *μ*L. The mobile phase was composed of 10 mM ammonium formate and 0.07% formic acid in water (solvent A) and 10 mM ammonium formate and 0.07% formic acid in 90% acetonitrile and 10% water (solvent B) (HPLC grade, Honeywell Burdick & Jackson, CA, USA). High purity ammonium formate and formic acid (98-100% chemical purity) were purchased from Sigma-Aldrich, St. Louis, MO, USA. Lactams were separated with the following gradient: decreased from 90%B to 70%B in 2 min, held at 70%B for 0.75 min, decreased from 70%B to 40%B in 0.25 min, held at 40%B for 1.25 min, increased from 40%B to 90%B for 0.25 min, held at 90%B for 1 min. The flow rate was varied as follows: 0.6 mL/min for 3.25 min, increased from 0.6 mL/min to 1 mL/min in 0.25 min, and held at 1 mL/min for 2 min. The total run time was 5.5 min.

The HPLC system was coupled to an Agilent Technologies 6520 quadrupole time-of-flight mass spectrometer (QTOF MS) with a 1:6 post-column split. Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and nebulizing gases were set to 12 L/min and 30 lb/in^2, respectively, and a drying gas temperature of 350°C was used throughout. Fragmentor, skimmer and OCT 1 RF voltages were set to 100 V, 50 V and 300 V, respectively. Electrospray ionization (ESI) was conducted in the positive-ion mode for the detection of $[M + H]^+$ ions with a capillary voltage of 4000 V. The collision energy voltage was set to 0 V. MS experiments were carried out in the full-scan mode (75–1100 *m/z*) at 0.86 spectra/s. The QTOF-MS system was tuned with the Agilent ESI-L Low concentration tuning mix in the range of 50-1700 *m/z*. Lactams were quantified by comparison with 8-point calibration curves of authentic chemical standards from 0.78125 *μ*M to 100 *μ*M. R^2 coefficients of ≥0.99[EB1] were achieved for the calibration curves. Data acquisition was performed by Agilent MassHunter Workstation (version B.05.00), qualitative assessment by Agilent MassHunter Qualitative Analysis (version B.05.00 or B.06.00), and data curation by Agilent Profinder (version B.08.00)

4.5.5. RB-TnSeq and Proteomics Analysis

RB-TnSeq experiments utilized *P. putida* library JBEI-1 which has been described previously [75]. Libraries of JBEI-1 were thawed on ice, diluted into 25 mL of LB medium with kanamycin and then grown to an OD_{600} of 0.5 at 30 °C at which point three 1 mL aliquots were removed, pelleted, and stored at -80 °C. Libraries were then washed once in MOPS minimal medium with no carbon source, and then diluted 1:50 in MOPS minimal medium with 10 mM valerolactam. Cells were grown in 500 µL of medium in 48-well plates (Falcon, 353072). Plates were sealed with a gas-permeable microplate adhesive film (VWR, USA), and then grown at 30 °C in a Tecan Infinite F200 microplate reader (Tecan Life Sciences, San Jose, CA), with shaking at 200 rpm. Two 500 µL aliquots were combined, pelleted, and stored at -80 °C until BarSeq analysis, which was performed as previously described [14,16]. All fitness data is publicly available at http://fit.genomics.lbl.gov.

Secreted proteins of *P. putida* were prepared by growing 500 mL of culture in MOPS minimal medium supplemented with either 10 mM glucose or 10 mM valerolactam for 24 hours at 30 °C, which were subsequently pelleted and filtered through a 0.22 µm filter and then concentrated 100x via a 10 kD MW cutoff filter. Cultures for intracellular proteomics analysis were grown in 10 mL cultures in the same conditions on either glucose, 5AVA, or valerolactam and were then pelleted and stored at -80 °C until sample workup and proteomic analysis. Proteins from secreted and intracellular samples were desalted and isolated using a variation of a previously-described chloroform/methanol extraction protocol [111]. For secreted proteins, 100-200 µL of the concentrated protein sample was used; for intracellular samples, cell pellets were thoroughly resuspended in 100 µL HPLC water. Then, the following reagents were added to each sample in sequential order with thorough vortexing after each addition: 400 µL of HPLC grade methanol, 100 µL of HPLC grade chloroform, 300 µL of HPLC grade water. Samples were centrifuged for 1 minute at ~21,000g in order to promote phase separation. After centrifugation, the entirety of the top layer (water and methanol) was removed and discarded, leaving on the protein pellet and chloroform layer remaining. Another 300 µL of HPLC grade methanol was added, then the samples were vortexed and centrifuged again for 2 minutes at ~21,000g. The remaining liquid was then removed and discarded, and the cell pellets were allowed to dry in a fume hood for 5 minutes. Protein pellets were then resuspended in freshly-prepared 100mM ammonium bicarbonate buffer in HPLC water containing 20% HPLC methanol. Protein concentrations in the resuspended samples were quantified using a DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). After quantification, 100 µg of protein was transferred to a PCR strip and tris(2-carboxyethyl)phosphine was added to a final concentration of 5mM. Samples were incubated at 22°C for 30 minutes, after which iodoacetamide was added (final concentration 10mM). Samples were again incubated at 22°C in the dark for 30 minutes. Finally, trypsin was added to a final ratio of 1:25 w/w trypsin:sample, and samples were digested at 37°C for 5-8 hours before being transferred to conical LC vials for LC-MS analysis. Peptides prepared for shotgun proteomic experiments were analyzed by using an Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA) coupled to an Agilent 1290 UHPLC system as described previously [47]. 20 µg of peptides were separated on a Sigma–Aldrich Ascentis Peptides ES-C18 column (2.1 mm × 100 mm, 2.7 µm particle size, operated at 60°C) at a 0.400 mL/min flow rate and eluted with the following gradient:

initial condition was 98% solvent A (0.1% formic acid) and 2% solvent B (99.9% acetonitrile, 0.1% formic acid). Solvent B was increased to 35% over 30 min, and then increased to 80% over 2 min, then held for 6 min, followed by a ramp back down to 2% B over 1 min where it was held for 4 min to re-equilibrate the column to original conditions. Peptides were introduced to the mass spectrometer from the LC by using a Jet Stream source (Agilent Technologies) operating in positive-ion mode (3,500 V). Source parameters employed gas temp (250°C), drying gas (14 L/min), nebulizer (35 psig), sheath gas temp (250°C), sheath gas flow (11 L/min), VCap (3,500 V), fragmentor (180 V), OCT 1 RF Vpp (750 V). The data were acquired with Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.06.01 operating in Auto MS/MS mode whereby the 20 most intense ions (charge states, 2–5) within 300–1,400 m/z mass range above a threshold of 1,500 counts were selected for MS/MS analysis. MS/MS spectra (100–1,700 m/z) were collected with the quadrupole set to "Medium" resolution and were acquired until 45,000 total counts were collected or for a maximum accumulation time of 333 ms. Former parent ions were excluded for 0.1 min following MS/MS acquisition. The acquired data were exported as mgf files and searched against the latest *P. putida* KT2440 protein database with Mascot search engine version 2.3.02 (Matrix Science). The resulting search results were filtered and analyzed by Scaffold v 4.3.0 (Proteome Software Inc.).

4.5.6. Protein purification and biochemical analysis of OplBA

Both *oplB* and *oplA* were cloned into the expression vector pET28 harboring N-terminal 6x-histidine purification tags. Protein expression and purification were carried out as described previously [75]. To characterize activity of OplBA we used conditions that were previously described to characterize 5-oxoprolinase with minor changes [112]. Briefly, 10 μM of each OplB and OplA or boiled controls were incubated for 4 hours at 30 °C with 2 mM valerolactam in 100 mM Tris-HCl pH 7.0, 4 mM $MgCl_2$, with or without 2 mM ATP. Reactions were quenched with ice-cold methanol, filtered through a 3kDa MWCO filter, diluted 40-fold with 50% methanol/50% H2O, and stored at - 20 °C until analysis via LC-MS.

4.5.7. Bioinformatic analyses

All statistical analyses were carried out using either the Python Scipy or Numpy libraries [113,114]. For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder as implemented on IQ-tree [58] 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 OplBA were identified with CORASON-BGC [60] and manually colored and annotated.

4.6. Miscellaneous

4.6.1. Acknowledgements

We thank Jesus Barajas for his careful reading of this manuscript and helpful suggestions during preparation. We also thank Morgan Price for assistance in analyzing

RB-TnSeq data. We would like to thank the Amgen Scholars program for providing funding for Alexandria Velasquez.

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, 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. 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).

4.6.2. Collaborator contributions

Conceptualization, M.G.T.; Methodology, M.G.T., L.E.V.,J.M.B, P.C.M.,E.E.K.B.; Investigation, M.G.T., L.E.V., J.M.B, P.C.M, V.T.B, W.A.S., A.N.P., A.E.V.; Writing – Original Draft, M.G.T.; Writing – Review and Editing, All authors.; Resources and supervision, C.J.P., A.M.D., J.D.K.

4.6.3. Competing interests

JDK has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

Chapter 5. A Storm of Sensing: Identification, characterization and application of a highly sensitive lactam biosensor from *Pseudomonas putida*

Including material from published work from **Mitchell G Thompson**, Allison N Pearson, Jesus F Barajas, Pablo Cruz-Morales, Nima Sedaghatian, Zak Costello, Megan E Garber, Matthew R Incha, Luis E Valencia, Edward Baidoo, Hector Garcia Martin, Aindrila Mukhopadhyay, Jay D Keasling, "Identification, characterization and application of a highly sensitive lactam biosensor from *Pseudomonas putida*". BioRxiv, 2019

5.1. Abstract

Caprolactam is an important polymer precursor to nylon traditionally derived from petroleum and produced on a scale of 5 million tons per year. Current biological pathways for the production of caprolactam are inefficient with titers not exceeding 2 mg/L, necessitating novel pathways for its production. As development of novel metabolic routes often require thousands of designs and result in low product titers, a highly sensitive biosensor for the final product has the potential to rapidly speed up development times. Here we report a highly sensitive biosensor for valerolactam and caprolactam from *Pseudomonas putida* KT2440 which is >1000x more sensitive to exogenous ligand than previously reported sensors. Manipulating the expression of the sensor *oplR* (PP_3516) substantially altered the sensing parameters, with various vectors showing K_d values ranging from 700 nM to 1.2 mM. Our most sensitive construct was able to detect *in vivo* production of caprolactam above background at 11 μg/L. The high sensitivity and range of OplR is a powerful tool towards the development of novel routes to the biological synthesis of caprolactam.

5.2. Introduction

Caprolactam is an important chemical precursor to the polymer nylon 6, with a global demand approximately 5 million tons per year [115]. Currently the majority of caprolactam is synthesized from cyclohexanone, which in turn derived from petroleum [115]. In addition to being inherently unsustainable, the chemical process to synthesize caprolactam requires toxic reagents and produces unwanted byproducts such as ammonium sulfate [115]. Multiple attempts have been made to produce caprolactam biologically, however the highest titers achieved to date are no greater than 1-2 mg/L [115,115]. All current published strategies to produce caprolactam rely on the cyclization of 6-aminocaproic acid (6ACA) via a promiscuous acyl-coA ligase [115,115]. While similar strategies to make C4 butyrolactam and C5 valerolactam produce gram per liter titers of each, it is thought that both the entropy and enthalpy properties of 7-membered ring formation of caprolactam present an inherent barrier to the cyclization of 6ACA [115,115]. Clearly, novel routes to a renewable biological production of caprolactam are needed.

Genetically encoded biosensors can accelerate metabolic engineering efforts in many ways, the foremost of which is the ability to rapidly screen for desirable phenotypes beyond the throughput of analytical chemistry [115]. Multiple papers have reported transcription factors or riboswitches that respond to lactams with varying degrees of sensitivity and specificity [115,115,115]. One feature that unifies currently available lactam biosensors is that lactams are not the native ligand for any of the corresponding biosensor systems. Zhang et al. used ChnR from *Acinetobacter* sp. Strain NCIMB 9871 to sense multiple lactams with all ligands showing having a K_d of > 30 mM. However, ChnR natively regulates cyclohexanol catabolism and is activated by its natural ligand, cyclohexanone, at sub-millimolar concentrations [115,115]. Yeom et al. selected mutants of the NitR biosensor to sense caprolactam at concentrations as low as 100 μM when added exogenously, and leveraged this sensor to identify novel cyclases to convert 6ACA to caprolactam [115]. Natively, NitR regulates nitrile catabolism in *Rhodococcus rhodochrous* J1, and is responsive to micromolar concentrations of isovalernitrile [115,115]. It is reasonable to assume then that if natural lactam catabolic pathways are identified, highly sensitive biosensors could also be found.

Recently two groups have identified pathways of lactam degradation in both *P. putida* and *Pseudomonas jessenii* [115,115]. Work in *P. putida* demonstrated that the enzyme OplBA, putatively responsible for the hydrolysis of valerolactam, is upregulated by the lactam but not its cognate ω-amino acid [115]. These findings suggest there may be a lactam-sensitive transcription factor controlling the expression of the hydrolytic enzyme that can be used as a biosensor. In this work we demonstrate that the AraC-type regulator directly downstream of *oplBA* is indeed a lactam biosensor with unprecedented sensitivity towards both valerolactam and caprolactam. Through rational engineering we developed a suite of lactam sensing plasmids with dissociation constants ranging from 700 nM to 1.2 mM, allowing for a dramatic dynamic range of sensing. To demonstrate the utility of these sensors, we show that they are able to detect low titers of caprolactam produced biologically in an *Escherichia coli* system.

5.3. Results

5.3.1. Identification and development of *oplR* as a lactam biosensor

In *P. putida*, the *oplBA* locus is flanked by the LysR-family regulator PP_3513 upstream, and the AraC-family regulator PP_3516 downstream. To infer if either of these transcription factors regulates *oplBA*, we used publicly available fitness data to assess if either regulator is cofit with *oplBA* (http://fit.genomics.lbl.gov) [115]. While no cofitness was observed between PP_3513 and either *oplA* or *oplB*, PP_3516 was highly cofit with both genes (0.91:PP_3514, 0.80:PP_3515). To examine the hypothesis that PP_3516 is the regulator of *oplBA*, we examined the genomic contexts of the oxoprolinase loci across multiple bacteria (Figure 5-1). While both regulators were found in closely related species, in more distantly related *Pseudomonads*, such as *Pseudomonas aeruginosa*, only the AraC-family regulator is conserved (Figure 5-1). Using Multiple EM for Motif Elicitation (MEME) [115] we attempted to identify conserved putative binding sites upstream of *oplBA* as well as PP_3516. In closely related *Pseudomonads*, including *P. aeruginosa*, a conserved motif was identified upstream of both *oplBA* as well PP_3516 (Figure 5-1). Attempts to confirm this as the binding site of PP_3516 were hampered by the insolubility of PP_3516 when expressed heterologously (Figure 5-2), a common issue with AraC-family proteins [115,115].

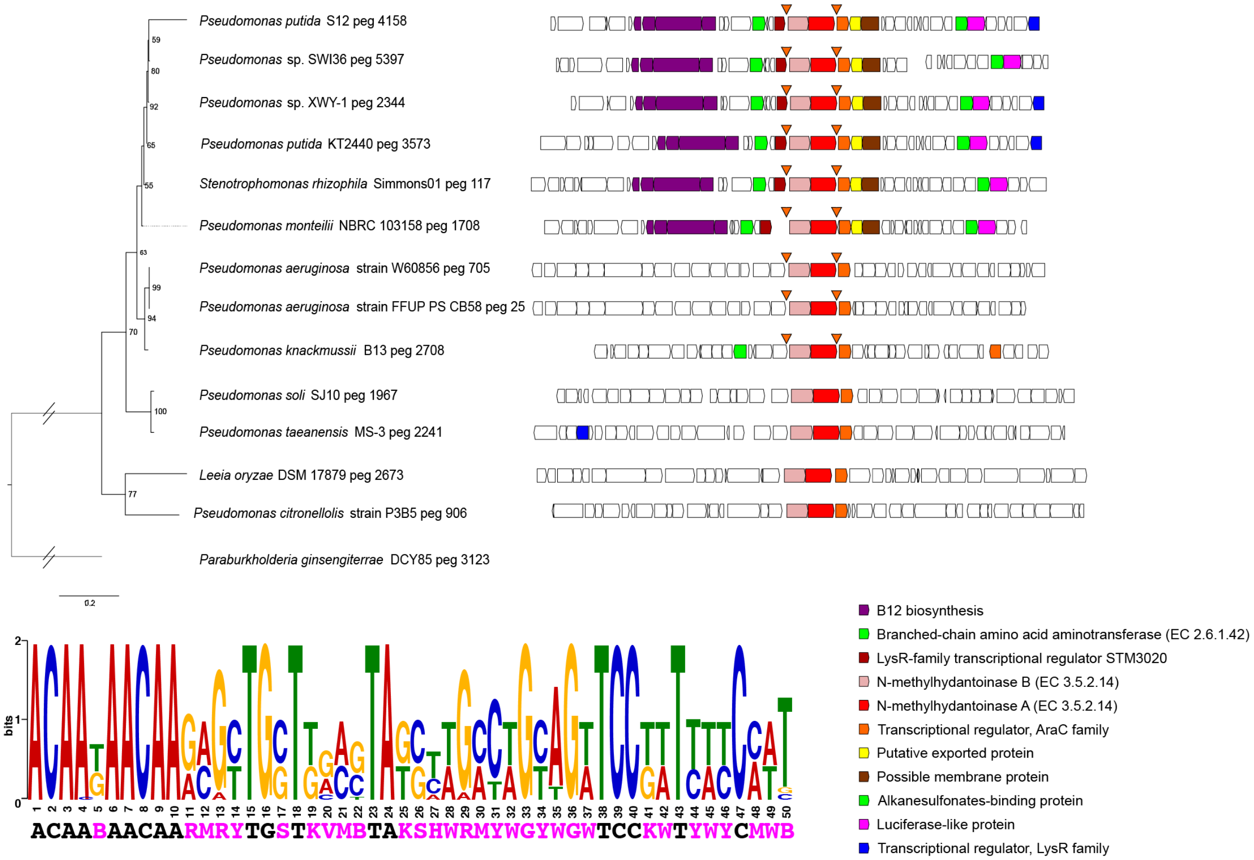


Figure 5-1 Synteny analysis of *oplBA* homologs across genomes of related species. - Triangles show the location of a conserved putative binding sites of OplR. A consensus putative OplR binding site is shown below**.**

In order to screen the ability of PP_3516 to sense lactams, we employed a two-plasmid test system wherein PP_3516 was cloned into an arabinose inducible medium-copy p15a plasmid and the 200-bp upstream of *oplB* (Figure 5-3A) was cloned upstream of RFP on a compatible medium-copy pBBR plasmid (Figure 5-3B). The relationship between fluorescence output, *oplR* expression, and ligand induction was tested via a checkerboard assay where the levels of arabinose and valerolactam were varied independently of one another in cultures of *E. coli* that harbored both the “reporter” and “regulator” plasmids. A dose-dependent expression of RFP was observed; both arabinose and valerolactam were required for high-level expression of RFP (Figure 5-3C). Our initial screen also showed that the fluorescence was far above background RFP expression even at the lowest concentration of valerolactam tested (10 μM).

To quantify the sensing properties of PP_3516 the regulator was induced with a fixed concentration of arabinose at 0.0125% w/v and concentrations of either valerolactam or caprolactam were varied from 1 mM to 12 nM (Figure 5-3D). PP_3516 proved to be extremely sensitive to both caprolactam and valerolactam with both ligands having a K_d ~5 uM, and limits of detection ≤ 12 nM (Table 5-1). Based on these findings we propose PP_3516 be named *oplR* for oxoprolinase regulator, which encodes a biosensor ~5000x more sensitive towards caprolactam than the next most sensitive published biosensor [115].

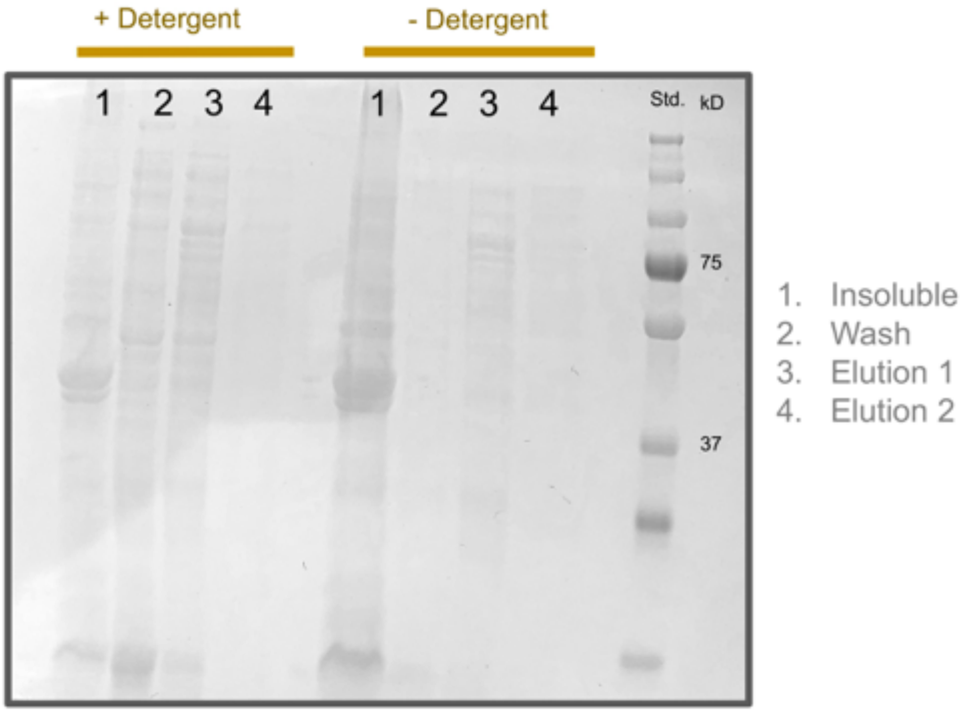


Figure 5-2 Insolubility of OplR expressed heterologously in *E. coli*. - OplR-6xHis expressed in the insoluble fraction (~40 kD) of *E. coli* in the presence and absence of the detergent Tween 20 (0.5% w/v).

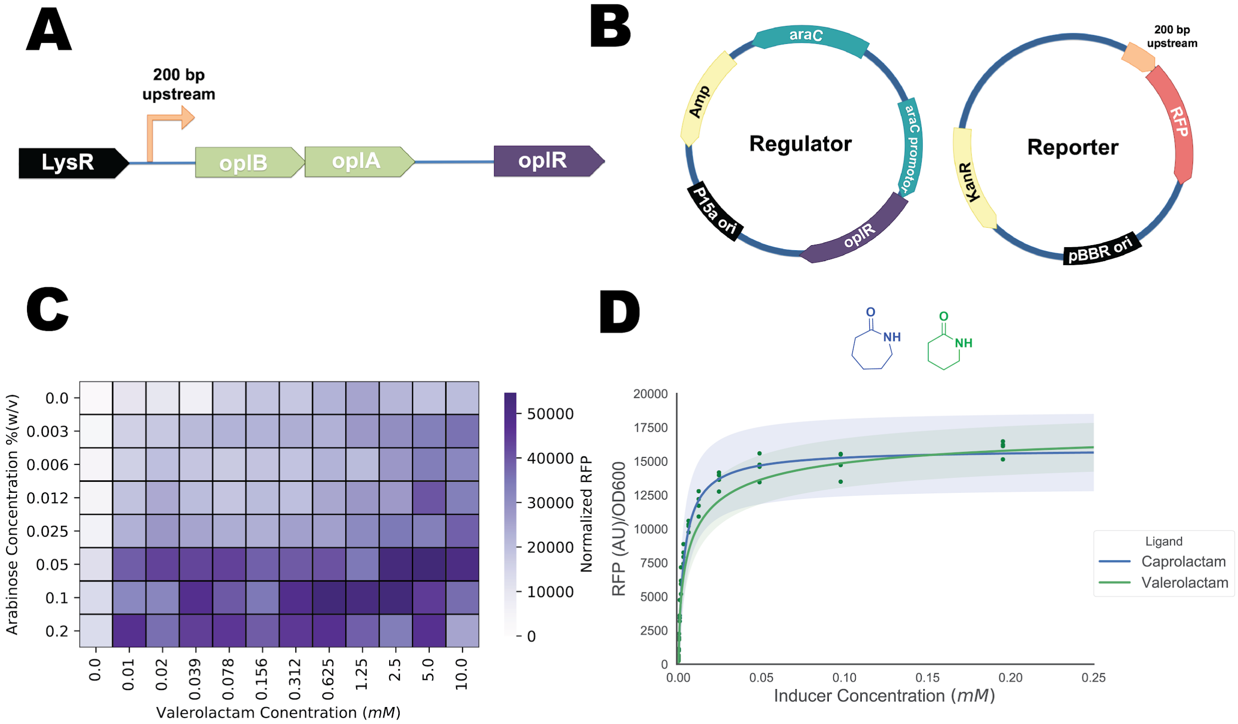


Figure 5-3 Development of an *oplR*-based lactam biosensor. A) Operonic structure of *oplBA* relative to *oplR* and putative promoter region used to construct the reporter. B) Diagram of the two-plasmid system used to test OplR lactam sensing C) Checkerboard screen of OplR biosensor two-plasmid system. Y-axis shows the concentration of arabinose (%w/v), X-axis shows the concentration of valerolactam (mM). Colorbar to right shows fluorescent intensity normalized to OD$_{600}$. D) Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for valerolactam and caprolactam. Points represent individual measurements. Shaded area represents (+/-) one standard deviation, n=4.

Table 5-1 Two-plasmid biosensor parameters. Max: Predicted maximal RFP, Hill Coef: Predicted Hill coefficient, K$_d$: Predicted K$_d$ in mM. LoD: Limit of detection determined experimentally. Standard deviation estimates are in parentheses.

Ligand	**Max**	**Hill Coef**	**K$_d$ (mM)**	**LoD (mM)**
Valerolactam	17817 (414)	0.62 (0.05)	0.007 (0.001)	≤ 0.000012
Caprolactam	15943 (483)	0.94 (0.16)	0.004 (0.001)	≤ 0.000012

5.3.2. Development of One Plasmid Systems

As a two-plasmid system is not convenient for engineering biological systems, we then sought to consolidate both the reporter and regulator into a single vector. Initial screening of OplR in the checkerboard assay suggested that varying the level of expression of OplR could dramatically influence the resulting sensing properties of the system (Figure 5-3C). We therefore constructed a family of plasmids, pLACSENS, where *oplR* was constitutively expressed from five promoters of increasing strength divergent from the RFP reporter (Figure 5-4A).

Biosensor performance of pLACSENS vectors was then assessed using valerolactam as a ligand across concentrations from 12.5 mM to 12.5 nM Figure (Figure 3B). As seen in the two-plasmid system, by varying the strength of *oplR* expression the characteristics of the biosensor changed dramatically (Table 5-2). The most sensitive vector, pLACSENS3, had an experimentally determined limit of detection (LoD) of $\leq$12 nM, and a K_d of 700 nM. The least sensitive vector, pLACSENS5, had a limit of detection of 1.5 μM, and a K_d of 1.5 mM, but drives the expression of *oplR* with the strongest predicted promoter. The maximal RFP expression also varied greatly with *oplR* expression, with the highest and lowest RFP expression observed in pLACSENS3 and pLACSENS5, respectively (26300 vs. 793 RFP (AU)/OD_{600}). Induction over background expression was also highly variable; pLACSENS1 and pLACSENS4 were both maximally induced at ~250x over background, while pLACSENS5 was only induced ~25x over background.

Given the wide range of biosensing parameters within the pLACSENS vectors we sought to characterize which ligand concentration ranges each vector is most suited to detect. To do this, we utilized a recently developed model to probabilistically relate inducer concentration and fluorescence data via Markov Chain Monte Carlo (MCMC) sampling [115]. A resolution window is defined as the concentrations of inducer that are statistically compatible with the fluorescence data fit to the Hill function at a 95% confidence interval (cI). Resolution windows for each pLACSENS plasmid were graphed from their experimentally determined LoD to ligand concentrations compatible with 75% maximal fluorescence (Figure 5-4C). Overall, the family of vectors showed high relative resolution from 5 nM to 1 mM valerolactam, with pLACSENS3 having the highest resolution at the lowest concentrations and pLACSENS5 having the best resolution at high concentrations (Figure 5-4C).

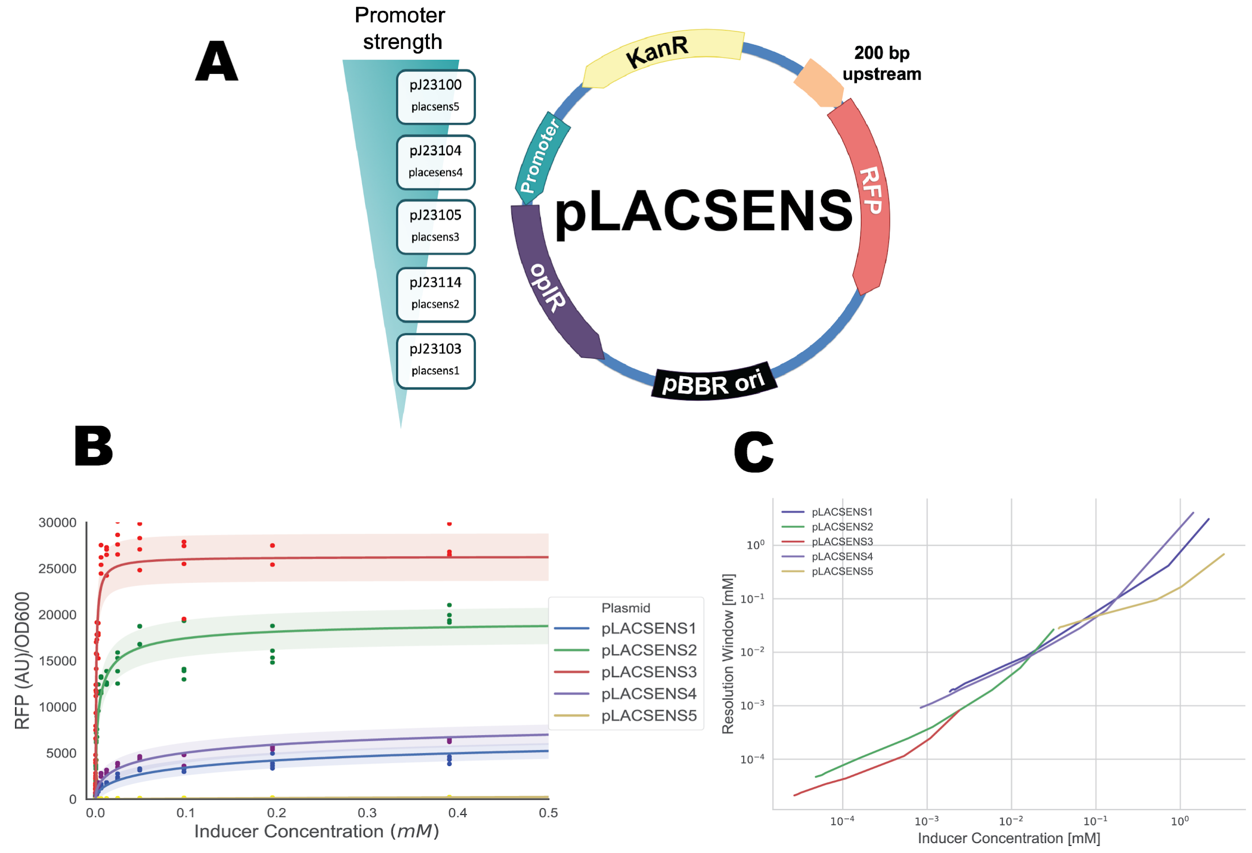


Figure 5-4 Development of one-plasmid *oplR*-based lactam biosensors. A) Diagram of the pLACSENS vector design, with the relative predicted strength of the promoter driving *oplR* on the left B) Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for valerolactam against all 5 pLACSENS vectors. Points represent individual measurements. Shaded area represents (+/-) one standard deviation, n=4. C) Resolution window of each pLACSENS vector over a range of valerolactam concentrations

Table 5-2 One-plasmid biosensor parameters. Max: Predicted maximal RFP, Hill Coef: Predicted Hill coefficient, K$_d$: Predicted K$_d$ in mM. LoD: Limit of detection determined experimentally. Standard deviation estimates are in parentheses. Induction: maximation induction over background based on experimental data

Plasmid	**Max**	**Hill Coef**	**K$_d$ (mM)**	**LoD (mM)**	**Induction**
pLACSENS1	8284 (358)	0.54 (0.05)	0.19 (0.05)	0.000024	253x

pLACSENS2	19614 (419)	0.63 (0.05)	0.004 (0.0006)	≤ 0.000012	93x
pLACSENS3	26300 (390)	0.90 (0.07)	0.0007 (0.0001)	≤ 0.000012	38x
pLACSENS4	9877 (421)	0.52 (0.05)	0.09 (0.02)	0.000024	256x
pLACSENS5	793 (25)	0.96 (0.10)	1.23 (0.15)	0.0015	25x

5.3.3. OplR is sensitive to valerolactam and caprolactam

To assess lactam specificity of OplR, we measured fluorescence induction of pLACSENS3 in the presence of lactams (laurolactam, caprolactam, valerolactam, butyrolactam, 5-oxoproline), ω-amino acids (4-aminobutyrate, 5AVA, 6ACA), the lactone valerolactone, and piperidine (Figure 5-5). Robust fluorescence induction was observed with caprolactam, valerolactam, and 5AVA (Table 5-3), but no other tested chemicals were capable of induction at the concentrations tested in the work (data not shown).

Given the dissimilarity in chemical structure of 5AVA and lactams, the ability of pLACSENS3 to detect 5AVA was surprising. Previously it has been shown in *E. coli* that 5AVA can spontaneously be converted into valerolactam by the activity of native acyl-coA ligases [115]. This led us to believe that rather than detecting 5AVA, pLACSENS3 was detecting valerolactam derived from the added 5AVA. LC-TOF analysis revealed that when 1 mM of 5AVA was fed to *E. coli* cultures harboring pLACSENS3 ~600 nM of valerolactam was produced, while no valerolactam could be detected in *E. coli* not supplemented with 5AVA (Figure 5-6). Valerolactam concentration at this level would explain the high levels of fluorescence observed when 5AVA is added to cultures.

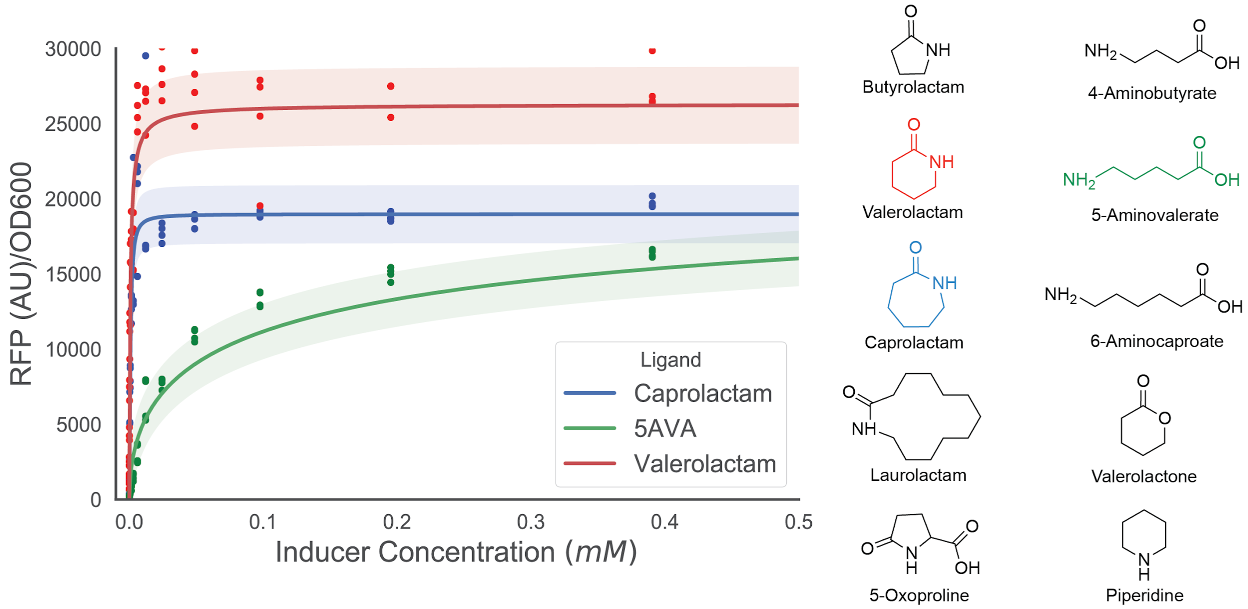


Figure 5-5 Ligand range of pLACSENS3. Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for ligands that activated pLACSENS3. Points represent individual measurements. Shaded area represents (+/-) one standard deviation, n=4. To the right, chemical structures of ligands that were tested.

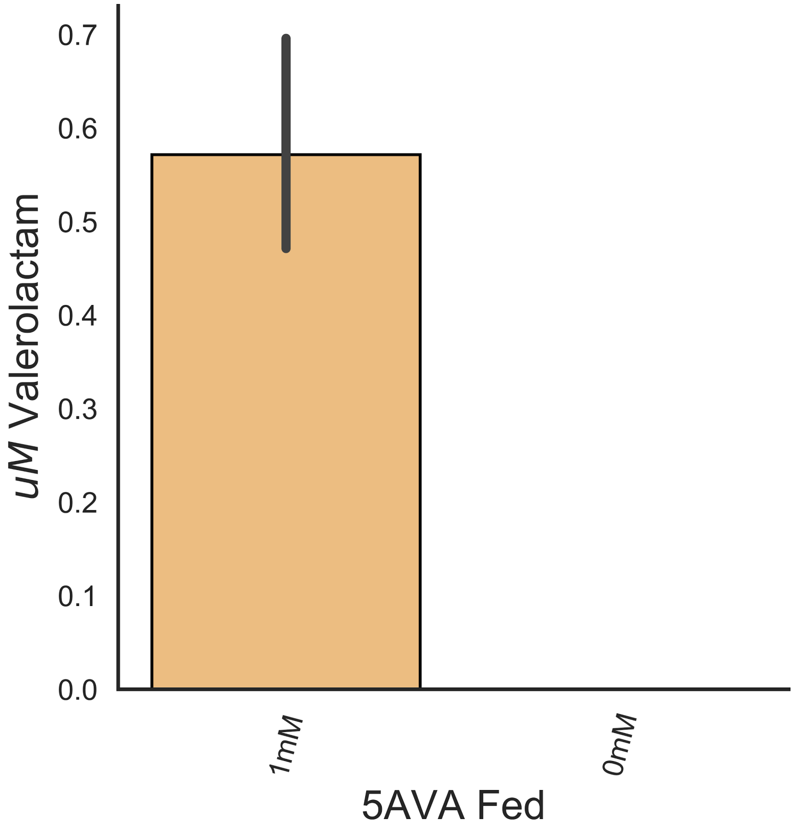


Figure 5-6 Valerolactam produced from 5AVA. Bars show the amount of valerolactam produced in E. coli harboring pLACSENS3 when fed either 1 mM or 0 mM 5AVA. Error bars represent 95% cI.

Table 5-3 pLACSENS3 biosensor parameters against different ligands. Max: Predicted maximal RFP, Hill Coef: Predicted Hill coefficient, K_d: Predicted K_d in mM.

Ligand	**Max**	**Hill Coef**	**K_d (mM)**
Valerolactam	26300 (390)	0.90 (0.07)	0.0007 (0.0001)
Caprolactam	18974 (272)	1.36 (0.14)	0.0009 (0.0001)
5AVA	23766 (1093)	0.52 (0.05)	0.12 (0.04)

5.3.4. Detection of caprolactam production *in vivo*

To demonstrate the utility of *oplR* based systems for metabolic engineering applications we introduced our most sensitive pLACSENS plasmid (pLACSENS3) into *E. coli* harboring various acyl-coA ligases on an IPTG-inducible orthogonal plasmid (Figure 5-7A). Multiple reports have utilized acyl-coA ligases to cyclize exogenously added 6ACA to produce low titers of caprolactam, with production ranging from 0.8 - 2 mg/L [115,115]. Strains harboring both plasmids were grown in LB medium with or without 10 mM 6ACA added for 24 hours. Cells grown in the presence of 6ACA demonstrated fluorescence greater than cells grown without 6ACA (Figure 5-7B). LC-MS analysis confirmed that no caprolactam was produced in cells grown without 6ACA, while cells grown with 6ACA had produced 11 ug/L (Figure 5-7B). The ability of pLACSENS3 to detect such minute production validates its utility as a means to rapidly and accurately screen novel pathways for caprolactam production.

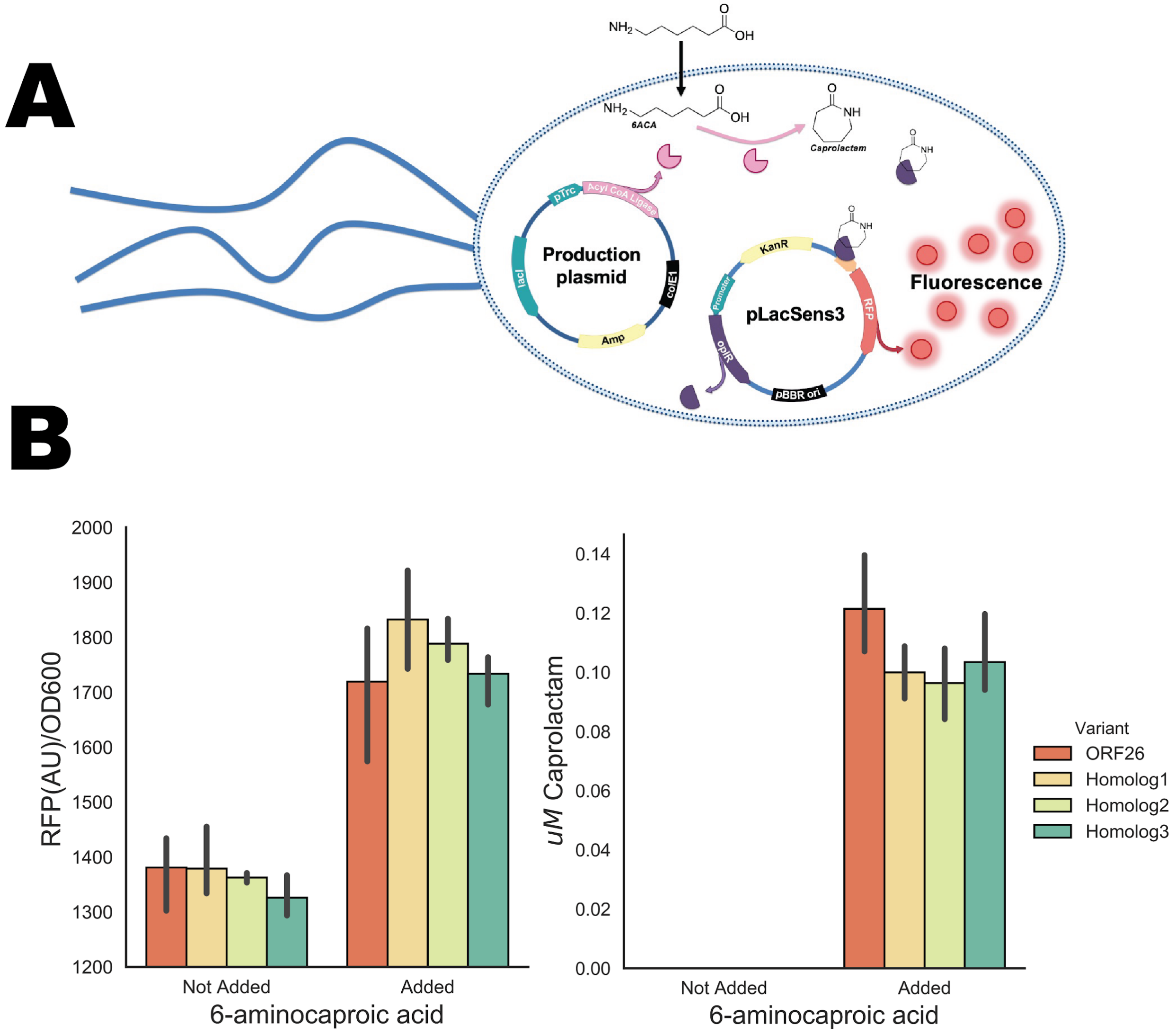


Figure 5-7 Detection of caprolactam production. A) Diagram of the pLACSENS vector design, with the relative predicted strength of the promoter driving *oplR* on the left B) Fluorescence data fit to the Hill equation to derive biosensor performance characteristics for valerolactam against all five pLACSENS vectors. Shaded area represents (+/-) one standard deviation, n=4. Different acyl-CoA ligases were tested for the *in vivo* cyclization of 6ACA to caprolactam. These included the previously studied ORF26 CoA-ligase (red) [115] and three additional homologs. Homolog 1 (yellow), Homolog 2 (light green) and homolog 3 (dark green).

5.4. Discussion

Previously, we have leveraged shotgun proteomics to infer local regulation within the lysine catabolism of *P. putida* [115,115,115,115]. From these data, multiple glutarate biosensors were engineered and used to measure relative metabolite amounts in the native host [115]. Here we again leveraged previously published proteomics data that showed OplBA to be specifically upregulated in the presence of valerolactam to develop a biosensor [115]. The AraC-family regulator tentatively named OplR was shown to have limits of detection for exogenously added valerolactam or caprolactam ≤ 12 nM when expressed at particular levels. This is remarkably more sensitive than the previously published valerolactam and caprolactam sensors, which showed limits of detection between 10-100 μM [115]. We attribute this high degree of sensitivity to the fact that, to our knowledge, it is the first sensor to control the catabolism of lactams specifically.

The range of detection of either caprolactam or valerolactam was highly dependent on the expression of *oplR*, with the K_d varying from 700 nM to 1.23 mM depending on the constitutive promoter used to drive expression. Previous work has also demonstrated that the sensing parameters of transcription factors can be readily modulated by changing the strength of transcription factor expression [115]. At both the highest and lowest levels of predicted expression OplR was less sensitive to lactams than when *oplR* was expressed more moderately. This may be explained by the fact that some AraC-family regulators are known to act as both positive and negative regulators, thus overexpression of OplR could result in hyper-repression [115]. Unfortunately, insoluble expression of OplR prevented further examination of the biochemical means of transcriptional control.

OplR was shown to be highly specific for valerolactam and caprolactam as ligands, but not butyrolactam or laurolactam. These findings are consistent with previous observations that showed *oplBA* mutants were not defective in their ability to hydrolyze butyrolactam. The inability of *oplR* to sense the annotated substrate of OplBA, 5-oxoproline, suggests that the natural function of the amidohydrolase is not that of a 5-oxoprolinase. While the 5-membered lactam rings tested here were not able to induce OplR-mediated expression, further work should be conducted to test derivatives of valerolactam and caprolactam. Additional functional groups added to these lactams could be used to produce both pharmaceutical precursors [115] and polymers with novel nylon properties. Identifying biological routes to their synthesis is highly attractive. Furthermore, recent work has shown that directed evolution may also be applied to broaden the ligand range [115], which could allow OplR to accommodate different lactam ligands.

Current published routes to caprolactam biosynthesis rely on the cyclization of 6ACA via the activity of promiscuous CoA-ligase activity [115,115]. While *oplR*-based biosensors may be able to aid in the selection of mutant acyl-coA ligases with enhanced activity, there remains a sizeable thermodynamic barrier for the cyclization of the 7-membered ring [115]. Novel routes to caprolactam or naturally occurring caprolactam-containing natural products that can mitigate this barrier would be ideal for high-level production. For example, a better understanding of pestalactam A-C biosynthesis may provide new and more efficient chemoenzymatic routes to 7-ring cyclization [115].

In addition to the value of OplR as a biosensor for metabolic engineering purposes, it may also be useful as an inducible system. Valerolactam is inexpensive (~$2/gram), and highly water soluble (291 mg/mL). The vector pLACSENS2 demonstrated ~100x induction over background, with a K_d toward valerolactam of 4 uM, and the second highest maximal expression of any single vector tested. Additional engineering of the system could improve upon these qualities. Furthermore, as OplR works well in *E. coli* and is derived from the distantly related *P. putida*, it may work well in other bacterial systems. Future work could evaluate which hosts are suitable for this inducible system.

Since naturally-occuring genetically encoded biosensors for chemicals of interest have the potential to be much more sensitive than those repurposed or evolved in the laboratory, it is critical to pursue rapid and efficient means of identifying them. The recent development of other high-throughput methods to associate genotypes with phenotypes, such as RB-TnSeq and CRISPRi, has created a large reservoir of data that can be easily mined for transcription factors useful in synthetic biology [115,115,115,115]. Bacteria often locally regulate catabolism, thus allowing inference of genetic control by adjacent transcription factors once a catabolic pathway has been discovered. Empirical evidence of catabolism is critical for assigning transcription factor function as orthologous transcription is often utilized differently by different species [115]. Future work to generalize approaches to develop useful synthetic biology tools from genome-wide fitness data has the potential to dramatically increase the genetically encoded chemical sensor space.

5.5. Materials and methods

5.5.1. Chemicals and media

General *E. coli* cultures were grown in Lysogeny Broth (LB) Miller medium (BD Biosciences, USA) at 37 °C. When indicated, *E. coli* was also grown on EZ-RICH medium (Teknova, Hollister, CA) supplemented with 1% glucose. Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA), when indicated. All other compounds were purchased through Sigma Aldrich (Sigma Aldrich, USA).

5.5.2. Strains and plasmids

All bacterial strains and plasmids used in this work are listed in Table 4. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/XXX). All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software [115,115,115]. Plasmids were assembled via Gibson Assembly using standard protocols [115], or Golden Gate Assembly using standard protocols [115]. Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA).

Table 5-4 Strains and plasmids used.

Strain	**JBEI Part ID**	**Reference**
E. coli DH10B		[115]
E. coli BL21(DE3)		Novagen
P. putida KT2440		ATCC 47054
Plasmids		
pBADT		[115]
pBADT-PP_3516p-RFP	JBEI-104519	This work
pBbA8a-PP_3516	JBEI-104517	This work
pLacSens1	JBEI-104506	This work
pLacSens2	JBEI-104505	This work
pLacSens3	JBEI-104504	This work
pLacSens4	JBEI-104503	This work
pLacSens5	JBEI-104502	This work
pBbE7a-ORF26	JBEI-104322	This work
pBbE7a-G2NX2_STRVO	JBEI-104323	This work
pBbE7a-A0A0D4DX08_9ACTN	JBEI-104324	This work
pBbE7a-A0A0C1VDH3_9ACTN	JBEI-104325	This work

5.5.3. Expression and purification of proteins

Proteins were purified as described previously [115]. The cultures were grown at 37 °C until the OD_{600} nm reached 0.8 and cooled on ice for 20 min. 1 mM IPTG was added to induce overexpression for 16 h at 18 °C. The cells were harvested by centrifugation (8000g, 10 min, 4 °C), resuspended in 40 mL of lysis buffer (50 mM HEPES, pH 8.0, 0.3 M NaCl, 10% glycerol (v/v) and 10 mM imidazole), and lysed by sonication on ice. Cellular debris was removed by centrifugation (20000g, 60 min, 4 °C).

The supernatant was applied to a fritted column containing Ni-NTA resin (Qiagen, USA) and the proteins were purified using the manufacturer's instructions. Fractions were collected and analyzed via SDS-PAGE.

5.5.4. Fluorescence biosensor assays

All assays were conducted in 96-deep well plates (Corning Costar, 3960), with each well containing 500 μL of medium with appropriate ligands, antibiotics, and/or inducers inoculated at 1% v/v from overnight cultures. Plates were sealed with AeraSeal film (Excel Scientific, AC1201-02 and incubated at 37 C in a 250 rpm shaker rack. After 24 hours, 100 μL from each well was aliquoted into a black, clear-bottom 96-well plate for measurements of optical density and fluorescence using an Infinite F200 (Tecan Life Sciences, San Jose, CA) plate reader. Optical density was measured at 600 nm (OD_{600}), while fluorescence was measured using an excitation wavelength of 535 nm, an emission wavelength of 620 nm, and a manually set gain of 60.

For the checkerboard assay of the two-plasmid system, LB medium supplemented with both kanamycin and carbenicillin was inoculated with E. coli containing the regulator and reporter plasmids grown overnight in the same medium. Arabinose concentration was decreased from 0.2 to 0% w/v along the y-axis, while valerolactam concentration was increased from 0-10 mM along the x-axis.

To find the Hill fit to the two-plasmid system, EZ-RICH medium containing kanamycin, carbenicillin, and 0.0125 w/v% arabinose was inoculated with an overnight culture of the two-plasmid system in LB medium supplemented with both kanamycin and carbenicillin. Both valerolactam and caprolactam were tested at concentrations ranging from 0 to 50 mM.

Characterization of the 5 variations of the one-plasmid pLACSENS plasmids was conducted using EZ-rich medium containing kanamycin and inoculated with overnight cultures of the appropriate E. coli strain. Valerolactam concentrations was varied from 0 to 50 mM. This same assay was repeated on pLACSENS3 with various ligands—using concentrations between 0 to 50 mM of caprolactam and 5AVA, and 0 to 10 mM of butyrolactam, 5-oxoproline, gamma-aminobutyric acid, 6ACA, valerolactone, and piperidine. The maximum concentration of laurolactam tested was 0.1 mM due to its poor solubility.

5.5.5. Production assays and analytical methods

Caprolactam production assays were carried out in 10 mL of LB medium supplemented with 10 mM 6ACA, as well as kanamycin, carbenicillin, and 1 mM IPTG. Cultures were inoculated 1:100 with overnight culture harboring both pLACSENS3 and expression vectors for acyl-coA ligases and grown at 30 °C shaking at 200 rpm for 24 hours. After 24 hours optical density was measured at 600 nm (OD_{600}), while fluorescence was measured using an excitation wavelength of 535 nm, an emission wavelength of 620 nm, and a manually set gain of 60. To sample for caprolactam production 200 μL of culture was quenched with an equal volume of ice-cold methanol and then stored at -20 °C until analysis.

Valerolactam and caprolactam were measured via LC-QTOF-MS as described previously [115]. Liquid chromatographic separation was conducted at 20°C with a Kinetex HILIC column (50-mm length, 4.6-mm internal diameter, 2.6-μm particle size; Phenomenex, Torrance, CA) using a 1260 Series HPLC system (Agilent Technologies, Santa Clara, CA, USA). The injection volume for each measurement was 5 μL. The mobile phase was composed of 10 mM ammonium formate and 0.07% formic acid in water (solvent A) and 10 mM ammonium formate and 0.07% formic acid in 90% acetonitrile and 10% water (solvent B) (HPLC grade, Honeywell Burdick & Jackson, CA, USA). High purity ammonium formate and formic acid (98-100% chemical purity) were purchased from Sigma-Aldrich, St. Louis, MO, USA. Lactams were separated with the following gradient: decreased from 90%B to 70%B in 2 min, held at 70%B for 0.75 min, decreased from 70%B to 40%B in 0.25 min, held at 40%B for 1.25 min, increased from 40%B to 90%B for 0.25 min, held at 90%B for 1 min. The flow rate was varied as follows: 0.6 mL/min for 3.25 min, increased from 0.6 mL/min to 1 mL/min in 0.25 min, and held at 1 mL/min for 2 min. The total run time was 5.5 min.

The HPLC system was coupled to an Agilent Technologies 6520 quadrupole time-of-flight mass spectrometer (QTOF MS) with a 1:6 post-column split. Nitrogen gas was used as both the nebulizing and drying gas to facilitate the production of gas-phase ions. The drying and nebulizing gases were set to 12 L/min and 30 lb/in^2, respectively, and a drying gas temperature of 350°C was used throughout. Fragmentor, skimmer and OCT 1 RF voltages were set to 100 V, 50 V and 300 V, respectively. Electrospray ionization (ESI) was conducted in the positive-ion mode for the detection of $[M + H]^+$ ions with a capillary voltage of 4000 V. The collision energy voltage was set to

0 V. MS experiments were carried out in the full-scan mode (75–1100 *m/z*) at 0.86 spectra/s. The QTOF-MS system was tuned with the Agilent ESI-L Low concentration tuning mix in the range of 50-1700 *m/z*. Lactams were quantified by comparison with 8-point calibration curves of authentic chemical standards from 0.78125 μM to 100 μM. R^2 coefficients of ≥0.99[EB1] were achieved for the calibration curves. Data acquisition was performed by Agilent MassHunter Workstation (version B.05.00), qualitative assessment by Agilent MassHunter Qualitative Analysis (version B.05.00 or B.06.00), and data curation by Agilent Profinder (version B.08.00)

5.5.6. Bioinformatics

For the phylogenetic reconstructions, the best amino acid substitution model was selected using ModelFinder as implemented on IQ-tree [115] 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 OplBA were identified with CORASON-BGC [115] and manually colored and annotated. DNA-binding sites were predicted with MEME [115].

5.5.7. Analysis of biosensor performance

Fluorescent measurements were fit to the Hill equation and biosensor parameters we estimated as described previously [115].

Induction above background was calculated by dividing the maximal experimental normalized RFP expression by background fluorescence in uninduced cultures. The experimental limit of detection was defined as the minimal concentration of inducer that produced normalized fluorescence that was statistically above uninduced cultures harboring the same plasmid via Student’s t-test ($p<0.05$).

5.6. Miscellaneous

5.6.1. Acknowledgements

We would like to thank the Koret Research Scholars Program for providing funding to ANP 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, 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. 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). HGM 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.

5.6.2. Collaborator contributions

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

5.6.3. Competing interests

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

Chapter 6. A Dance with DUFs: DUF1338 family enzymes are Fe(II) dependent oxygenases with highly conserved biochemical and physiological functions across multiple domains of life

6.1. Abstract

Due to its low abundance in cereal grains lysine must be produced on an industrial scale to meet agricultural needs. Despite intensive study and manipulation, the catabolism of lysine in plants past the step producing 2-oxoadipate (2OA) has yet to be elucidated. Recently a missing step in D-lysine catabolism was described in *Pseudomonas putida* in which 2OA is converted to 2-hydroxyglutarate (2HG) via an enzyme (HglS) belonging to the previously uncharacterized DUF1338 family of proteins. Here we solve the structure of HglS to 1.1Å resolution in the apo form and in complex with 2OA. Structural similarity to hydroxymandelate synthase suggested a mechanism by which 2HG is formed through successive decarboxylation and intramolecular hydroxylation in a Fe-(II)- and O_2-dependent manner, which we experimentally validate. Specificity for 2OA was in part mediated by a single arginine (R74), which was highly conserved across nearly all DUF1338 family proteins. In plants DUF1338 homologs localized to the apoplast were significantly coexpressed with the known enzymes of lysine catabolism, and a mutant of the *Arabidopsis thaliana* homolog showed significant defects in germination rates. Structural and biochemical analysis of the *Oryza sativa* homolog FLO7 revealed identical enzymology to the bacterial homolog despite low sequence level identity. Our results suggest that all DUF1338 containing enzymes may catalyze the same biochemical and physiological function.

6.2. Introduction

Lysine in an essential amino acid, and due to its low abundance in cereals and legumes is produced on a scale of one million tons a year by bacterial fermentation to supplement agricultural needs [116–118]. Significant work has been done to engineer crops such as rice and maize to produce greater quantities of lysine as a means to thwart malnutrition in the developing world [118,119]. High-level fortification of cereal grains with lysine requires the overexpression of the biosynthetic pathway in addition to disruption of lysine catabolism [118]. Mutants such as *opaque2* in maize have thus received considerable attention for their ability to accumulate lysine within their endosperm [120,121]. Despite its worldwide importance the catabolic pathways of plant lysine catabolism have yet to be resolved, with no consensus for catabolism past the formation of 2OA [122].

Recently we described a novel route of D-lysine catabolism in the bacterium *Pseudomonas putida* [75]. In this pathway from 2-aminoadipate (2AA) to 2-oxoglutarate (2OG) a unique 2OA to D2HG reaction was found to be catalyzed by the DUF1338 family enzyme HglS (Figure 6-1) [75]. Homologs of this enzyme are broadly distributed across multiple domains of life, and are found in every sequenced plant genome [29,75]. The only other published work to describe DUF1338 enzymes observed aberrant starch granule formation in *O. sativa flo7* mutants resulting in a floury phenotype of rice kernels similar to that of *opaque2* mutants in maize [29].

Conversion of 2OA directly to D2HG requires two chemical steps, both a decarboxylation and hydroxylation, making the chemical mechanism of the enzyme somewhat puzzling [75]. We have previously indicated that hydroxymandelate synthase (HMS) performs a similar reaction involving a decarboxylation followed by an intramolecular hydroxylation in a Fe(II)- and O_2-dependent manner [27,123]. Here we leverage structural and biochemical analyses to postulate a logical chemical mechanism and characterize the specificity of HglS, and we show that critical residues involved in catalysis have been highly conserved across nearly all DUF1338 family proteins. We further show that despite very low sequence identity that plant homologs also catalyze a 2OA to 2HG decarboxylation and are likely the last unknown step of plant lysine catabolism.

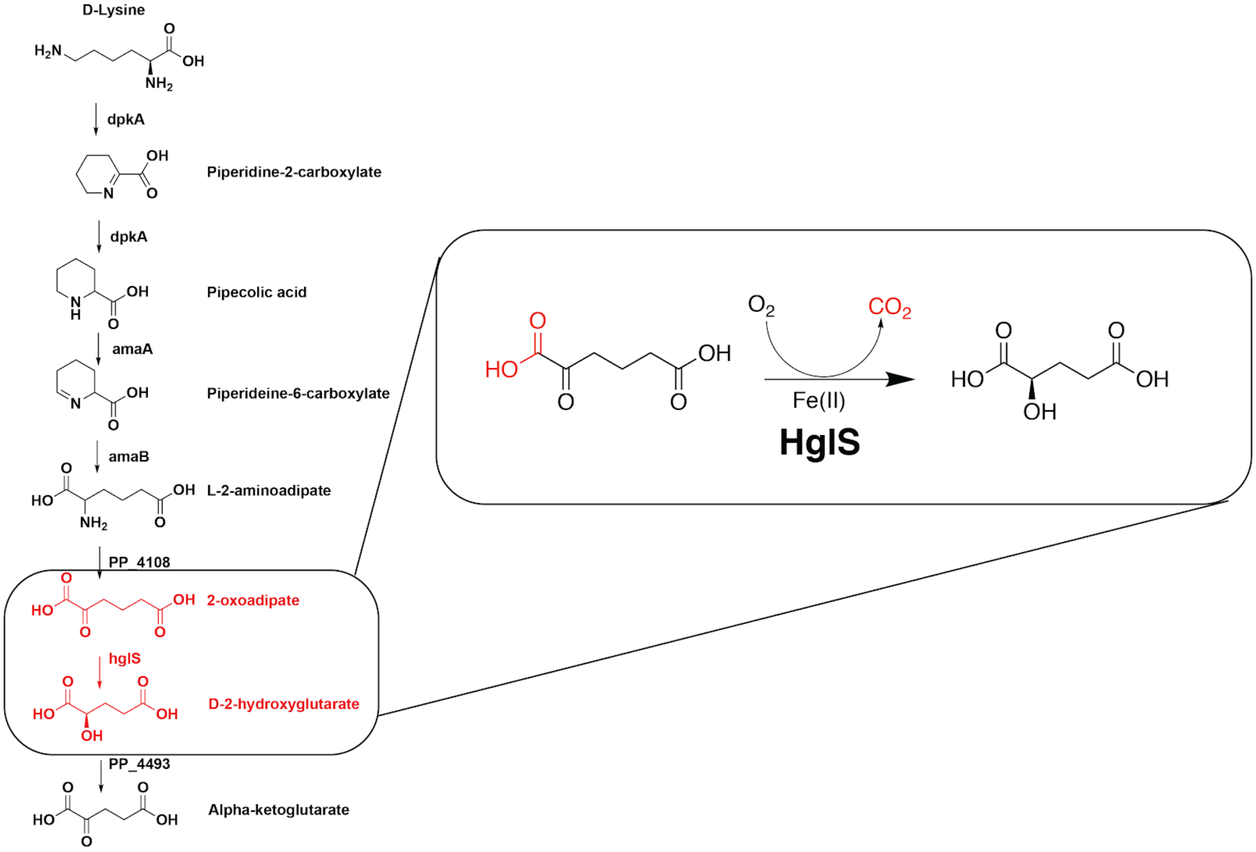


Figure 6-1 D-lysine catabolic pathway of *P. putida*. - D-lysine is catabolized independently of L-lysine in *P. putida*. Previously we identified HglS, which catalyzed a previously undescribed 2OA to D-2HG decarboxylation. D-2HG is then brought into the TCA cycle via oxidation to 2KG.

6.3. Results

6.3.1. Structural and biochemical analyses of HglS

In order to better understand the chemical mechanism of HglS, we crystallized the protein in its *apo* form. When diffracted, the crystals produced high resolution data which was used to solve the structure to 1.1Å resolution. The enzyme possesses a central jelly roll fold that is conserved in the three other DUF1338 structures deposited in the Protein Data Bank (PDB, Figure 6-2A) [124–126]. Further analysis also revealed the presence of a conserved metal cofactor-binding motif consisting of 3 residues: His 70, His226, and Glu 294. While the central domain and metal-binding motif of the proteins is conserved, the surrounding α-helices of each protein diverge significantly, suggesting that only the central region participates in the catalytic function of the enzymes. However, none of the DUF1338 structures deposited in the PDB have been biochemically characterized, making the proposal of a chemical mechanism based on structural comparison of HglS with these homologs challenging.

Instead, we performed a search for characterized proteins with similar structures using the Vector Alignment Search Tool (VAST) [127]. The top hits from the VAST search were the three solved DUF1338 structures, followed by the hydroxymandelate synthase (HMS) and 4-hydroxyphenyl pyruvate dioxygenase (HPPD) structures (Table 6-1) [123,128]. We had previously noted the similarity between the set of reactions catalyzed by HglS and HMS. Both enzymes first decarboxylate an α-ketoacid substrate then subsequently hydroxylate the resulting acid at the α-position. HPPD catalyzes a similar overall reaction with the hydroxylation occurring at a different position. Comparison of the structure of HglS with HMS revealed that the two enzymes share the central jelly roll fold common to the DUF1338 structures (Figure 6-3). More importantly, the jelly roll domain of HMS contains the active site of the enzyme which consists of three metal-binding residues (two histidines and a glutamate) which bind the Fe(II) cofactor in nearly the same orientation as in HglS. However, the identity of the metal cofactor in the HglS structure was determined to be nickel through a fluorescent scan of the protein crystal which resulted in the emission of a photon near 7475 eV, the K-alpha emission of nickel (Figure 6-4). While the metal cofactors present in the structures of HMS and HglS differ, we had previously determined that HglS utilizes Fe(II), not Ni(II), for catalysis. It is likely that the nickel bound to HglS is simply an artefact of protein purification via nickel affinity chromatography.

The structural and biochemical similarity of HglS and HMS led us to propose a chemical mechanism for HglS that is similar to the HMS mechanism (Figure 1B). More specifically, the catalytic cycle begins when the oxygens of the carboxylate and α-keto groups of the substrate and molecular oxygen bind to the 3 open coordination sites of Fe(II). A radical rearrangement results in a decarboxylation and the formation of a Fe(IV)-oxo and an α-radical species. Finally, continued radical rearrangement results in the formation of the hydroxylated product 2-hydroxyglutarate. If HglS follows this mechanism, it should consume O_2 and 2-oxoadipate in a 1:1 ratio, and two oxygens in the product should be derived from molecular O_2. In order to gather evidence to support this mechanism, we determined the stoichiometry of substrate to O_2 consumption. The enzyme consumed the same concentration of O_2 as the concentration of substrate provided (Figure 6-2C). In order to determine the source of oxygen atoms in the produced 2-hydroxyglutarate, a set of enzyme assays were performed under an atmosphere of O^{18}_2. Analysis of the assays using LC-HRMS revealed the presence of a species that co-eluted with a 2-hydroxyglutarate standard and had a m/z of 151.022, corresponding to 2-hydroxyglutarate containing two O^{18}-labeled oxygens (Figure 6-2D). No feedback inhibition was observed in the presence of 1 mM D-2HG, or L-2HG (Figure

6-5A), and kinetic parameters as determined by oxygen consumption were similar to those previously reporter (Figure 6-5B)

To further corroborate our proposed mechanism, we attempted to obtain structural data to confirm the mode of substrate binding. Soaking HglS crystals with 2-oxoadipate produced crystals that, when diffracted, allowed us to solve another structure of HglS which contained a new density occupying two of the coordination sites of the metal. Structural refinement confirmed that the unknown density corresponded to 2-oxoadipate, the substrate. The presence of nickel in the active site fortuitously meant that the enzyme was unable to turn over without its Fe(II) cofactor, resulting in a substrate-bound structure and no observation of product density. As hypothesized, the oxygens of the carboxylate and α-keto groups of the substrate are coordinated to the metal. The co-crystal structure also revealed several residues that participate in substrate binding. Specifically, arginine 74 forms a salt bridge with the distal end of the substrate (Figure 6-2E). Additionally, residues valine 402 and serine 403 form hydrogen bonds with the α-keto group and distal carboxylate group, respectively (Figure 6-2F). Residues 402 and 403 are located on a loop that displays significant movement between the apo and substrate-bound states. Arginine 74 appears to not only participate in favorable substrate binding interactions, but it also appears to influence the strict substrate specificity of the enzyme. When assayed against a panel of α-ketoacids, HglS was found to be only active on 2OA, with no other 2-ketoacid able to be turned over by the enzyme when measured via an enzyme coupled assay (Figure 6-2G). Mutational analysis revealed that, while valine 402 and arginine 74 are not completely essential for catalysis, they contribute significantly to enzyme activity. When each residue was mutated to alanine, enzyme turnover was significantly slower (Figure 6-2H).

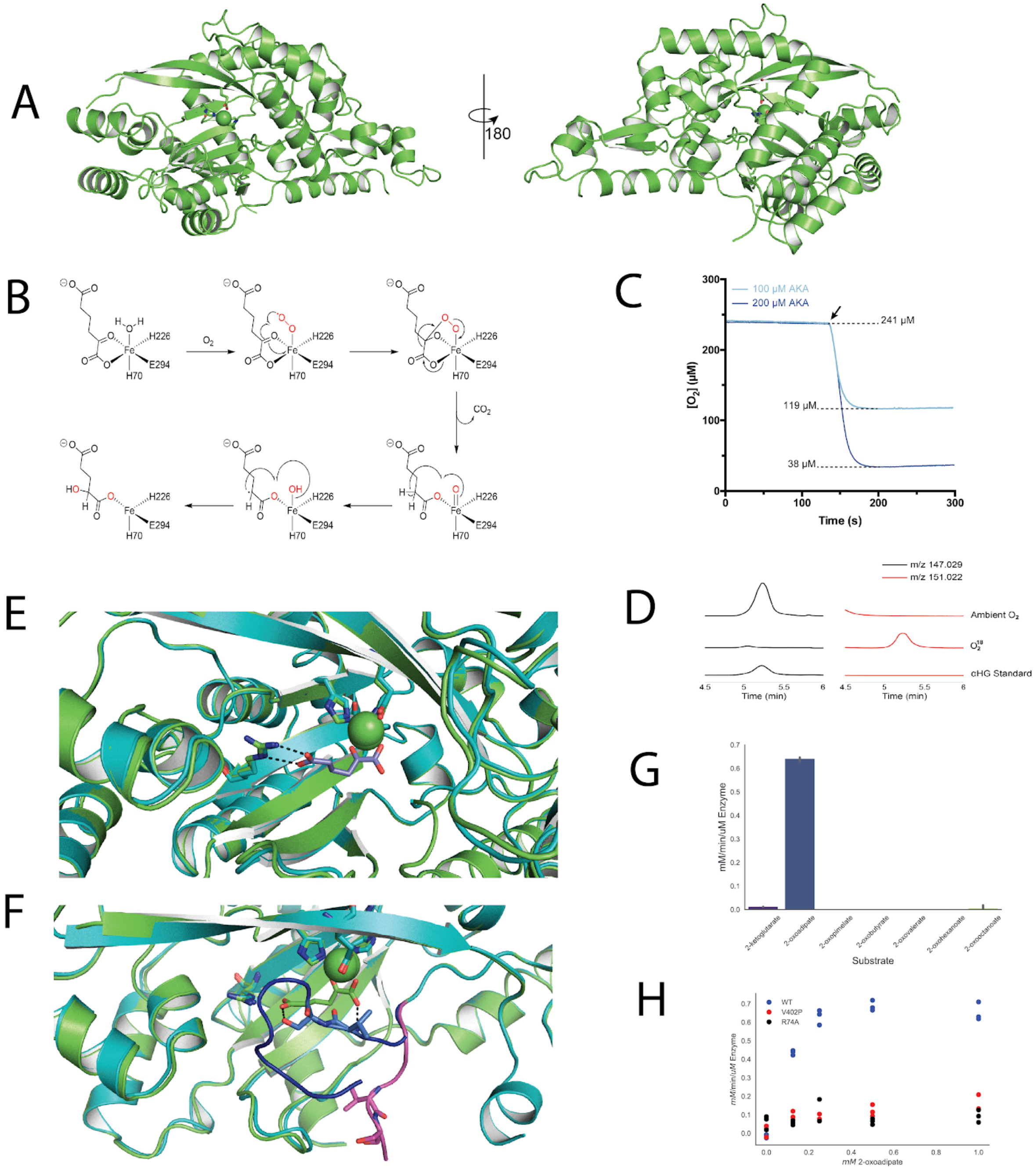


Figure 6-2 Structural and mechanistic analysis of HglS. - Overall structure of PP_5260 displayed in cartoon form. On the left, the entrance to the active site which contains the metal cofactor (in spheres) and the central jelly roll fold of the protein. A 180° rotation on the right shows the overall structure of the protein. **B)** Proposed chemical mechanism of PP_5260. Oxygens derived from O_2 are shown in red. **C)** Oxygen stoichiometry of HglS is 1:1 with 2OA. The concentration of dissolved oxygen (y-axis) was measured with a Clark-type oxygen probe. Either 200 μM or 100 μM of 2OA was added to initiate reaction with equimolar oxygen being consumed. **D)** LC-HRMS extracted ion chromatograms (EICs)

showing the incorporation of labeled O^{18} into the 2-hydroxyglutarate product of PP5260. On the left (black lines) are EICs for ions with m/z 147.029, representing 2-hydroxyglutarate containing O^{16}. On the right (red lines) are EICs for ions with m/z 151.022, representing 2-hydroxyglutarate containing two O^{18} atoms. A control reaction performed under ambient O_2 conditions (n=3) is compared to a reaction performed in a O^{18}_2 atmosphere (n=3) and a 2-hydroxyglutarate standard. **E)** Overlay of apo (teal) and substrate-bound (green) structures of PP5260 displaying the enzyme active site and the ionic interaction between Arg74 and the distal carboxylate group of the 2-oxoadipate substrate. **F)** Overlay of apo (teal) and substrate-bound (green) structures of PP5260 displaying the enzyme active site and the interaction of Val402 and Ser403 with the 2-oxoadipate substrate. The loop containing Val402 and Ser403 is shown in the apo (blue) and substrate-bound (magenta) states. **G)** Reaction rates of HglS using different 2-oxoacids as substrates. All substrates were supplied at 1 mM, and rates were measured via an enzyme coupled decarboxylation assay. Error bars represent the 95% confidence interval, n=3. **H)** Reaction rates of WT HglS (blue dots) as well as R74A (black dots) and V402P (red dots) mutants as measured by enzyme coupled decarboxylation assay with 2OA as a substrate, n=3.

Table 6-1 Results of VAST analysis of HglS.

PDB	Ali. Len.	SCORE	P-VAL	RMSD	%Id	Description
2RJB A	408	43.6	10e-43.8	1.3	59.1	Crystal Structure Of Uncharacterized Protein Ydcj (Sf1787) From Shigella Flexneri Which Includes Domain Duf1338. Northeast Structural Genomics Consortium Target Sfr276
2RJB A 1	306	28.4	10e-25.6	0.9	62.4	Crystal Structure Of Uncharacterized Protein Ydcj (Sf1787) From Shigella Flexneri Which Includes Domain Duf1338. Northeast Structural Genomics Consortium Target Sfr276
3IUZ A	256	22.1	10e-12.8	2.9	20.3	Crystal Structure Of Putative Glyoxalase Superfamily Protein (Yp_299723.1) From Ralstonia Eutropha Jmp134 At 1.90 A Resolutionÿ

3LHO A	218	21.4	10e-14.3	2.7	21.1	Crystal Structure Of Putative Hydrolase (Yp_751971.1) From Shewanella Frigidimarina Ncimb 400 At 1.80 A Resolutionÿ
2R5V A 2	133	13.7	10e-5.3	2.3	12.8	Hydroxymandelate Synthase Crystal Structure
1CJX B	132	14.7	0.0095	2.4	18.2	Crystal Structure Of Pseudomonas Fluorescens Hppd
5EC3 A 2	127	12.4	0.0047	2.5	15	Structural Insight Into The Catalyitc Mechanism Of Human 4-Hydroxyphenylpyruvate Dioxygenase
5EC3 A 1	125	10.8	0.0028	2.9	12	Structural Insight Into The Catalyitc Mechanism Of Human 4-Hydroxyphenylpyruvate Dioxygenase
1CJX B 2	125	13.8	10e-5.5	2.3	19.2	Crystal Structure Of Pseudomonas Fluorescens Hppd
1JC5 B	124	10.5	0.0089	3	12.9	Crystal Structure Of Native Methylmalonyl-Coa Epimerase
4HUZ A 1	119	12	0.0003	2.9	16.8	2,6-dichloro-p-hydroquinone 1,2-dioxygenase
1SP8 D 1	118	12.8	10e-5.2	2.9	11.9	4-Hydroxyphenylpyruvate Dioxygenase
1SP8 D 2	114	11.5	0.0079	2.9	17.5	4-Hydroxyphenylpyruvate Dioxygenase
2R5V A 1	114	11.6	0.0002	2.7	13.2	Hydroxymandelate Synthase Crystal Structure
3OA4 A	111	9.3	0.0199	3	11.7	Crystal Structure Of Hypothetical Protein

						Bh1468 From Bacillus Halodurans C-125ÿ
2ZI8 A	107	13.2	0.0209	2.9	15	Crystal Structure Of The Hsac Extradiol Dioxygenase From M. Tuberculosis In Complex With 3,4-Dihydroxy-9,10-Seconandrost-1,3,5(10)-Triene-9,17-Dione (Dhsa)ÿ
1ZSW A 1	104	12.2	10e-5.2	2.3	13.5	Crystal Structure Of Bacillus Cereus Metallo Protein From Glyoxalase Family
3LM4 D 2	104	11.1	0.0034	3	8.7	Crystal Structure Of 2,3-Dihydroxy Biphenyl Dioxygenase From Rhodococcus Sp. (Strain Rha1)ÿ
3VB0 A 2	101	11.7	10e-4.2	2.5	13.9	Crystal Structure Of 2,2',3-trihydroxybiphenyl 1,2-dioxygenase From Dibenzofuran-degrading Sphingomonas Wittichii Strain Rw1
3B59 A 1	100	11.9	10e-5.3	2.7	18	Crystal Structure Of The Mn(Ii)-Bound Glyoxalase From Novosphingobium Aromaticivorans
2ZI8 A 2	99	12.8	10e-6.0	2.7	15.2	Crystal Structure Of The Hsac Extradiol Dioxygenase From M. Tuberculosis In Complex With 3,4-Dihydroxy-9,10-Seconandrost-1,3,5(10)-Triene-9,17-Dione (Dhsa)ÿ
1CJX B 1	99	10.7	0.0031	2.9	15.2	Crystal Structure Of Pseudomonas Fluorescens Hppd
1MPY A 1	98	11.4	0.0003	2.6	15.3	Structure Of Catechol 2,3-Dioxygenase

						(Metapyrocatechase) From Pseudomonas Putida Mt-2
2EHZ A 2	97	10.6	0.024	2.7	12.4	Anaerobic Crystal Structure Analysis Of 1,2-Dihydroxynaphthalene Dioxygenase From Pseudomonas Sp. Strain C18 Complexed With 4-Methylcatechol
3LM4 D 1	97	12.2	10e-6.6	2.6	8.2	Crystal Structure Of 2,3-Dihydroxy Biphenyl Dioxygenase From Rhodococcus Sp. (Strain Rha1)ÿ
2EHZ A 1	96	9.9	0.0229	2.5	11.5	Anaerobic Crystal Structure Analysis Of 1,2-Dihydroxynaphthalene Dioxygenase From Pseudomonas Sp. Strain C18 Complexed With 4-Methylcatechol
1EIL A 2	96	11.6	0.0007	2.4	13.5	2,3-Dihydroxybiphenyl-1,2-Dioxygenase
3B59 A 2	95	10.8	0.0013	2.4	16.8	Crystal Structure Of The Mn(Ii)-Bound Glyoxalase From Novosphingobium Aromaticivorans
2RJB A 3	94	8.8	10e-4.8	0.8	52.1	Crystal Structure Of Uncharacterized Protein Ydcj (Sf1787) From Shigella Flexneri Which Includes Domain Duf1338. Northeast Structural Genomics Consortium Target Sfr276
2ZI8 A 1	93	10.9	0.0007	2.5	12.9	Crystal Structure Of The Hsac Extradiol Dioxygenase From M. Tuberculosis In Complex With 3,4-Dihydroxy-9,10-Seconandrost-1,3,5(10)-Triene-9,17-Dione (Dhsa)ÿ

1EIL A 1	93	11.4	0.0001	2.2	10.8	2,3-Dihydroxybiphenyl-1,2-Dioxygenase
3VB0 A 1	92	11.7	0.0005	2.6	9.8	Crystal Structure Of 2,2',3-trihydroxybiphenyl 1,2-dioxygenase From Dibenzofuran-degrading Sphingomonas Wittichii Strain Rw1
1F1U A 1	91	10.8	10e-4.1	2.7	12.1	Crystal Structure Of Homoprotocatechuate 2,3-Dioxygenase From Arthrobacter Globiformis (Native, Low Temperature)
5BWH B 1	91	9.7	0.005	2.8	9.9	Structure Of H200c Variant Of Homoprotocatechuate 2,3-dioxygenase From B.fuscum In Complex With Hpca At 1.46 Ang Resolution
2RK9 B	89	9.5	0.0086	3.1	6.7	The Crystal Structure Of A GlyoxalaseBLEOMYCIN RESISTANCE ProteinDIOXYGENASE SUPERFAMILY MEMBER FROM VIBRIO Splendidus 12b01
3E0R A	84	12	0.0087	2.2	7.1	Crystal Structure Of Cppa Protein From Streptococcus Pneumoniae Tigr4
5BWH B 2	84	9.8	0.0327	2.4	13.1	Structure Of H200c Variant Of Homoprotocatechuate 2,3-dioxygenase From B.fuscum In Complex With Hpca At 1.46 Ang Resolution
3E0R A 1	82	12	10e-7.0	2.3	7.3	Crystal Structure Of Cppa Protein From Streptococcus Pneumoniae Tigr4
1MPY A 2	82	10.2	0.0491	2.2	12.2	Structure Of Catechol 2,3-Dioxygenase

						(Metapyrocatechase) From Pseudomonas Putida Mt-2
4RT5 A	78	9.1	0.0038	2.6	10.3	The Crystal Structure Of A Glyoxalase/bleomycin Resistance Protein/dioxygenase Protein From Planctomyces Limnophilus Dsm 3776
3FCD B	76	9.6	0.0072	2.6	7.9	Crystal Structure Of A Putative Glyoxalase From An Environmental Bacteria
3P8A A 2	73	9.3	0.0003	2.5	6.8	Crystal Structure Of A Hypothetical Protein From Staphylococcus Aureusÿ

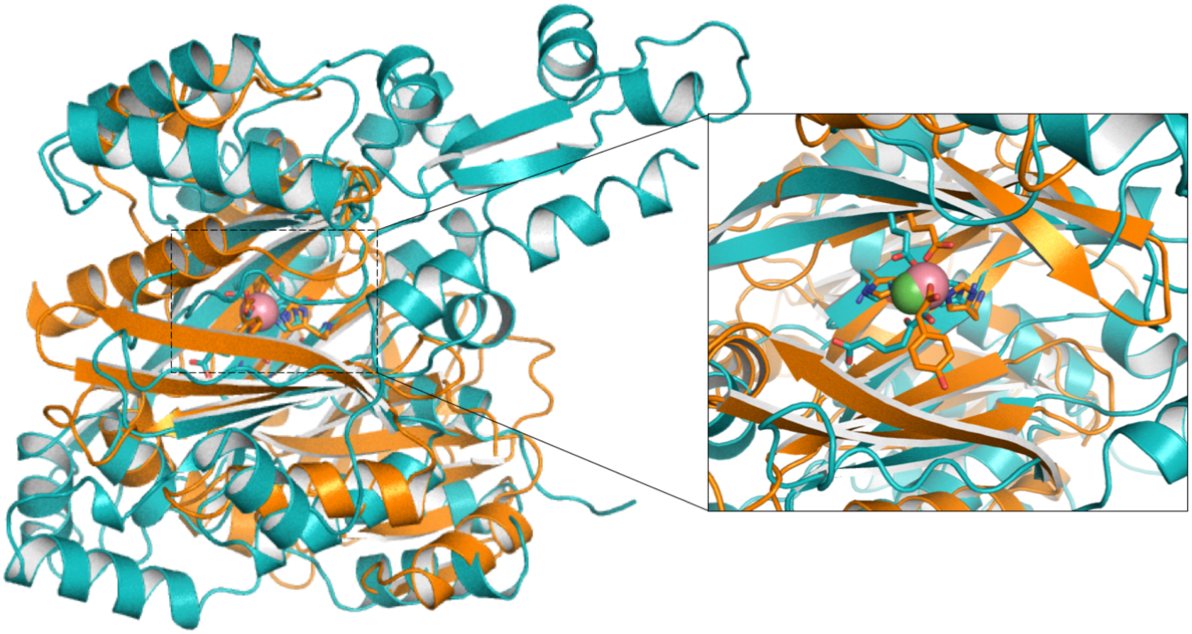


Figure 6-3 HglS/HMS overlay. Overlay of the active sites of PP_5260 (teal) and HMS (orange) showing the conservation of the central jelly roll fold and active site architecture. Metal-binding residues and substrates are shown as sticks. Metal cofactors are shown as spheres.

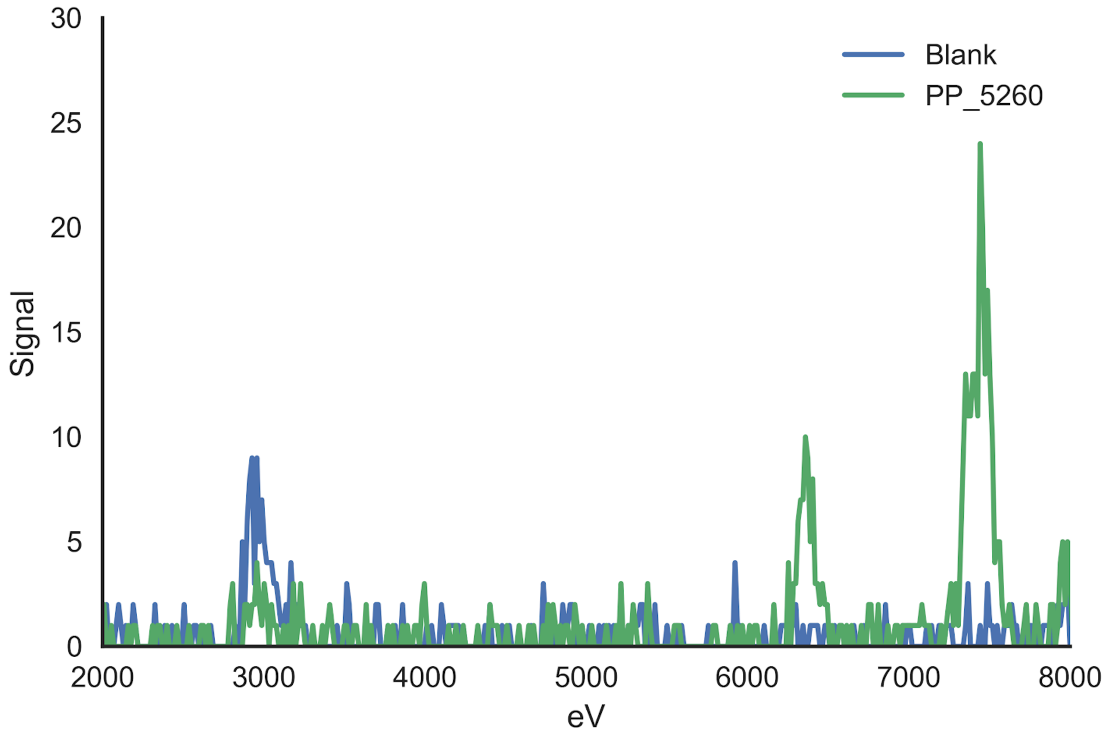


Figure 6-4 K-alpha emission spectra of HglS crystals. K-alpha emission analysis of purified HglS showed emissions characteristic of Ni (near 7475 eV), as well as Fe (near 6400 eV).

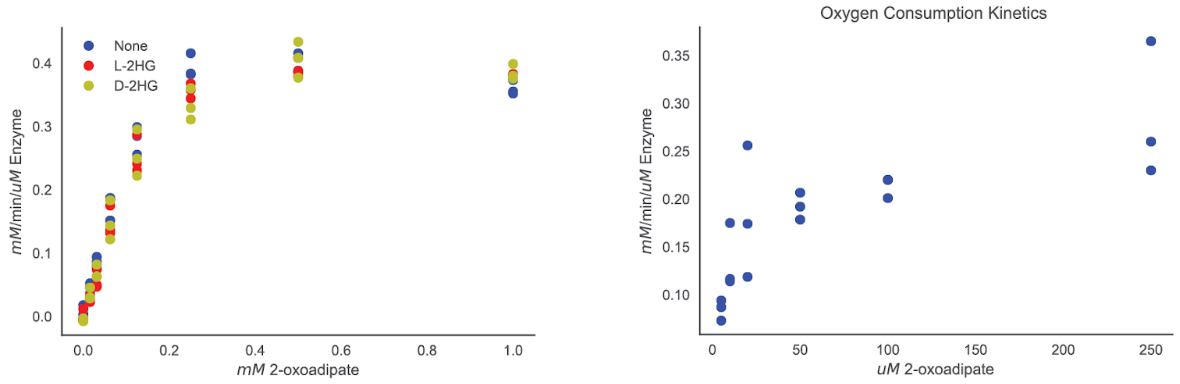


Figure 6-5 Product feedback inhibition and oxygen consumption kinetics of HglS. Left: Reaction rates of HglS in the presence of either L-2HG or D-2HG at 1 mM concentration as measured by enzyme coupled decarboxylation assay with 2OA as a substrate, n=3. Right: Reaction rates of HglS when measured by oxygen consumption, n=3. The Km was determined to be 0.01 mM with a Vmax of 0.273 mM/min/ μM enzyme.

6.3.2. Plant HglS homologs encode the final missing step of plant lysine catabolism.

Though there is little primary sequence identity between bacterial homologs and plant homologs (Figure 6-6), the residues critical that coordinate the Fe (II) as well as the carboxylate coordinating arginine are both highly conserved (Figure 6-7A). Across all DUF1338 homologs >99% of those that had the canonical HHE metal binding triad also maintained the arginine that mediates specificity (Figure 6-7B). Given the substrate specificity displayed by the *P. putida* homolog we hypothesized that all homologs that maintained the conserved arginine may act on 2OA as a substrate.

When we examined other metabolic pathways that go through 2OA we found that plant lysine catabolism is also known to produce 2OA, though the remaining steps of plant lysine catabolism are not known. Previous work suggested a route through the 2-ketoglutarate dehydrogenase complex (2KGD) resulting in the formation of glutaryl-coA (Figure 6-8) [122]. Correlation analysis of transcripts in *Arabidopsis thaliana*, however, revealed that transcription of known lysine catabolic enzymes was not correlated with the expression 2KGD, but highly correlated with the expression of the DUF1338-containing protein AT1G07040 (Table 6-2). Previous work with another plant homolog in *Oryzae sativa*, FLO7, revealed the protein to be localized to the apoplasm with mutants developing a floury starch phenotype similar to that of *opaque2* mutants in maize [29]. There has been much agricultural interest in *opaque2* mutants as they are known to accumulate high levels of lysine in their kernels [118,120,121]. In *A. thaliana* AT1G07040 mutants displayed significantly delayed germination, consistent with perturbations to apoplasm during seed development (Figure 6-9).

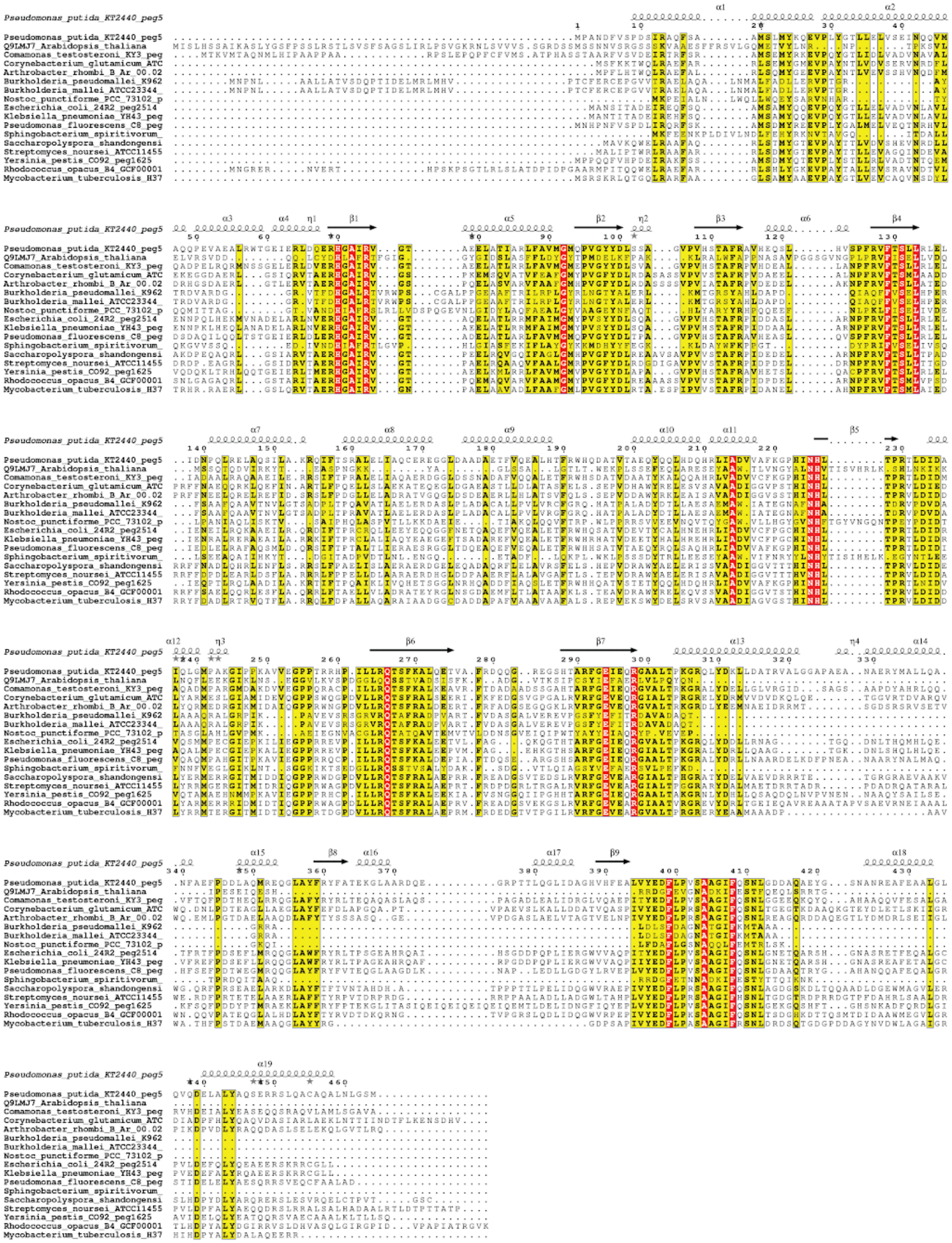

Pseudomonas_putida_KT2440_peg5
Pseudomonas_putida_KT2440_peg5
Q9LMJ7_Arabidopsis_thaliana
Comamonas_testosteroni_KY3_peg
Corynebacterium_glutamicum_ATC
Arthrobacter_rhombi_B_Ar_00.02
Burkholderia_pseudomallei_K962
Burkholderia_mallei_ATCC23344_
Nostoc_punctiforme_PCC_73102_p
Escherichia_coli_24R2_peg2514
Klebsiella_pneumoniae_YH43_peg
Pseudomonas_fluorescens_C8_peg
Sphingobacterium_spiritivorum_
Saccharopolyspora_shandongensi
Streptomyces_noursei_ATCC11455
Yersinia_pestis_CO92_peg1625
Rhodococcus_opacus_B4_GCF00001
Mycobacterium_tuberculosis_H37

Figure 6-6 Alignment of bacterial and plant homologs of *P. putida* HglS. Conserved residues are shown in yellow, while perfectly conserved residues are shown in red. Above alignments is the secondary structure of HglS.

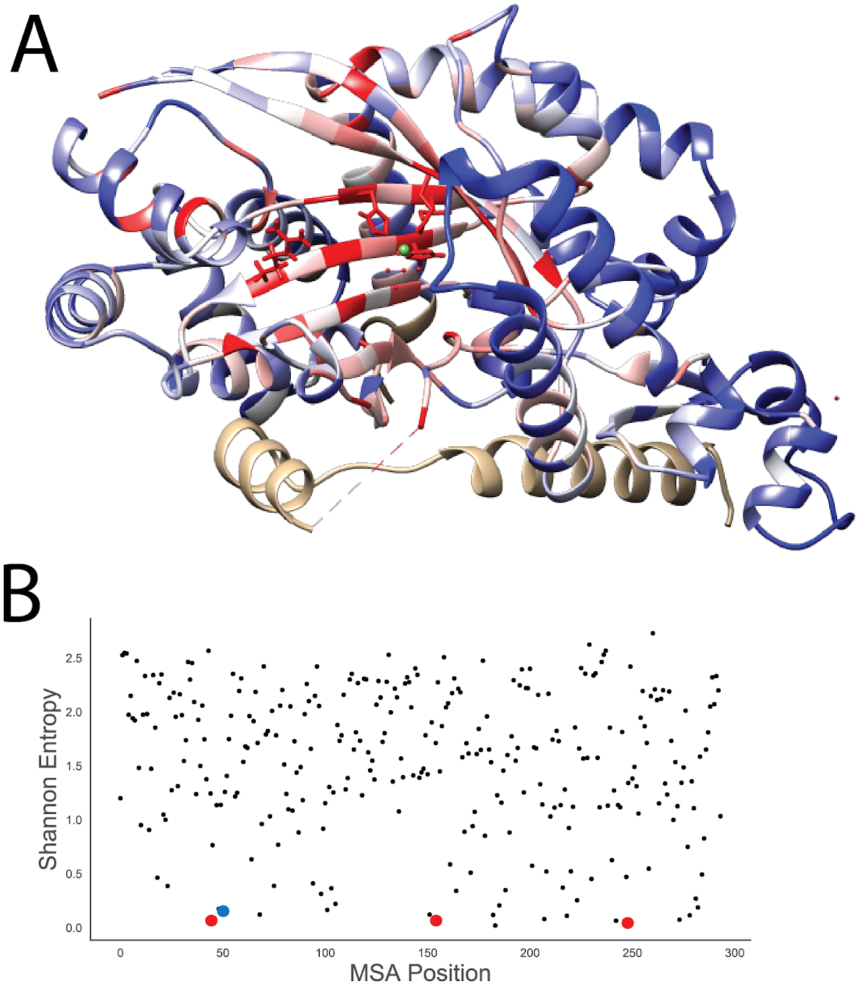


Figure 6-7 Conservation of critical residues in DUF1338 proteins. A) Structure of HglS with residues colored by conservation, red showing the highest level of conservation with blue showing the least amount of conservation. B) Shannon entropy of an alignment of the representative DUF1338 containing proteins in the Pfam database. The Fe (II) coordinating HHE triad residues are highlighted as red dots, while the carboxylate coordinating arginine is highlighted as a blue dot.

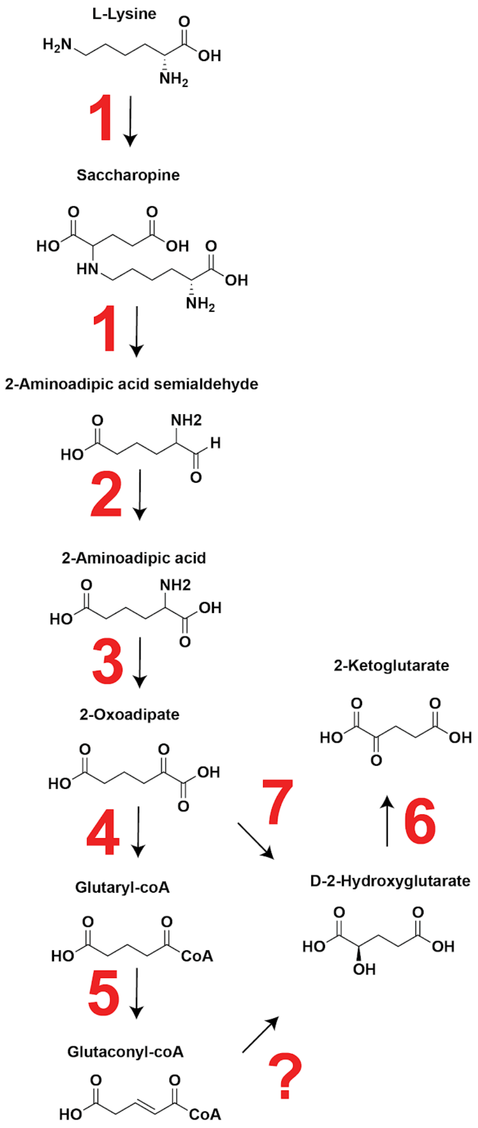
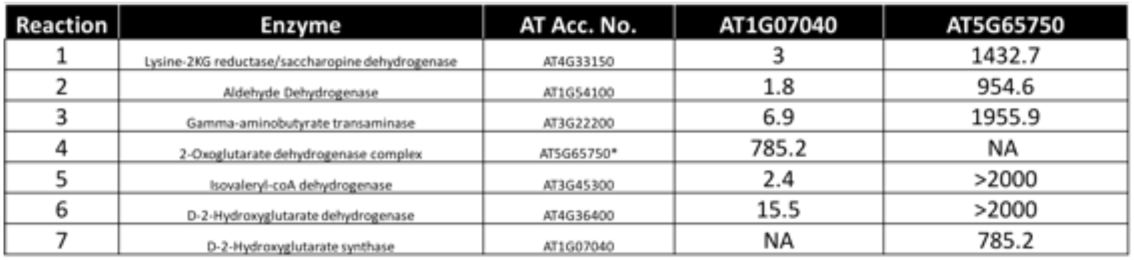


Figure 6-8 Putative route of lysine catabolism in plants. Enzymes putatively involved in plant lysine catabolism according to [115] 1) Lysine-2KG reductase/saccharopine dehydrogenase 2) Aldehyde dehydrogenase 3) Gamma-aminobutyrate aminotransferase 4) 2KGD 5) Isovaleryl-coA dehydrogenase 6) D-2-hydroxyglutarate dehydrogenase. ? represents a putative enzyme that would convert glutaconyl-coA to D2HG. We propose that HglS homologs (7) catalyze the conversion of 2OA to 2HG in plants, rather than go through the 2KGD complex.

Table 6-2 Coexpression analysis of *A. thaliana* lysine catabolic enzymes. Expression of enzymes that catalyze reactions from Figure SX were analyzed for their coexpression, reported by their Mutual Rank Index, with either HglS or 2KGD in *A. thaliana*. HglS showed to be highly correlated with all known lysine catabolic enzymes (column AT1G07040), while there was little correlation with the 2KGD complex (column AT5G65750).



Reaction	Enzyme	AT Acc. No.	AT1G07040	AT5G65750
1	Lysine-2KG reductase/saccharopine dehydrogenase	AT4G33150	3	1432.7
2	Aldehyde Dehydrogenase	AT1G54100	1.8	954.6
3	Gamma-aminobutyrate transaminase	AT3G22200	6.9	1955.9
4	2-Oxoglutarate dehydrogenase complex	AT5G65750*	785.2	NA
5	Isovaleryl-coA dehydrogenase	AT3G45300	2.4	>2000
6	D-2-Hydroxyglutarate dehydrogenase	AT4G36400	15.5	>2000
7	D-2-Hydroxyglutarate synthase	AT1G07040	NA	785.2

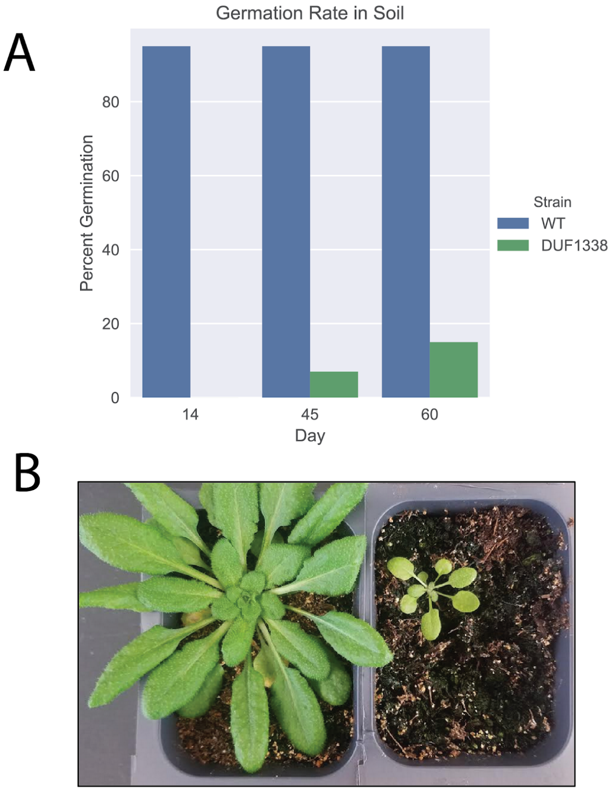


Figure 6-9 Germination rates of WT and DUF1338 mutants of *A. thaliana*. A) Mutants of AT1G07040 showed delayed germination relative to WT *A. thaliana*. No germination in mutants took place after 14 days after planting while 95% of WT seeds had germinated, n=96. By 60 days after planting only 15% of mutant seeds had germinated. B) Gross physiology of WT (left) and mutant AT1G07040 plants after 45 days after planting.

To evaluate the biochemical activity of plant homologs we purified both *A. thaliana* and *O. sativa* homologs without their N-terminal localization sequences. Both homologs were catalyzed the conversion of 2OA to 2HG, with the FLO7 homolog having demonstrating a Km of 0.55 mM and a Vmax of 0.89 mM/min/ μM enzyme (Figure 6-10). The high expression level of FLO7 resulted in enough protein production to set up crystallographic screens of the enzyme in order to investigate the conservation of DUF1338 domain-containing proteins in higher eukaryotes. Initial screening resulted in crystals which, after optimization of crystallization conditions and soaking of 2-oxoadipate, produced diffraction data which was used to solve the substrate-bound crystal structure at 1.85Å resolution. The FLO7 crystal structure, like the other structures of DUF1338 domain-containing proteins, displayed the conserved central jelly roll fold containing the active site and metal-binding center (Figure 6-11A). Comparison of the PP5260 and FLO7 structures revealed that the orientation of the metal-coordinating residues and 2-oxoadipate in both structures was nearly identical, even though the proteins display low (~15%) sequence homology (Figure 6-11B). In addition, the FLO7 structure also contains a substrate-binding arginine, Arg64, which is located in nearly the same position as Arg74 of PP5260. However, unlike PP5260, FLO7 does not possess any other residues, such as Val402 and Ser403 of PP5260, that interact with the substrate. The structure of FLO7 revealed the remarkable conservation of the central fold and catalytic domain of DUF1338 domain-containing proteins across widely divergent domains of life. As all sequenced green plants appear to have a homolog of this enzyme [29], it is tempting to speculate that this may be a universal step in plant lysine catabolism.

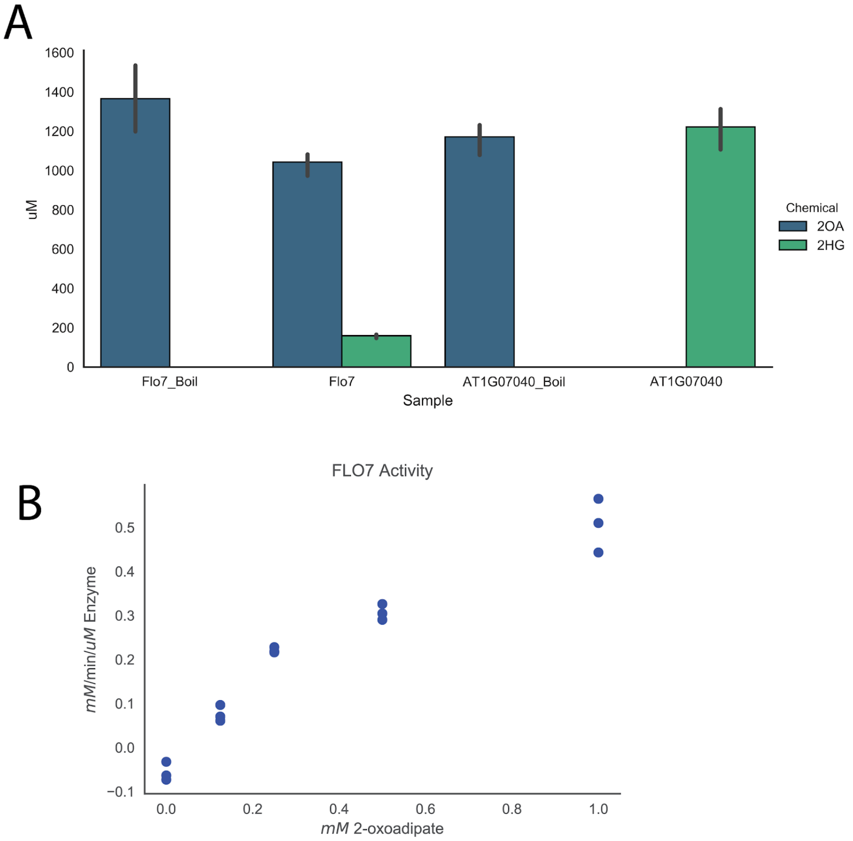


Figure 6-10 Biochemical analysis of plant HglS homologs. A) Flo7 was purified without its N-terminal localization sequence and incubated with 1 mM of 2OA for 5 min either native or boiled. AT1G07040 was also purified without its N-terminal localization sequence 1 mM of 2OA for 16 hours either native or boiled. Concentrations of 2HG and 2OA were then determined by LCMS. Error bars represent the 95% confidence intervals, n=3. B) Reaction rates of Flo7 as measured by enzyme coupled decarboxylation assay with 2OA as a substrate, n=3.

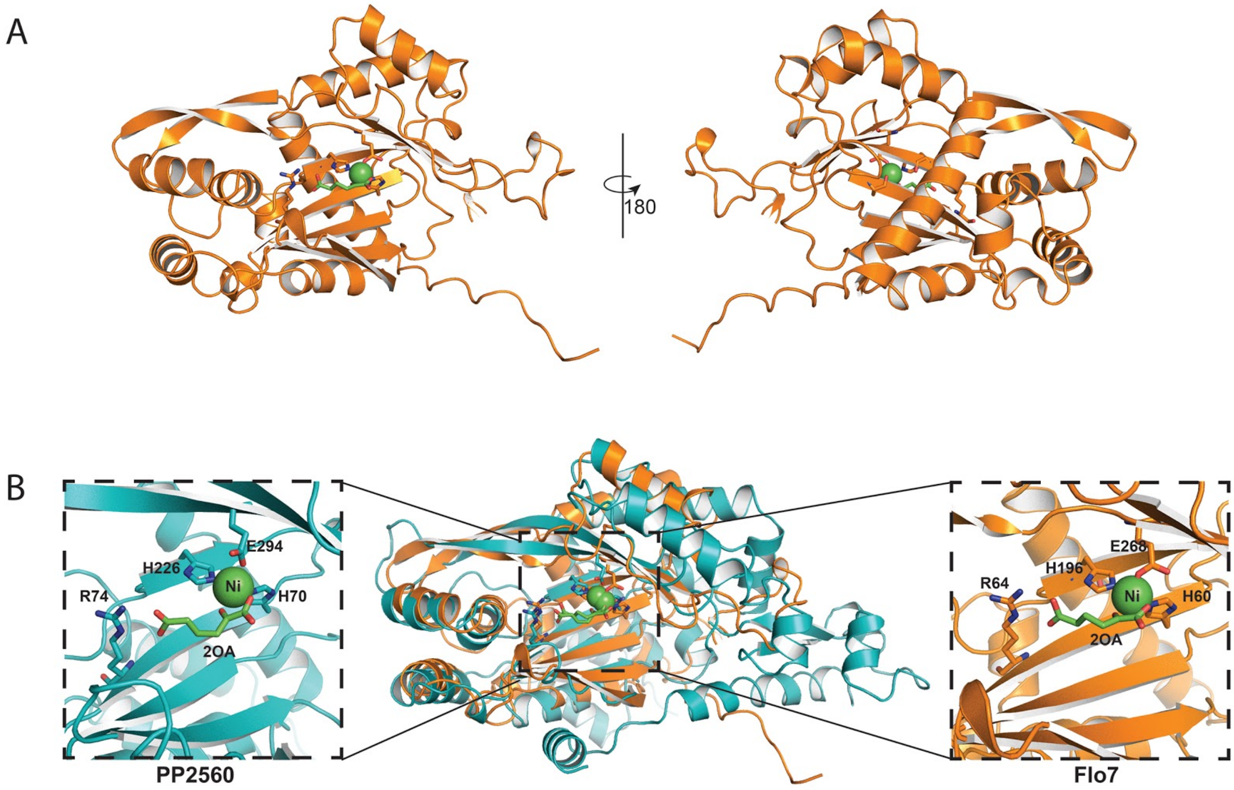


Figure 6-11 Crystal structure of *O. sativa* FLO7. Crystal structure of FLO7 displayed as cartoon sheets and helices. Metal-coordinating and substrate-binding residues are shown as orange sticks, while the 2-oxoadipate substrate is shown as green sticks. The nickel bound in the active site is shown as a green sphere. **B)** Center: overlay of the structures of PP5260 (teal) and FLO7 (orange) structures. Left and right: active sites of PP5260 (left) and FLO7 (right) displaying conserved residues and substrate-binding mode.

6.4. Discussion

Lysine is one of the most limiting essential amino acids in both cereal and legume crops, and its deficiency can lead to protein-energy malnutrition which the World Health Organization estimates up to 30% of the population in the developing world suffer from [118,129–132]. Fortification of staple crops with lysine is believed to require not only overexpression of the biosynthetic route or lysine abundant proteins, but also the elimination of the catabolic pathway [131]. However changes to the lysine metabolism in maize, rice, and soybeans has resulted in low germination rates, abnormal endosperm, and reduced grain weights [118,133–136]. With a more complete picture of the catabolic route of lysine now known, it may be possible to better understand the cause of these pleiotropic effects.

Microorganisms share ~800 of the ~1500 known DUFs found within eukaryotic organisms [137]. Within these shared unknown proteins represents a wealth of novel enzymatic potential, and druggable targets [138,139]. Recent advances in functional genomics to rapidly characterize genes in diverse microbial species has the potential to unlock the function of many of the proteins [15]. Our de-orphaning of the *P. putida* enzyme HglS using RB-TnSeq [75], and our subsequent determination of its biochemical mechanism and physiological function in plant homologs will hopefully serve as a roadmap for ongoing research to characterize the biological unknown.

6.5. Materials and methods

6.5.1. Chemicals and media

Routine bacterial cultures were grown in Luria-Bertani (LB) Miller medium (BD Biosciences, USA). *E. coli* was grown at 37 °C. Cultures were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA). All compounds were purchased through Sigma Aldrich.

6.5.2. DNA manipulation

All bacterial strains and plasmids used in this work are listed in Table 6-3. All strains and plasmids created in this work are available through the public instance of the JBEI registry. (https://public-registry.jbei.org/folders/XXX). All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software [40–42]. Plasmids were assembled via Gibson Assembly using standard protocols [43], or Golden Gate Assembly using standard protocols [44]. Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Site directed mutants were created as previously described [140]. The geneblock for the *E. coli* codon optimized *O. sativa* was purchased through Integrated DNA Technologies (IDT, Coralville, IA). Arabidopsis cDNA was used to amplify AT1G07040.

Table 6-3 Strains and plasmids used.

Strains	**JBEI Part ID**	**Reference**
E. coli DH10B		[115]
E. coli BL21(DE3)		Novagen
Plasmids		
pET21b		Novagen
pET21b *hglS*		[115]
pET21b *hglS* R74A		This work
pET21b *hglS* V402P		This work
pET21b *FLO7* Δ1-76		This work
pET21b *AT1G07040* Δ1-56		This work

6.5.3. Protein purification

A 5 mL overnight culture of *E. coli* BL21 (DE3) containing the expression plasmid was used to inoculate a 500 mL culture of LB. Cells were grown at 37 °C to an OD of 0.6 then induced with Isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The temperature was lowered to 30 °C and cells were allowed to express for 18 hours before being harvested via centrifugation. Cell pellets were stored at -80 °C until purification. For purification, cell pellets were 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 40,000xg). 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 protein was dialyzed overnight at 4 °C against 50 mM HEPES pH 7.5, 5% glycerol.

6.5.4. Crystalization

An initial crystallization screen was set up using a Phoenix robot (Art Robbins Instruments, Sunnyvale, CA) using the sparse matrix screening method [141]. Purified PP5260 was concentrated to 20 mg/mL and Flo7 was concentrated to 10 mg/mL prior to crystallization using the sitting drop method in 0.4 μL drops containing a 1:1 ratio of protein sample to crystallization solution. For PP5260, the crystallization solution consisted of 0.2 M Ammonium Fluoride and 20 % PEG 3,350, while the crystallization solution for Flo7 contained 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6 and 1.8 M Lithium sulfate monohydrate. Crystals were transferred to crystallization solution containing 20 % glycerol prior to flash freezing in liquid nitrogen.

6.5.5. X-ray data collection and model refinement

X-ray diffraction data for HglS/PP5260 was collected at the Berkeley Center for Structural Biology on beamline 5.0.2 and 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Lab. Diffraction data for Flo7 was collected at Stanford Synchrotron Radiation Lightsource on beamline 12-2. The HglS/PP5260 and Flo7 structures were determined by the molecular-replacement method with the program *PHASER* [142] using uncharacterized protein YdcJ (SF1787) from *Shigella flexneri* (PDB ID: 2RJB) and the putative hydrolase (YP_751971.1) from *Shewanella frigidimarina* (PDB ID: 3LHO) as the search models, respectively. Structure refinement was performed by *phenix.refine* program [143]. Manual rebuilding using COOT [144] and the addition of water molecules allowed for construction of the final model. The R-work and R-free values for the final models of all structures are listed in Table 6-4. Root-mean-square deviations from ideal geometries for bond lengths, angles, and dihedrals were calculated with Phenix [145]. The overall stereochemical quality of the final models was assessed using the MolProbity program [146]. Structural analyses were performed in Coot [144], PyMOL [147], and UCSF Chimera [148].

Table 6-4 X-ray collection statistics.

	5260 apo	**Flo7 substrate-bound**
Resolution range	50.37 - 1.14 (1.181 - 1.14)	38.65 - 1.85 (1.916 - 1.85)
Space group	P 21 21 21	P 21 21 21
Unit cell	42.4486 100.746 104.765 90 90 90	92.3 98.2 203.8 90 90 90
Total reflections	1121754 (103068)	1141314 (116637)
Unique reflections	164034 (16167)	156493 (15364)
Multiplicity	6.8 (6.4)	7.3 (7.6)
Completeness (%)	99.94 (99.40)	98.94 (98.68)
Mean I/sigma(I)	12.33 (2.10)	11.52 (1.39)

Wilson B-factor	13	29.19
R-merge	0.0615 (0.6688)	0.1162 (1.592)
R-meas	0.06648 (0.7282)	0.1251 (1.707)
R-pim	0.0249 (0.2833)	0.04555 (0.6084)
CC1/2	0.999 (0.666)	0.998 (0.517)
CC*	1 (0.894)	1 (0.825)
Reflections used in refinement	164033 (16166)	156467 (15364)
Reflections used for R-free	8263 (802)	7752 (777)
R-work	0.1494 (0.2425)	0.1690 (0.2802)
R-free	0.1738 (0.2760)	0.1927 (0.3062)
CC(work)	0.967 (0.857)	0.963 (0.771)
CC(free)	0.962 (0.827)	0.959 (0.756)
Number of non-hydrogen atoms	4198	11051
macromolecules	3461	9800
ligands	1	128
solvent	736	1123
Protein residues	434	1236
RMS(bonds)	0.012	0.009
RMS(angles)	1.16	1.15
Ramachandran favored (%)	98.36	98.03
Ramachandran allowed (%)	1.64	1.89
Ramachandran outliers (%)	0	0.08
Rotamer outliers (%)	1.39	0
Clashscore	4.2	3.54
Average B-factor	23.06	37.7
macromolecules	20.95	36.15
ligands	13.52	89.45
solvent	33.01	45.4
Number of TLS groups	N/A	27

6.5.6. Enzyme kinetics and oxygen consumption

Enzyme coupled decarboxylation assays were carried out as previously described [20]. Reaction mixtures contained 100 mM Tris-HCl (pH 7), 10 mM $MgCl_2$, 0.4 mM NADH, 4 mM phosphoenol pyruvate (PEP), 100U/mL pig heart malate dehydrogenase(Roche), 2U/mL microbial PEP carboxylase (Sigma), and 10 mM 2OA. Reactions were initiated by the addition of purified PP_5260 or boiled enzyme controls, and absorbance at 340 nm was measured via a SpectraMax M4 plate reader (Molecular Devices, USA).

Initial rate measurements were directly recorded monitoring the consumption of O_2 using a Clarke-type electrode (Hansatech Oxygraph). Reaction mixtures containing $FeSO_4$(10 μM) and 2-oxoadipic acid (10-200 μM) in 100 mM Tris pH 7.0 were allowed to equilibrate to room temperature determined by a stable O_2 concentration reading of 240 μM. Addition of purified apo enzyme (100 nM) to the sealed reaction vial initiated the reaction. Initial rates were determined from the linear portion of consumption of O_2 corresponding to up to 10% consumption of the limiting reactant. No burst or lag phases were observed. All assays were performed in triplicate.

O_2 consumption measurements used to determine the reaction stoichiometry were also measured using a Clarke-type electrode. Reactions mixtures containing $FeSO_4$(10 μM) and 2-oxoadipic acid (100 or 200 μM) in 100 mM Tris pH 7.0 were equilibrate to room temperature ($[O_2]$ = 240 μM) and monitored for at least 2 minutes prior to the addition of apo enzyme (xx μM). Oxygen consumption was monitored until the signal plateaued and the O_2 concentration was stable. The observed rate was determined by fitting the data to a single exponential decay model. The reaction stoichiometry was determined by taking the ratio of the moles of O_2 consumed and the concentration of the 2-oxoadipic acid present in the reaction. Each reaction condition was performed in triplicate.

6.5.7. Oxygen labelling experiments

All reagents were exhaustively purged with argon on a Schlenk link to remove $^{16}O_2$. Anaerobic buffer was subsequently saturated with $^{18}O_2$ via gentle bubbling with $^{18}O_2$ (Sigma, 99 atom % ^{18}O). Reaction mixtures containing enzyme (1 μM), 2-oxoadipic acid (1 mM), $FeSO_4$ (10 μM) and $^{18}O_2$ saturated buffer (1.2 mM) were mixed in anaerobic sealed reaction vials using gastight syringes. The reaction was initiated by the addition of 2-oxoadipic acid. The headspace was filled with $^{18}O_2$gas. Reactions were incubated at room temperature for 2 hours before being quenched with an equal volume of methanol. Control experiments, replacing $^{18}O_2$ with $^{16}O_2$, were run in parallel. Quenched reaction mixtures were analyzed by LC-MS.

6.5.8. LCMS Analysis

All *in vitro* reactions to be analyzed via LC-MS were quenched with an equal volume of ice cold methanol and stored at -80C until analyses. Detection of 2OA and 2HG were described previously [75]. Briefly, HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using positive mode and an Atlantis HILIC Silica 5 μm column (150 x 4.6 mm) with a linear of 95 to

50% acetonitrile (v/v) over 8 minutes in water with 40 mM ammonium formate, pH 4.5, at a flow rate of 1 mL minute^{-1}.

6.5.9. Synthesis of 2-oxohexanoic acid

Synthesis of 2-oxohexanoic acid was carried out as described previously [149]. Briefly, a solution of the Grignard reagent, prepared from 1- bromobutane (500 mg, 3.68 mmol) and the suspension of magnesium (178 mg, 7.33 mmol) in THF (5 mL), was added dropwise under N_2 atmosphere to a solution of diethyloxalate (487 mg, 3.33 mmol) in THF (4 mL) at -78 °C. After the addition was complete, the reaction mixture was stirred at –78 °C for an additional 5 h. The reaction was quenched with 2 N HCl, the aqueous layer was extracted with ether, and the combined organic layer was washed with brine, dried over $MgSO_4$, and evaporated. The crude product was dissolved in acetic acid (20 mL) and conc. HCl (5 mL). After 11 h, the reaction was concentrated directly, and the residue was purified by distillation under reduced pressure to give the pure product (203 mg, 43%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) 2.92 (t, 2H), 1.81 – 1.54 (m, 2H), 1.48 – 1.28 (m, 2H), 0.92 (t, 3H) (Figure 6-12).

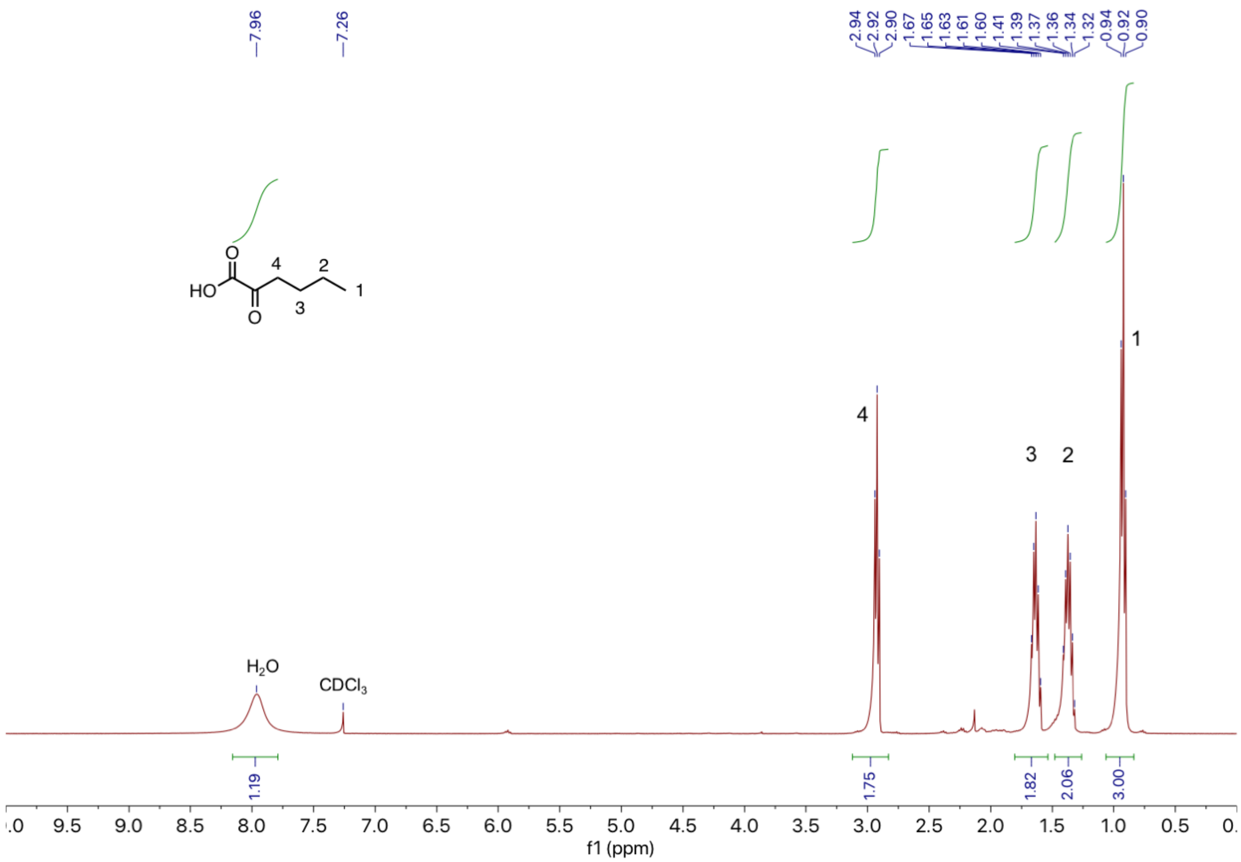


Figure 6-12 NMR of synthesized 2-oxohexanoic acid.

6.5.10. Plant Growth

Arabidopsis thaliana DUF1338 mutant, Salk_103299C, seeds were ordered from the Arabidopsis Biological Resource Center (Columbus, OH, USA). Seeds were surface sterilized, rinsed with sterile water, and planted in Premier Pro-Mix mycorise pro soil, two seeds per well in a twenty-four well tray. Trays with seeds were then stratified at 4°C for 3 days, then grown in a Percival-Scientific growth chambers at 22°C in 10/14-hour light/dark short-day cycles with 60% humidity. After the formation of the full rosette, plants were genotyped to confirm homozygosity for the mutation of interest. Primers were designed using the automated method from http://signal.salk.edu/cgi-bin/tdnaexpress, and genotyping was done in accordance with a previously described protocol [150]. After confirmation of homozygous mutants, plants were moved into individual wells and transferred to long-day conditions, 22°C in 16/8-hour light/dark cycles with 60% humidity, to induce flowering for seed collection.

6.5.11. Bioinformatics

Sequences of DUF1338 homologs were downloaded from Pfam (https://pfam.xfam.org/family/PF07063). To compare the 3D structure of PP_5260 with other protein structures we used Vector Alignment Search Tool (VAST) [127] . All alignments were done using the MAFFT-LINSI algorithm [52], and alignments were compared to secondary structures and visualized using Easy Sequencing in PostScrip (http://espript.ibcp.fr) [151]. Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 [148] . Calculation of Shannon-Entropy for DUF1338 sequence conservation was carried out using the python library Protein Dynamics and Sequence Analysis (ProDy) [152,153]. Co-expression analysis of *A. thalina* transcripts was performed using the ATTED-II database [154].

6.6. Miscellaneous

6.6.1. Acknowledgements

The authors would like to thank Dr. Robert Haushalter for his helpful discussions in preparing this manuscript. We would also like to thank Johan Jaenisch for generously providing *A. thaliana* cDNA. Python code to analyze kinetics data was provided by Sam Curran. 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. 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.

6.6.2. Collaborator contributions

Conceptualization, M.G.T., J.M.B, and J.H.P.; Methodology, M.G.T., J.M.B, J.H.P., J.F.B., P.C.M., S.C.C., N.C.H, C.B.E, E.E.K.B,.; Investigation, M.G.T., J.M.B, W.N.S., R.A.K, J.F.B., V.T.B, P.C.M., J.W.G, C.J.P, N.C.H., E.E.K.B.; Writing – Original Draft, M.G.T, J.M.B.; Writing – Review and Editing, All authors.; Resources and supervisionP.M.S., M.M., .P.D.A., and J.D.K.Competing Interests

6.6.3. Competing interests

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

Chapter 7. References

Bibliography

1. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ Microbiol* **4,** 799–808 (2002).

2. Nikel, P. I., Chavarría, M., Danchin, A. & de Lorenzo, V. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Curr Opin Chem Biol* **34,** 20–29 (2016).

3. Loeschcke, A. & Thies, S. Pseudomonas putida-a versatile host for the production of natural products. *Appl Microbiol Biotechnol* **99,** 6197–6214 (2015).

4. Jiménez, J. I., Miñambres, B., García, J. L. & Díaz, E. Genomic analysis of the aromatic catabolic pathways from Pseudomonas putida KT2440. *Environ Microbiol* **4,** 824–841 (2002).

5. Ragauskas, A. J. *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science* **344,** 1246843 (2014).

6. Chang, Y. F. & Adams, E. Induction of separate catabolic pathways for L- and D-lysine in Pseudomonas putida. *Biochem Biophys Res Commun* **45,** 570–577 (1971).

7. Chae, T. U., Ko, Y.-S., Hwang, K.-S. & Lee, S. Y. Metabolic engineering of Escherichia coli for the production of four-, five- and six-carbon lactams. *Metab Eng* **41,** 82–91 (2017).

8. Zhang, J. *et al.* Application of an Acyl-CoA Ligase from Streptomyces aizunensis for Lactam Biosynthesis. *ACS Synth Biol* **6,** 884–890 (2017).

9. Zhang, M. *et al.* Increased glutarate production by blocking the glutaryl-CoA dehydrogenation pathway and a catabolic pathway involving L-2-hydroxyglutarate. *Nat Commun* **9,** 2114 (2018).

10. Kim, H. T. *et al.* Metabolic engineering of Corynebacterium glutamicum for the production of glutaric acid, a C5 dicarboxylic acid platform chemical. *Metab Eng* **51,** 99–109 (2018).

11. Revelles, O., Espinosa-Urgel, M., Fuhrer, T., Sauer, U. & Ramos, J. L. Multiple and interconnected pathways for L-lysine catabolism in Pseudomonas putida KT2440. *J Bacteriol* **187,** 7500–7510 (2005).

12. Revelles, O., Wittich, R.-M. & Ramos, J. L. Identification of the initial steps in D-lysine catabolism in Pseudomonas putida. *J Bacteriol* **189,** 2787–2792 (2007).

13. Perfetti, R., Campbell, R. J., Titus, J. & Hartline, R. A. Catabolism of pipecolate to

glutamate in Pseudomonas putida. *J Biol Chem* **247,** 4089–4095 (1972).

14. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6,** e00306–15 (2015).

15. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557,** 503–509 (2018).

16. Rand, J. M. *et al.* A metabolic pathway for catabolizing levulinic acid in bacteria. *Nature microbiology* **2,** 1624–1634 (2017).

17. Okuno, E., Tsujimoto, M., Nakamura, M. & Kido, R. 2-Aminoadipate-2-oxoglutarate aminotransferase isoenzymes in human liver: a plausible physiological role in lysine and tryptophan metabolism. *Enzyme Protein* **47,** 136–148 (1993).

18. Wilding, M., Peat, T. S., Newman, J. & Scott, C. A β-Alanine Catabolism Pathway Containing a Highly Promiscuous ω-Transaminase in the 12-Aminododecanate-Degrading Pseudomonas sp. Strain AAC. *Appl Environ Microbiol* **82,** 3846–3856 (2016).

19. Kopchick, J. J. & Hartline, R. A. alpha-Hydroxyglutarate as an intermediate in the catabolism of alpha-aminoadipate by Pseudomonas putida. *J Biol Chem* **254,** 3259–3263 (1979).

20. Witkowski, A., Joshi, A. K. & Smith, S. Mechanism of the β-Ketoacyl Synthase Reaction Catalyzed by the Animal Fatty Acid Synthase†. *Biochemistry* **41,** 10877–10887 (2002).

21. Martinez, S. & Hausinger, R. P. Biochemical and Spectroscopic Characterization of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme from Pseudomonas syringae pv. phaseolicola PK2. *Biochemistry* **55,** 5989–5999 (2016).

22. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J Biol Chem* **290,** 20702–20711 (2015).

23. Revelles, O., Espinosa-Urgel, M., Molin, S. & Ramos, J. L. 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 (2004).

24. Danhauser, K. *et al.* DHTKD1 mutations cause 2-aminoadipic and 2-oxoadipic aciduria. *Am J Hum Genet* **91,** 1082–1087 (2012).

25. REITZ, M. & RODWELL, V. a-Hydroxyglutarate Oxidoreductase of Pseudomonas putida. *J Bacteriol* (1969).

26. Wójcik, A., Broclawik, E., Siegbahn, P. E. M. & Borowski, T. Mechanism of

benzylic hydroxylation by 4-hydroxymandelate synthase. A computational study. *Biochemistry* **51,** 9570–9580 (2012).

27. Di Giuro, C. M. L. *et al.* Chiral hydroxylation at the mononuclear nonheme Fe(II) center of 4-(S) hydroxymandelate synthase--a structure-activity relationship analysis. *PLoS ONE* **8,** e68932 (2013).

28. Lequette, Y., Odberg-Ferragut, C., Bohin, J.-P. & Lacroix, J.-M. Identification of mdoD, an mdoG paralog which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic glucan backbone structures. *J Bacteriol* **186,** 3695–3702 (2004).

29. Zhang, L. *et al.* FLOURY ENDOSPERM7 encodes a regulator of starch synthesis and amyloplast development essential for peripheral endosperm development in rice. *J Exp Bot* **67,** 633–647 (2016).

30. Vijayakumar, S. R. V., Kirchhof, M. G., Patten, C. L. & Schellhorn, H. E. RpoS-regulated genes of Escherichia coli identified by random lacZ fusion mutagenesis. *J Bacteriol* **186,** 8499–8507 (2004).

31. Knorr, S. *et al.* Widespread bacterial lysine degradation proceeding via glutarate and L-2-hydroxyglutarate. *Nat Commun* **9,** 5071 (2018).

32. Nishijyo, T., Haas, D. & Itoh, Y. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. *Mol Microbiol* **40,** 917–931 (2001).

33. Valentini, M. *et al.* Hierarchical management of carbon sources is regulated similarly by the CbrA/B systems in Pseudomonas aeruginosa and Pseudomonas putida. *Microbiology (Reading, Engl)* **160,** 2243–2252 (2014).

34. Bouffartigues, E. *et al.* Transcription of the oprF gene of Pseudomonas aeruginosa is dependent mainly on the SigX sigma factor and is sucrose induced. *J Bacteriol* **194,** 4301–4311 (2012).

35. Blanka, A. *et al.* Identification of the alternative sigma factor SigX regulon and its implications for Pseudomonas aeruginosa pathogenicity. *J Bacteriol* **196,** 345–356 (2014).

36. Neshich, I. A. P., Kiyota, E. & Arruda, P. Genome-wide analysis of lysine catabolism in bacteria reveals new connections with osmotic stress resistance. *ISME J* **7,** 2400–2410 (2013).

37. Brinkman, A. B., Bell, S. D., Lebbink, R. J., de Vos, W. M. & van der Oost, J. The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277,** 29537–29549 (2002).

38. Thaw, P. *et al.* Structural insight into gene transcriptional regulation and effector

binding by the Lrp/AsnC family. *Nucleic Acids Res* **34,** 1439–1449 (2006).

39. LaBauve, A. E. & Wargo, M. J. Growth and laboratory maintenance of Pseudomonas aeruginosa. *Curr Protoc Microbiol* **Chapter 6,** Unit 6E.1. (2012).

40. Ham, T. S. *et al.* Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res* **40,** e141 (2012).

41. Chen, J., Densmore, D., Ham, T. S., Keasling, J. D. & Hillson, N. J. DeviceEditor visual biological CAD canvas. *J Biol Eng* **6,** 1 (2012).

42. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation software. *ACS Synth Biol* **1,** 14–21 (2012).

43. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6,** 343–345 (2009).

44. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3,** e3647 (2008).

45. Shanks, R. M. Q., Kadouri, D. E., MacEachran, D. P. & O'Toole, G. A. New yeast recombineering tools for bacteria. *Plasmid* **62,** 88–97 (2009).

46. Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics*. 510 (Wiley, John & Sons, Incorporated, 2012).

47. González Fernández-Niño, S. M. *et al.* Standard flow liquid chromatography for shotgun proteomics in bioenergy research. *Frontiers in bioengineering and biotechnology* **3,** 44 (2015).

48. Batth, T. S. *et al.* A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. *Metab Eng* **26,** 48–56 (2014).

49. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968 (2010).

50. Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* **8,** 430–435 (2011).

51. George, K. W. *et al.* Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing Escherichia coli. *Metab Eng* **47,** 60–72 (2018).

52. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30,** 772–780 (2013).

53. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 — approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).

54. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44,** W242–5 (2016).

55. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9,** 75 (2008).

56. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32,** 1792–1797 (2004).

57. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

58. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. von & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14,** 587–589 (2017).

59. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32,** 268–274 (2015).

60. Cruz-Morales, P. *et al.* Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore functional characterization, unveil new desferrioxamine traits. *FEMS Microbiol Ecol* **93,** (2017).

61. Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *Neuroimage* **86,** 446–460 (2014).

62. Dietrich, J. A., Shis, D. L., Alikhani, A. & Keasling, J. D. Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth Biol* **2,** 47–58 (2013).

63. Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem* **79,** 563–590 (2010).

64. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for metabolic engineering. *Frontiers in bioengineering and biotechnology* **3,** 135 (2015).

65. 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 Synth Biol* **7,** 227–239 (2018).

66. Lei, C. *et al.* A feedback regulatory model for RifQ-mediated repression of rifamycin export in Amycolatopsis mediterranei. *Microb Cell Fact* **17,** 14 (2018).

67. Zhang, J. *et al.* Development of a Transcription Factor-Based Lactam Biosensor.

ACS Synth Biol **6,** 439–445 (2017).

68. Gilbert, M. in *Brydson's Plastics Materials* 487–511 (Elsevier, 2017). doi:10.1016/B978-0-323-35824-8.00018-9

69. Chae, T. U. *et al.* Metabolic engineering for the production of dicarboxylic acids and diamines. *Metab Eng* (2019). doi:10.1016/j.ymben.2019.03.005

70. Lee, S. Y. *et al.* A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis* **2,** 18–33 (2019).

71. Li, Z., Shen, Y.-P., Jiang, X.-L., Feng, L.-S. & Liu, J.-Z. Metabolic evolution and a comparative omics analysis of Corynebacterium glutamicum for putrescine production. *J Ind Microbiol Biotechnol* **45,** 123–139 (2018).

72. Li, M. *et al.* Improving the secretion of cadaverine in Corynebacterium glutamicum by cadaverine-lysine antiporter. *J Ind Microbiol Biotechnol* **41,** 701–709 (2014).

73. Picataggio, S. *et al.* Metabolic engineering of candida tropicalis for the production of long–chain dicarboxylic acids. *Nat Biotechnol* **10,** 894–898 (1992).

74. Thompson, M. *et al.* Host engineering for improved valerolactam production in Pseudomonas putida. *BioRxiv* (2019). doi:10.1101/660795

75. Thompson, M. G. *et al.* Massively Parallel Fitness Profiling Reveals Multiple Novel Enzymes in Pseudomonas putida Lysine Metabolism. *MBio* **10,** (2019).

76. Madhuri Indurthi, S., Chou, H.-T. & Lu, C.-D. 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 (2016).

77. Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34,** W369–73 (2006).

78. Jain, D. Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *IUBMB Life* **67,** 556–563 (2015).

79. Anderson, J. C. *et al.* BglBricks: A flexible standard for biological part assembly. *J Biol Eng* **4,** 1 (2010).

80. Siedler, S., Stahlhut, S. G., Malla, S., Maury, J. & Neves, A. R. Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into Escherichia coli. *Metab Eng* **21,** 2–8 (2014).

81. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57,** 97–109 (1970).

82. Sagawa, S., Price, M. N., Deutschbauer, A. M. & Arkin, A. P. Validating regulatory predictions from diverse bacteria with mutant fitness data. *PLoS ONE* **12,** e0178258 (2017).

83. Nitta, K. R. *et al.* Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *elife* **4,** (2015).

84. Suvorova, I. A., Korostelev, Y. D. & Gelfand, M. S. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS ONE* **10,** e0132618 (2015).

85. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* **157,** 151–161 (2014).

86. Wehrs, M. *et al.* Engineering Robust Production Microbes for Large-Scale Cultivation. *Trends Microbiol* **27,** 524–537 (2019).

87. Park, S. J. *et al.* Metabolic engineering of Escherichia coli for the production of 5-aminovalerate and glutarate as C5 platform chemicals. *Metab Eng* **16,** 42–47 (2013).

88. Rohles, C. M., Gießelmann, G., Kohlstedt, M., Wittmann, C. & Becker, J. Systems metabolic engineering of Corynebacterium glutamicum for the production of the carbon-5 platform chemicals 5-aminovalerate and glutarate. *Microb Cell Fact* **15,** 154 (2016).

89. Hagen, A. *et al.* Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth Biol* **5,** 21–27 (2016).

90. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87,** 4645–4649 (1990).

91. Bi, C. *et al.* Development of a broad-host synthetic biology toolbox for Ralstonia eutropha and its application to engineering hydrocarbon biofuel production. *Microb Cell Fact* **12,** 107 (2013).

92. Jackson, D. R. *et al.* Structural Insights into Anthranilate Priming during Type II Polyketide Biosynthesis. *ACS Chem Biol* **11,** 95–103 (2016).

93. Thompson, M. G. *et al.* Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci Rep* **8,** 1590 (2018).

94. Ravi, K., García-Hidalgo, J., Gorwa-Grauslund, M. F. & Lidén, G. Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost. *Appl Microbiol Biotechnol* **101,** 5059–5070 (2017).

95. Nogales, J., Gudmundsson, S., Duque, E., Ramos, J. L. & Palsson, B. O. Expanding

The Computable Reactome In *Pseudomonas putida* Reveals Metabolic Cycles Providing Robustness. *BioRxiv* (2017). doi:10.1101/139121

96. Kukurugya, M. A. *et al.* Multi-omics analysis unravels a segregated metabolic flux network that tunes co-utilization of sugar and aromatic carbons in Pseudomonas putida. *J Biol Chem* (2019). doi:10.1074/jbc.RA119.007885

97. Nikel, P. I., Chavarría, M., Fuhrer, T., Sauer, U. & de Lorenzo, V. Pseudomonas putida KT2440 Strain Metabolizes Glucose through a Cycle Formed by Enzymes of the Entner-Doudoroff, Embden-Meyerhof-Parnas, and Pentose Phosphate Pathways. *J Biol Chem* **290,** 25920–25932 (2015).

98. Clarke, P. H. The metabolic versatility of pseudomonads. *Antonie Van Leeuwenhoek* **48,** 105–130 (1982).

99. George, K. W. & Hay, A. G. Bacterial strategies for growth on aromatic compounds. *Adv Appl Microbiol* **74,** 1–33 (2011).

100. Otzen, M., Palacio, C. & Janssen, D. B. Characterization of the caprolactam degradation pathway in Pseudomonas jessenii using mass spectrometry-based proteomics. *Appl Microbiol Biotechnol* **102,** 6699–6711 (2018).

101. Seddon, A. P., Li, L. & Meister, A. in *Glutamate, glutamine, glutathione, and related compounds* **113,** 451–458 (Elsevier, 1985).

102. Baral, N. *et al.* Techno-economic analysis and life-cycle greenhouse gas mitigation cost of five routes to bio-jet fuel blendstocks. *Energy Environ. Sci.* (2019). doi:10.1039/C8EE03266A

103. Becker, J. & Wittmann, C. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotechnol Adv* (2019). doi:10.1016/j.biotechadv.2019.02.016

104. Rodriguez, A. *et al.* Base-Catalyzed Depolymerization of Solid Lignin-Rich Streams Enables Microbial Conversion. *ACS sustainable chemistry & engineering* **5,** 8171–8180 (2017).

105. Kohlstedt, M. *et al.* From lignin to nylon: Cascaded chemical and biochemical conversion using metabolically engineered Pseudomonas putida. *Metab Eng* **47,** 279–293 (2018).

106. Wehrmann, M., Billard, P., Martin-Meriadec, A., Zegeye, A. & Klebensberger, J. Functional Role of Lanthanides in Enzymatic Activity and Transcriptional Regulation of Pyrroloquinoline Quinone-Dependent Alcohol Dehydrogenases in Pseudomonas putida KT2440. *MBio* **8,** (2017).

107. Chen, Y. *et al.* A rapid methods development workflow for high-throughput quantitative proteomic applications. *PLoS ONE* **14,** e0211582 (2019).

108. Nærdal, I., Netzer, R., Ellingsen, T. E. & Brautaset, T. Analysis and manipulation of aspartate pathway genes for L-lysine overproduction from methanol by Bacillus methanolicus. *Appl Environ Microbiol* **77,** 6020–6026 (2011).

109. Rajvanshi, M., Gayen, K. & Venkatesh, K. V. Lysine overproducing Corynebacterium glutamicum is characterized by a robust linear combination of two optimal phenotypic states. *Syst Synth Biol* **7,** 51–62 (2013).

110. Thompson, M. G. *et al.* Glutarate metabolism in Pseudomonas putida is regulated by two distinct glutarate sensing transcription factors. *BioRxiv* (2019). doi:10.1101/557751

111. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138,** 141–143 (1984).

112. Li, L. Y., Seddon, A. P. & Meister, A. Interaction of the protein components of 5-oxoprolinase. Substrate-dependent enzyme complex formation. *J Biol Chem* **263,** 6495–6501 (1988).

113. Jones, E., Oliphant, T., Peterson, P. & Others. SciPy: Open source scientific tools for Python.

114. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng* **13,** 22–30 (2011).

115.

116. Galili, G. New insights into the regulation and functional significance of lysine metabolism in plants. *Annu Rev Plant Biol* **53,** 27–43 (2002).

117. Galili, G. & Höfgen, R. Metabolic engineering of amino acids and storage proteins in plants. *Metab Eng* **4,** 3–11 (2002).

118. Galili, G. & Amir, R. Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol J* **11,** 211–222 (2013).

119. Yang, Q.-Q. *et al.* Biofortification of rice with the essential amino acid lysine: molecular characterization, nutritional evaluation, and field performance. *J Exp Bot* **67,** 4285–4296 (2016).

120. Hartings, H., Lauria, M., Lazzaroni, N., Pirona, R. & Motto, M. The Zea mays mutants opaque-2 and opaque-7 disclose extensive changes in endosperm metabolism as revealed by protein, amino acid, and transcriptome-wide analyses. *BMC Genomics* **12,** 41 (2011).

121. Locatelli, S., Piatti, P., Motto, M. & Rossi, V. Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm

development. *Plant Cell* **21,** 1410–1427 (2009).

122. Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L. & Braun, H.-P. Amino acid catabolism in plants. *Mol Plant* **8,** 1563–1579 (2015).

123. Brownlee, J., He, P., Moran, G. R. & Harrison, D. H. T. Two roads diverged: the structure of hydroxymandelate synthase from Amycolatopsis orientalis in complex with 4-hydroxymandelate. *Biochemistry* **47,** 2002–2013 (2008).

124. Joint Center for Structural Genomics (JCSG). Crystal structure of putative hydrolase (YP_751971.1) from Shewanella frigidimarina NCIMB 400 at 1.80 A resolution. *TO BE PUBLISHED*

125. Joint Center for Structural Genomics (JCSG). Crystal structure of Putative glyoxalase superfamily protein (YP_299723.1) from RALSTONIA EUTROPHA JMP134 at 1.90 A resolution. *TO BE PUBLISHED*

126. Seetharaman, J. *et al.* Crystal structure of uncharacterized protein YdcJ (SF1787) from Shigella flexneri which includes domain DUF1338. *TO BE PUBLISHED*

127. Madej, T. *et al.* MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res* **42,** D297–303 (2014).

128. Serre, L. *et al.* Crystal structure of Pseudomonas fluorescens 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure* **7,** 977–988 (1999).

129. Grover, Z. & Ee, L. C. Protein energy malnutrition. *Pediatr Clin North Am* **56,** 1055–1068 (2009).

130. Wenefrida, I., Utomo, H. S., Blanche, S. B. & Linscombe, S. D. Enhancing essential amino acids and health benefit components in grain crops for improved nutritional values. *Recent Pat DNA Gene Seq* **3,** 219–225 (2009).

131. Le, D. T., Chu, H. D. & Le, N. Q. Improving nutritional quality of plant proteins through genetic engineering. *Curr Genomics* **17,** 220–229 (2016).

132. WHO. *Protein and amino acid requirements in human nutrition*. (World Health Organization, 2007).

133. Kawakatsu, T., Wang, S., Wakasa, Y. & Takaiwa, F. Increased lysine content in rice grains by over-accumulation of BiP in the endosperm. *Biosci Biotechnol Biochem* **74,** 2529–2531 (2010).

134. Falco, S. C. *et al.* Transgenic canola and soybean seeds with increased lysine. *Biotechnology (NY)* **13,** 577–582 (1995).

135. Betrán, F. J., Bockholt, A., Fojt, F., Rooney, L. & Waniska, R. Registration of Tx802.

Crop Sci **43,** 1891–a (2003).

136. Mertz, E. T., Bates, L. S. & Nelson, O. E. Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science* **145,** 279–280 (1964).

137. Goodacre, N. F., Gerloff, D. L. & Uetz, P. Protein domains of unknown function are essential in bacteria. *MBio* **5,** e00744–13 (2013).

138. Wood, V. *et al.* Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? *Open Biol* **9,** 180241 (2019).

139. Fellmann, C., Gowen, B. G., Lin, P.-C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat Rev Drug Discov* **16,** 89–100 (2017).

140. Yan, P., Gao, X., Shen, W., Zhou, P. & Duan, J. Parallel assembly for multiple site-directed mutagenesis of plasmids. *Anal Biochem* **430,** 65–67 (2012).

141. Jancarik, J. & Kim, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* **24,** 409–411 (1991).

142. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

143. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68,** 352–367 (2012).

144. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60,** 2126–2132 (2004).

145. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).

146. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35,** W375–83 (2007).

147. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. . at <https://pymol.org/2/>

148. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

149. Rapf, R. J. *et al.* Photochemical Synthesis of Oligomeric Amphiphiles from Alkyl Oxoacids in Aqueous Environments. *J Am Chem Soc* **139,** 6946–6959 (2017).

150. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301,** 653–657 (2003).

151. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42,** W320–4 (2014).

152. Bakan, A. *et al.* Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **30,** 2681–2683 (2014).

153. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27,** 1575–1577 (2011).

154. Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y. & Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol* **59,** e3 (2018).