ENGINEERING AND RAPID PROTOTYPING FOR BIOLOGY IN EXTREME CONDITIONS

Thesis by

Joseph T. Meyerowitz

In Partial Fulfillment of the Requirements for the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended October 05, 2022

Joseph T. Meyerowitz

ORCID: 0000-0002-3426-0885

ACKNOWLEDGEMENTS

To birth family and chosen family. You are near and far in time and space, and you are much loved.

--JTM

ABSTRACT

In this thesis we show three projects in which biological systems are engineered for increased robustness to environmental stressors such as toxic small molecules. Several lignocellulose-derived growth inhibitors commonly found in industrial feedstocks for fermentation were used to grow a panel of yeast knockouts for several efflux pumps and detoxifying enzymes. Some specific knockout strains showed slowed growth on specific growth inhibitors, while other knockout strains showed the same growth rate as the wild-type. One efflux pump was identified for vanillin, YHK8, and was overexpressed in yeast. The overexpression strain did not show an improved tolerance to vanillin, and grew more slowly than the wild-type. To regulate the expression of the vanillin pump, a sensor for vanillin was created. The starting enzyme was the wild-type qacR transcription factor, and several variants were generated using computational protein design. The designs were synthesized and tested using *in vitro* transcription-translation (TX-TL) as part of a rapid prototyping process. This rapid prototyping considerably sped up the design-build-test process. Finally, four bacteria, *Pseudomonas synxantha* 2-79, *Pseudomonas chlororaphis* PCL1391, *Pseudomonas aureofaciens* 30-84, and *E. coli* are tested against the same lignocellulose growth inhibitors. The *Pseudomonas spp.* show an improved tolerance to the growth inhibitors. We then develop some ability to engineer and prototype in all three species. A panel of promoter parts were integrated into the *P. synxantha* genome to produce a collection of test strains. These same promoter parts were also used as DNA templates for TX-TL reactions. The *in vivo* measurements of promoter strength and *in vitro* measurements show similar relative strengths between the parts, showing the *Pseudomonas*-based TX-TL systems can be used for design-build-

test activities in these non-model organisms. This alternate approach to developing tolerance, starting with a species that already has a working tolerance to the stressor in question, changes the problem to one of building engineering capabilities in the new chassis.

PUBLISHED CONTENT AND CONTRIBUTIONS

Santos, Emmanuel L. C. de los, Joseph T. Meyerowitz, Stephen L. Mayo, and Richard M. Murray. 2016. "Engineering Transcriptional Regulator Effector Specificity Using Computational Design and *In Vitro* Rapid Prototyping: Developing a Vanillin Sensor." *ACS Synthetic Biology* 5 (4): 287–295 https://doi.org/10.1021/acssynbio.5b00090

J.T.M. participated in the conception of the project, identified the protein and its inducers to be used for the computational design, participated in experimental design and testing of the wild-type protein and the engineered variants, and participated in data analysis and writing.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

The field of synthetic biology has advanced considerably over the past years. From scar-free DNA assembly (1) to *in vitro* transcription-translation (2), these advances have made it possible to tackle increasingly complex problems in engineering and control of biological systems. Artemisinin synthesis (3), vinblastine synthesis (4), and opioid synthesis (5) are examples of biochemical pathways that are now amenable to bottom-up engineering approaches despite containing dozens of individual enzymes and metabolic intermediates all requiring optimization for an acceptable yield of the final molecular product.

There are substantial challenges still for practical application of synthetic biology. While the cost of DNA sequencing and DNA synthesis continues to decline (6,7), and databases of genetic information (8), protein structure (9), and characterized genetic parts (10,11,12) continue to expand, composing these separate advances into engineered biological systems is complicated and incompletely systematized. A number of approaches have been successful on a case-by-case basis, and techniques for rapid prototyping can be used to fill the experimental gaps that invariably arise when building the next new biological capability (13).

In this thesis we first pose a challenge in improving whole-organism tolerance to growth inhibitors found in low-cost industrial feedstocks used for growing and powering large-scale fermentations of microbes. In applications such as biofuel synthesis, the cost of

feedstock is a significant fraction of the cost of the end-product (14). The concentrations of these growth inhibitors can be sufficient to sicken or kill normally robust industrial strains of yeast and bacteria, slowing or stopping production, and purification of feedstock is cost prohibitive.

In Chapter 2, we describe original work characterizing some growth inhibitors found in acid-treated lignocellulose, a common source of plant-derived sugars used for growing cells. These growth inhibitors include reactive lignin monomers and furans derived from dehydration of sugars. By screening a series of efflux pump knockouts against individual growth inhibitors, we identify possible cognate transporters and a detoxifying enzyme that could aid in improving tolerance to these small molecule stressors. Overexpression of the YHK8 vanillin transporter, however, produced a slow growing strain without improved tolerance to vanillin. The data in Chapter 2 have not been published to date.

In Chapter 3, we develop a vanillin sensor. Regulating the expression of the vanillin efflux pump might reduce the growth defect seen in the strain constitutively expressing the pump. QacR, a tetR family transcription factor, has previously been sequenced and characterized, and was used as the starting point for developing a vanillin concentration sensitive transcriptional regulator. The existing literature described a known DNA binding sequence and associated promoter, inducers that cause de-repression of the downstream gene, and crystal structures in DNA-bound and inducer-bound conformations.

In collaboration with Emzo de los Santos, we produced computational redesigns of the qacR protein to make it responsive to vanillin, one of the growth inhibitors characterized in Chapter 2. The wild-type gene did not respond to vanillin. The DNA templates for these variants were synthesized. Using an *in vitro* transcription-translation (TX-TL) system to synthesize the vanillin sensor designs, two designs were successfully identified as responsive to vanillin. This TX-TL based approach allowed us to rapidly tune the parameters of our test system, and operate a faster design-build-test cycle.

TX-TL reactions consist of a processed lysate, typically from *E. coli*, that can be frozen and thawed for later use. The lysate is a cytoplasmic fraction of soluble proteins, ribosomes, and possibly inverted membrane vesicles. This lysate is mixed with an "energy solution" containing a variety of co-factors such as nucleotide triphosphates and tRNAs, an amino acid mixture, and template DNA such as a purified plasmid. Re-comprising all of the necessary intracellular machinery of the original bacterium, the TX-TL components use the DNA template to produce RNA, then protein, at useful yields. Similar systems were used to identify the codon table in early studies of transcription and translation (15). With the open nature of the TX-TL mixture, DNA templates and inducers at different concentrations can be tested in a wide variety of conditions without the time and cost intensive steps of genetically modifying a microbe with multiple variants of a DNA-encoded circuit. This work was published in a peer-reviewed journal (16).

In Chapter 4 we describe an alternate approach to our challenge. Despite different parts of our tolerance-building work succeeding, together those parts did not produce improved tolerance overall. In collaboration with Elin Larsson, we identified three specific *Pseudomonas* species, *P. synxantha* 2-79, *P. chlororaphis* PCL1391, and *P. aureofaciens* 30-84 for testing. These soil-derived bacteria have evolved in the presence of decaying plant matter in the soil, an environment that includes exposure to these toxins. All three bacteria have been studied before for their production of anti-fungal phenazines and use in agriculture (17, 18). We hypothesized these bacteria would be more robust "chassis" to build on compared to gut-derived *E. coli*. With growth tests, we demonstrate these soil-derived pseudomonads were able to grow at concentrations that stopped the previously tested yeast and *E. coli*. Having an existing tolerance mechanism in place, the challenge changes to making these bacteria tractable for later engineering. An existing genomic integration protocol was successful in producing variants of *P. synxantha* with a variety of test DNA parts, showing at least one is tractable for *in vivo* engineering.

We produced and tested *in vitro* TX-TL systems from all three bacteria, where instead of *E. coli* lysate we use *Pseudomonas* lysate. Producing TX-TL systems from non-model organisms has been attempted before, including with *Pseudomonas putida* and *Vibrio natriegens* (19, 20). This is the first report of successful *in vitro* transcription-translation with *P. synxantha* 2-79, *P. chlororaphis* PCL1391, and *P. aureofaciens* 30-84. We evaluate several parameters of the lysate production process and reaction conditions to optimize production of a fluorescent reporter protein. We prepared DNA templates of the

same panel of parts integrated into the *P. synxantha* genome, and show that these parts produce similar strength fluorescence signals *in vivo* and *in vitro*.

Unlike some TX-TL studies, we perform these tests with full biological triplicates of lysate batches grown and processed on separate days, and with individual experimental conditions run in technical triplicate. The comprehensive dataset produced includes hundreds of experimental conditions tested across several 384-well plates, with all data collected within a few weeks after making the first batches of lysate. Including time series data, hundreds of thousands of measurements were made very quickly. Our ability to rapidly prototype in these organisms means we can test and use many of the parts and protocols that see mature and routine use in engineering *E. coli*. The work in Chapter 4 is intended for later publication in a peer-reviewed journal.

Overall, we identified an existing challenge within the domain of synthetic biology. We successfully solved individual problems related to the challenge of improving growth in extreme environments with growth-stopping levels of toxic small molecules, but didn't completely resolve the challenge. Finally, we approach the challenge from a different angle, successfully finding three new bacteria that can already grow in the extreme environment we target, and we demonstrate substantial advances in making these bacteria tractable for engineering and rapid prototyping with modern techniques.

Citations:

1) Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison 3rd, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nature Methods* 6 (5): 343–45.

2) Yim, Sung Sun, Nathan I. Johns, Jimin Park, Antonio Lc Gomes, Ross M. McBee, Miles Richardson, Carlotta Ronda, et al. 2019. “Multiplex Transcriptional Characterizations across Diverse Bacterial Species Using Cell-Free Systems.” *Molecular Systems Biology* 15 (8): e8875.

3) Ro, Dae-Kyun, Eric M. Paradise, Mario Ouellet, Karl J. Fisher, Karyn L. Newman, John M. Ndungu, Kimberly A. Ho, et al. 2006. “Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast.” *Nature* 440 (7086): 940–43.

4) Zhang, Jie, Lea G. Hansen, Olga Gudich, Konrad Viehrig, Lærke M. M. Lassen, Lars Schrübbers, Khem B. Adhikari, et al. 2022. “A Microbial Supply Chain for Production of the Anti-Cancer Drug Vinblastine.” *Nature* 609 (7926): 341–47.

5) Galanie, Stephanie, Kate Thodey, Isis J. Trenchard, Maria Filsinger Interrante, and Christina D. Smolke. 2015. “Complete Biosynthesis of Opioids in Yeast.” *Science* 349 (6252): 1095–1100.

6) Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata. Accessed 15 November 2022.

7) Hughes, Randall A., and Andrew D. Ellington. 2017. “Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology.” *Cold Spring Harbor Perspectives in Biology* 9 (1). https://doi.org/10.1101/cshperspect.a023812.

8) Sayers, Eric W., Mark Cavanaugh, Karen Clark, Kim D. Pruitt, Conrad L. Schoch, Stephen T. Sherry, and Ilene Karsch-Mizrachi. 2021. “GenBank.” *Nucleic Acids Research* 49 (D1): D92–96.

9) Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596 (7873): 583–89.

10) Nielsen, Alec A. K., Bryan S. Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A. Strychalski, David Ross, Douglas Densmore, and Christopher A. Voigt. 2016. “Genetic Circuit Design Automation.” *Science* 352 (6281): aac7341.

11) Meyer, Adam J., Thomas H. Segall-Shapiro, Emerson Glassey, Jing Zhang, and Christopher A. Voigt. 2019. “*Escherichia Coli* ‘Marionette’ Strains with 12 Highly Optimized Small-Molecule Sensors.” *Nature Chemical Biology* 15 (2): 196–204.

12) Canton, Barry, Anna Labno, and Drew Endy. 2008. “Refinement and Standardization of Synthetic Biological Parts and Devices.” *Nature Biotechnology* 26 (7): 787–93.

13) Cole, Stephanie D., Aleksandr E. Miklos, Abel C. Chiao, Zachary Z. Sun, and Matthew W. Lux. 2020. “Methodologies for Preparation of Prokaryotic Extracts for Cell-Free Expression Systems.” *Synthetic and Systems Biotechnology* 5 (4): 252–67.

14) Bastian, Sabine, Xiang Liu, Joseph T. Meyerowitz, Christopher D. Snow, Mike M. Y. Chen, and Frances H. Arnold. 2011. "Engineered Ketol-Acid Reductoisomerase and Alcohol Dehydrogenase Enable Anaerobic 2-Methylpropan-1-Ol Production at Theoretical Yield in *Escherichia Coli*." *Metabolic Engineering* 13 (3): 345–52.

15) Nirenberg, M. W., and J. H. Matthaei. 1961. "The Dependence of Cell-Free Protein Synthesis in *E. Coli* upon Naturally Occurring or Synthetic Polyribonucleotides." *Proceedings of the National Academy of Sciences of the United States of America* 47 (October): 1588–1602.

16) Santos, Emmanuel L. C. de los, Joseph T. Meyerowitz, Stephen L. Mayo, and Richard M. Murray. 2016. "Engineering Transcriptional Regulator Effector Specificity Using Computational Design and *In Vitro* Rapid Prototyping: Developing a Vanillin Sensor." *ACS Synthetic Biology* 5 (4): 287–95.

17) Mazzola, M., R. J. Cook, L. S. Thomashow, D. M. Weller, and L. S. Pierson 3rd. 1992. "Contribution of Phenazine Antibiotic Biosynthesis to the Ecological Competence of Fluorescent Pseudomonads in Soil Habitats." *Applied and Environmental Microbiology* 58 (8): 2616–24.

18) Chin-A-Woeng, T. F., J. E. Thomas-Oates, B. J. Lugtenberg, and G. V. Bloemberg. 2001. "Introduction of the phzH Gene of *Pseudomonas Chlororaphis* PCL1391 Extends the Range of Biocontrol Ability of Phenazine-1-Carboxylic Acid-Producing *Pseudomonas Spp*. Strains." *Molecular Plant-Microbe Interactions: MPMI* 14 (8): 1006–15.

19) Des Soye, Benjamin J., Samuel R. Davidson, Matthew T. Weinstock, Daniel G. Gibson, and Michael C. Jewett. 2018. "Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from *Vibrio Natriegens*." *ACS Synthetic Biology* 7 (9): 2245–55.

20) Wang, He, Jian Li, and Michael C. Jewett. 2018. "Development of a *Pseudomonas Putida* Cell-Free Protein Synthesis Platform for Rapid Screening of Gene Regulatory Elements." *Synthetic Biology* 3 (1): ysy003.

Chapter 2

ENGINEERING A SYNTHETIC MICROBIAL STRESS RESPONSE

Obtaining liquid fuels from renewable resources is one of the most important challenges facing society. Resolving this challenge would provide broad benefits, including greater national security, a sustainable transportation fuel supply, and reduced greenhouse gas emissions. Production of fuels using engineered microbial pathways is an emerging approach for improving the efficiency of biofuel production from non-food resources, but one that faces many challenges in order to be economically viable at large scale (1).

For example, one of the most important problems is the conversion of bulk cellulosic biomass to feedstock for fermentative organisms. The industrial processes used to convert lignocellulose into substituent sugars have the undesired effect of producing compounds that inhibit microbial growth and fuel production. Current techniques for removing these growth inhibitors from cellulose-derived sugar are expensive and inefficient (2).

In this work we demonstrate a high-throughput method for quantifying growth inhibition in the presence of lignocellulosic breakdown products. A high-throughput mechanism for quantitatively assaying the degree of growth inhibition in various hydrolysate and genetic backgrounds would help disentangle the multiple effects at play and identify the most important growth inhibitors and genetic contributions to tolerance. In this chapter I show that this aim has been accomplished.

We apply this method to find candidate mechanisms for engineering tolerance to lignocellulose-derived growth inhibitors**.** By screening for exacerbated growth defects of single-gene deletion strains in the presence of growth inhibitors, it may be possible to find efflux pumps or degradative enzymes that play a role in the native yeast response to these growth inhibitors. We provide preliminary data suggesting that two efflux pumps and one cinnamyl alcohol dehydrogenase may act on some target substrates.

Background

Acid pre-treatment of lignocellulose is used before fermentation to hydrolyse polysaccharide chains in the cellulose fibers of recalcitrant biomass into substituent free sugars. This process produces a number of growth inhibiting byproducts. The dominant inhibitory byproducts can be classified as carboxylic acids, furans, and phenols. Managing these inhibitors is one of the central obstacles to adoption of acid hydrolysis for general cellulose to monosaccharide conversion, and current techniques for removing these growth inhibitors from cellulose-derived sugar are expensive and inefficient (2).

In the literature, acetate and formate vary in concentration from 2-10 g/L in standard feedstock. The furans, including 2-furanoic acid, furfural, and 5-hydroxymethyl-furfural (5-HMF) are derived from overdehydration of sugars during acid hydrolysis. Furans range from 0.5-6 g/L in standard feedstock. Phenols range from 0.6-3 g/L and primarily derive from three ringed monomers of lignin, p-hydroxyphenyl, guaiacyl, and syringyl. Their derivatives include acids, aldehydes, phenol monomers, phenol dimers, and ketones (3).

These byproducts can inhibit growth and/or ethanol production in yeast at concentrations ranging from 1mM to 50mM. Growth inhibitors may produce stronger growth suppression in combination, but these effects are not well quantified. Different feedstocks and pre-treatment methods produce different quantities of different inhibitors.

To build a fementative organism that is tolerant of these growth inhibitors, we describe here some mechanisms for small molecule tolerance to another class of molecules toxic to microbes. Antibiotic resistance can serve as a good proxy for understanding different successful adaptations to small molecule stressors. Antibiotics use a variety of techniques to suppress microbial growth and many microbes have well-characterized evolutionary adaptations for coping with small-molecule stress imposed by antibiotics. From a survey of the literature we can readily see three general classes of evolved response to exogenous small molecule stressors.

The first class of response is through target mutation. By mutating a targeted structure to a resistant form, the growth inhibiting effect of a small molecule can be reduced or stopped entirely. One example of this comes from the nalidixic acid and DNA gyrase system. Several mutations in DNA gyrase have been found to stop nalidixic acid from binding to DNA gyrase, ablating a normally-lethal interaction for the microbe (4). The target mutation approach is limited to situations where the growth inhibitor's mechanism of action primarily targets a single mutable structure within the cell. For the case of lignocellulosic growth inhibitors, the precise mechanisms of growth inhibition are not known and may impact multiple components of the intracellular machinery simultaneously. In addition, partition into the cell's membranes of lipophilic growth

inhibitors may comprise an important mechanism of toxicity. The membrane as a whole involves multiple metabolic processes working in tandem, and is expected to be more difficult to engineer than a single protein due to its high degree of interconnection.

A second class of response is enzymatic degradation. For several commonly used antibiotics including hygromycin B, nourseothricin, and chloramphenicol, acetyltransferases or phosphotransferases can convert a growth inhibiting small molecule to a non-toxic one in a single step (5, 6, 7). Members of the β-lactam family of antibiotics can be inactivated by cleavage of the lactam ring by β-lactamases (8). Enzymatic degradation is an attractive option if there are readily available chemical transformations that convert a toxic molecule into a substantially less toxic molecule; however, it may be that multiple reactions are required before a growth inhibitor is effectively degraded to a non-toxic form, increasing the complexity of engineering tolerance via enzymatic degradation.

A final class of small molecule stress response comes from small molecule transporters. By pumping the toxic substance out of the cell, the growth inhibition is stopped. Carbenicillin, erythromycin, and fluoroquinolone resistances have all been found in transport-mediated forms (9). There are a number of multi-drug efflux pump superfamilies as well, a promising sign that an effective pump for lignocellulosic growth inhibitors might already exist or be readily evolved from any number of potential parent proteins (10).

There are examples of multiple classes of response arising for the same antibiotic. Tetracycline-resistant ribosomes, tetracycline degradation pathways, and tetracycline

efflux pumps have all been identified and characterized in clinical and environmental isolates (11,12). The kind of small molecule stress response resulting from evolutionary adaptation in the environment does not appear to be closely coupled to the chemical structure of the growth inhibitor, but may be a mix of environmental factors and evolutionary fortuity.

Quantifying growth inhibition

Many of the significant growth inhibitors present in lignocellulosic hydrolysate are commercially available in purified form from chemical companies including Sigma Aldrich. To quantify the degree of growth inhibition from each individual substituent of lignocellulosic hydrolysate, several growth inhibitors were purchased. They were chosen from the list of the strongest growth inhibitors described in the literature and close analogs thereof (Table 2) (3). Due to the substantial existing literature on lignocelluosic growth inhibition in *S. cerevisiae* and due to the numerous industrial applications of *S. cerevisiae*, yeast was chosen as a model organism for this work.

Compound	Type	Compound	Type
2-furaldehyde	furan	5-hydroxymethylfurfural	furan
vanillin	phenolic aldehyde	syringaldehyde	phenolic aldehyde
2-furoic acid	furan	4-hydroxybenzaldehyde	phenolic aldehyde
4-hydroxybenzoic acid	phenolic acid	2-furfuryl alcohol	furan

Table 2.1: Lignocellulosic growth inhibitors used in initial screening.

To develop a repeatable method for measuring growth inhibition in varied conditions, a technique for performing continuously monitored yeast growth in 96-well plate format

was pursued. Yeast-peptone-dextrose (YPD) medium was prepared per established protocols and 300 uL of YPD with selective agent G418 at 100 µg/mL was added to each well of a 96-well clear plastic microtiter plate. Each well was inoculated with an equal quantity of BY4741 Mat-a haploid G418-resistant *S. cerevisiae* previously resuspended from a YPD-agar plate in YPD. The yeast in the plate were allowed to grow for 24 hours in a Victor X3 plate reader heated to 30 degrees centigrade. During growth, every 6 minutes the plate reader would shake the plate vigorously for 3 seconds, then measure the optical density at 600 nanometers.

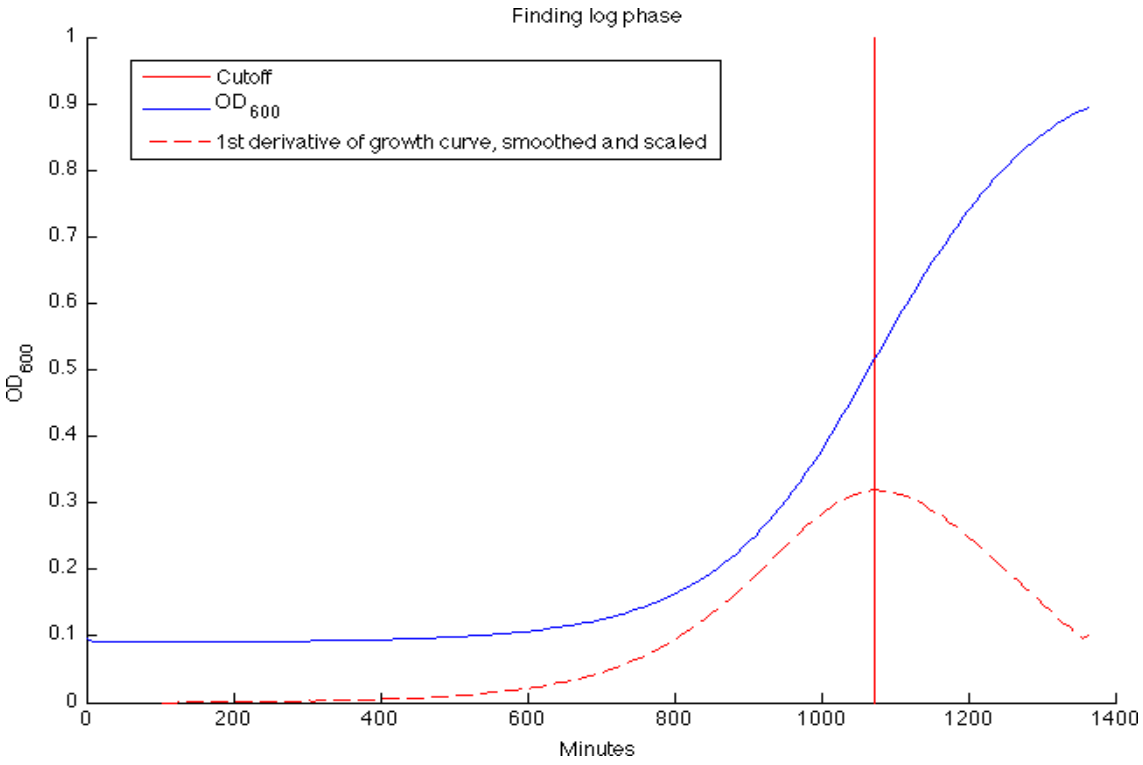


Figure 2.2: Data from one of the 96 wells in the plate. The optical density of the well can be seen in blue, the first derivative in dashed red, and the cutoff separating exponential growth from post-exponential growth in solid red.

The data from the Victor X3 plate reader was imported into an Excel format, and from that Excel format imported into MATLAB R2010a for analysis. Each well showed a classic growth curve with a lag phase, exponential phase, and a post-exponential phase. By finding the peak of the first derivative of the OD curve with respect to time, a cutoff separating the exponential phase from the post-exponential phase could be established (Figure 2.2). Curve fitting using the *fminsearch* tool in MATLAB allowed for extraction of the doubling time from the exponential phase interval on each optical density curve (Figure 2.3).

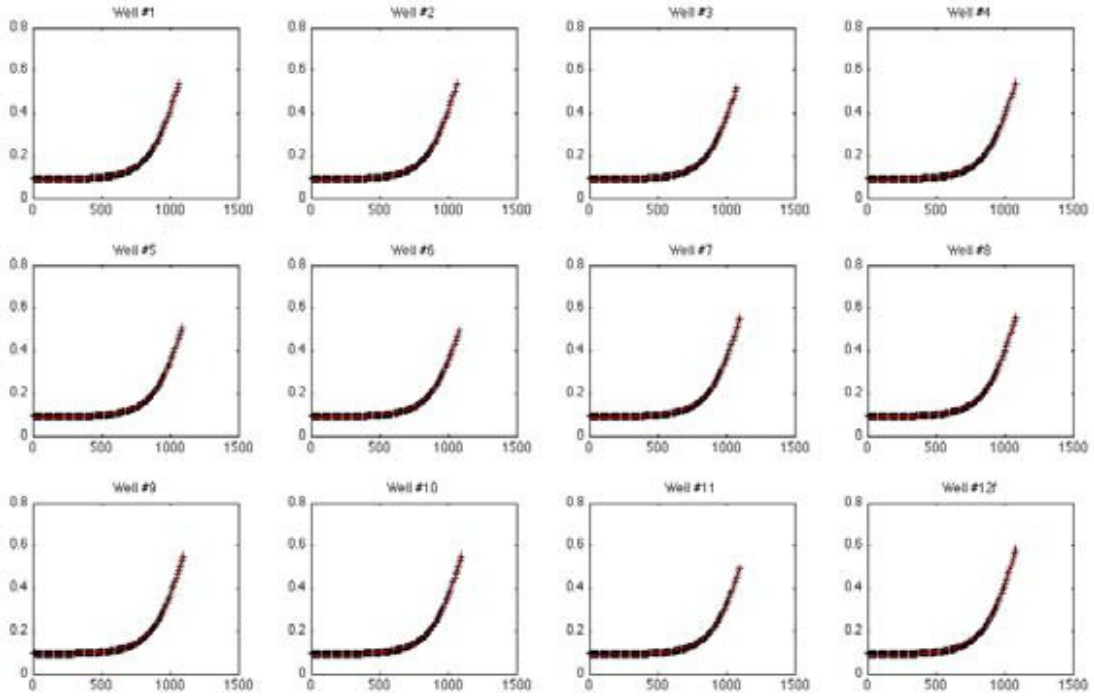


Figure 2.3: Twelve wells, truncated to show only the exponential phase of growth, all show good fits to an exponential function. The red line is the fitted curve, and the black crosses are individual optical density measurements over time.

Several issues were noted, including substantial evaporation from the wells by the end of the growth period. After some trial and error, it was discovered that starting with a larger inoculum allowed for a shorter growth period of 16 hours, reducing the amount of time

for evaporation to take place. Further, a gas permeable membrane recommended in the literature dramatically reduced evaporation and increased repeatability of results (data not shown) (18). The gas permeable membrane was not perfectly transparent, but because the transparency of the gas permeable membrane does not change over time, its impact is seen only in the time-invariant parameters in the exponential fit, leaving the doubling time unperturbed (Figure 2.4).

Figure 2.4: A filled 96-well microtiter plate with a gas permeable membrane affixed to the top.

When all 96 wells are assessed using this process under a gas permeable membrane with an initial inoculum of OD_{600} = 0.03 over a growth period of 16 hours, a narrow distribution around a doubling time of 120 minutes is evident in the data (Figure 2.5). Using this 96-well microtiter format, the degree of growth inhibition due to various lignocellulose growth inhibitors was assessed via growth measurements in varied concentrations of growth inhibitor prepared by serial dilution. Growths were performed with biological triplicates.

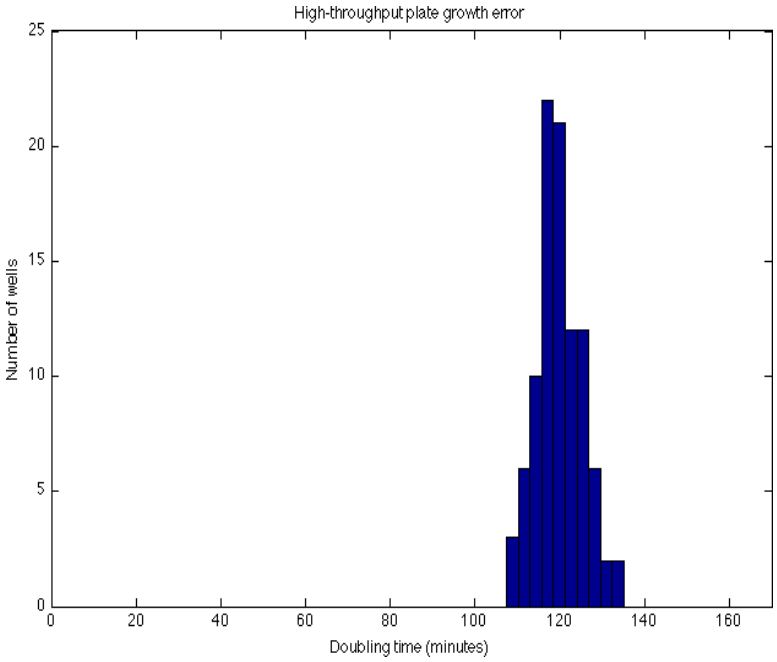


Figure 2.5: Distribution of extracted doubling times from 96 wells of the identical strain of yeast in the same media and at the same temperature in each well.

Among the furans, the results matched the expected qualitative aspects of growth inhibition as reported in the literature; namely, furfural (2-furaldehyde) and 5-hydroxymethylfurfural are moderately toxic, and similar molecules furfuryl alcohol and furoic acid are substantially less toxic (Figure 2.6). This procedure was repeated for the phenolic compounds previously mentioned. 4- hydroxybenzoic acid was insufficiently soluble in YPD to be tested for growth inhibition over 25 millimolar. The other substances showed substantial growth inhibition at lower concentrations. Of note, the three most toxic molecules tested in this panel were all phenolic aldehydes, representing each of the different lignin monomers p-hydroxyphenyl, guaiacyl, and syringyl respectively (Figure 2.7).

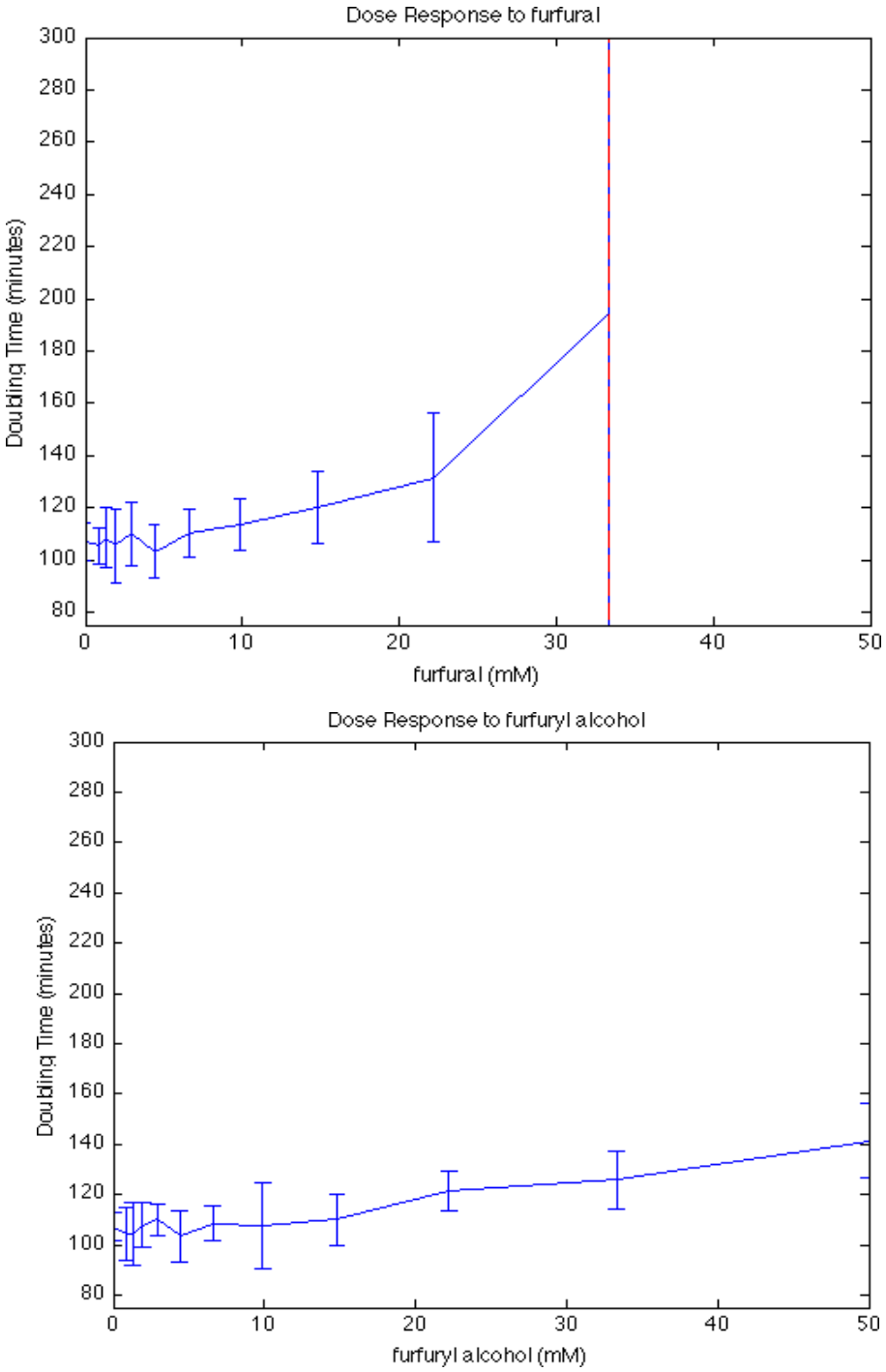


Figure 2.6: Growth inhibition due to various furans. The blue lines are interpolations between doubling times measured at the intersections with the error bars. Errors were calculated between three biological replicates measured on separate days. The red dotted line indicates the concentration above which no measurable growth was seen over 16 hours.

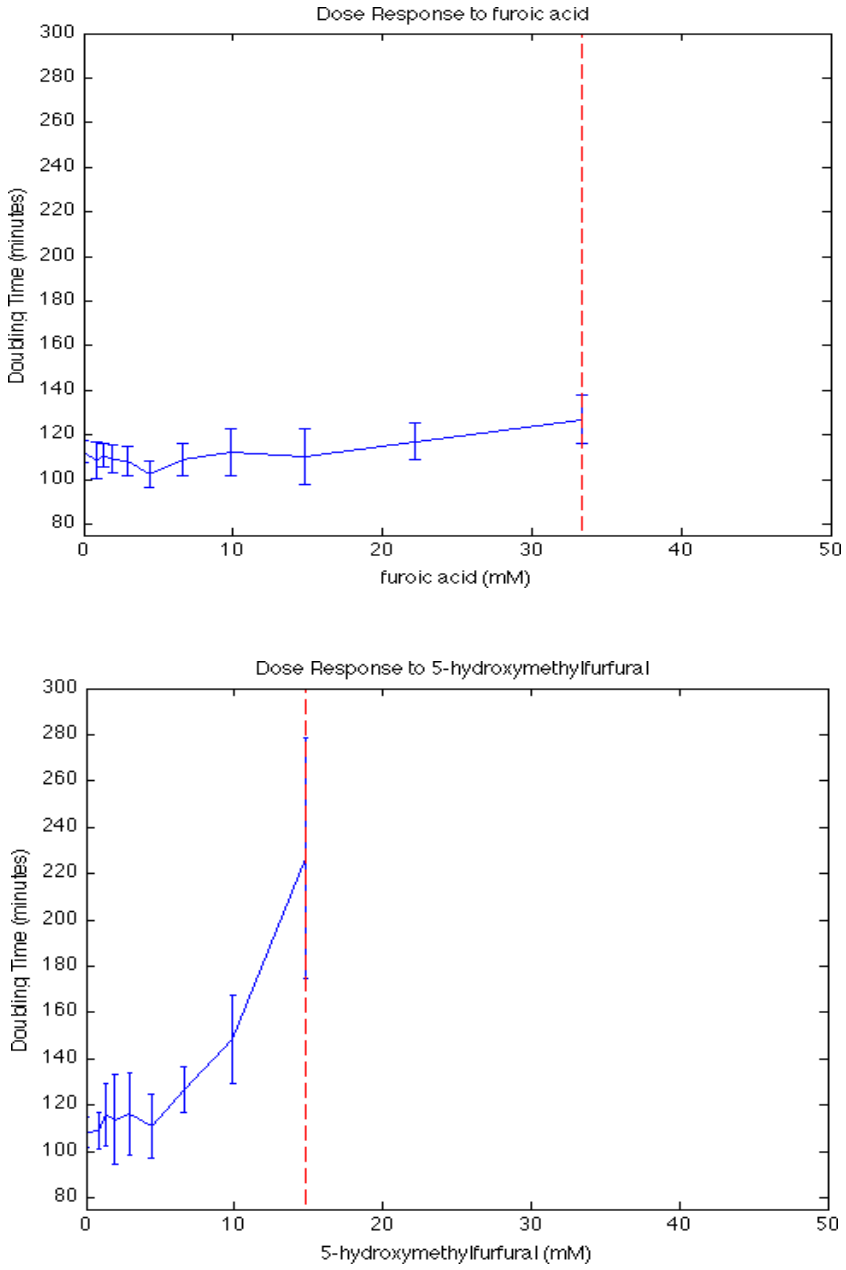


Figure 2.6 (cont'd): Growth inhibition due to various furans. The blue lines are interpolations between doubling times measured at the intersections with the error bars. Errors were calculated between three biological replicates measured on separate days. The red dotted line indicates the concentration above which no measurable growth was seen over 16 hours.

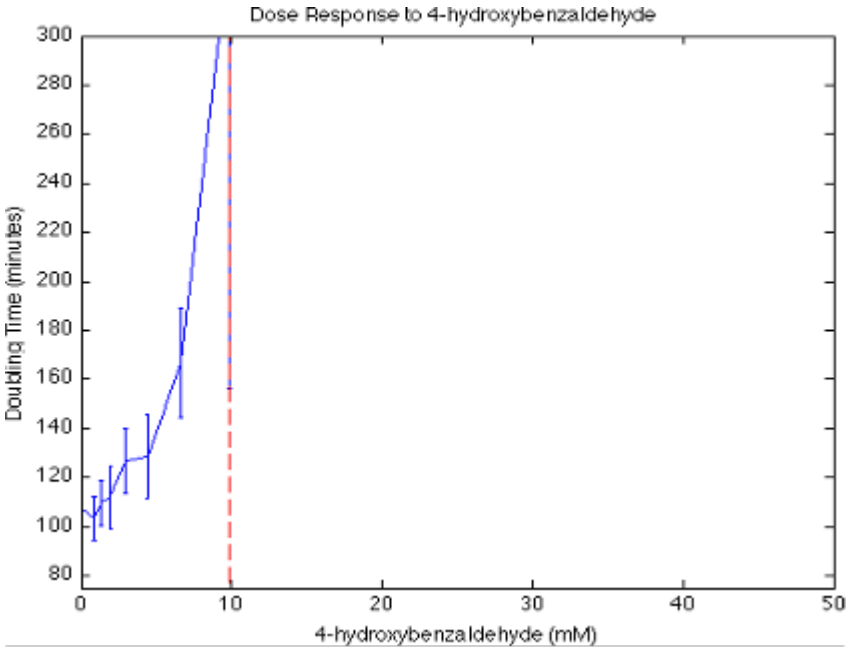


Figure 2.7: Growth inhibition due to various phenolic compounds. The blue lines are interpolations between doubling times measured at the intersections with the error bars. Errors were calculated between three biological replicates measured on separate days. The red dotted line indicates the concentration above which no measurable growth was seen over 16 hours.

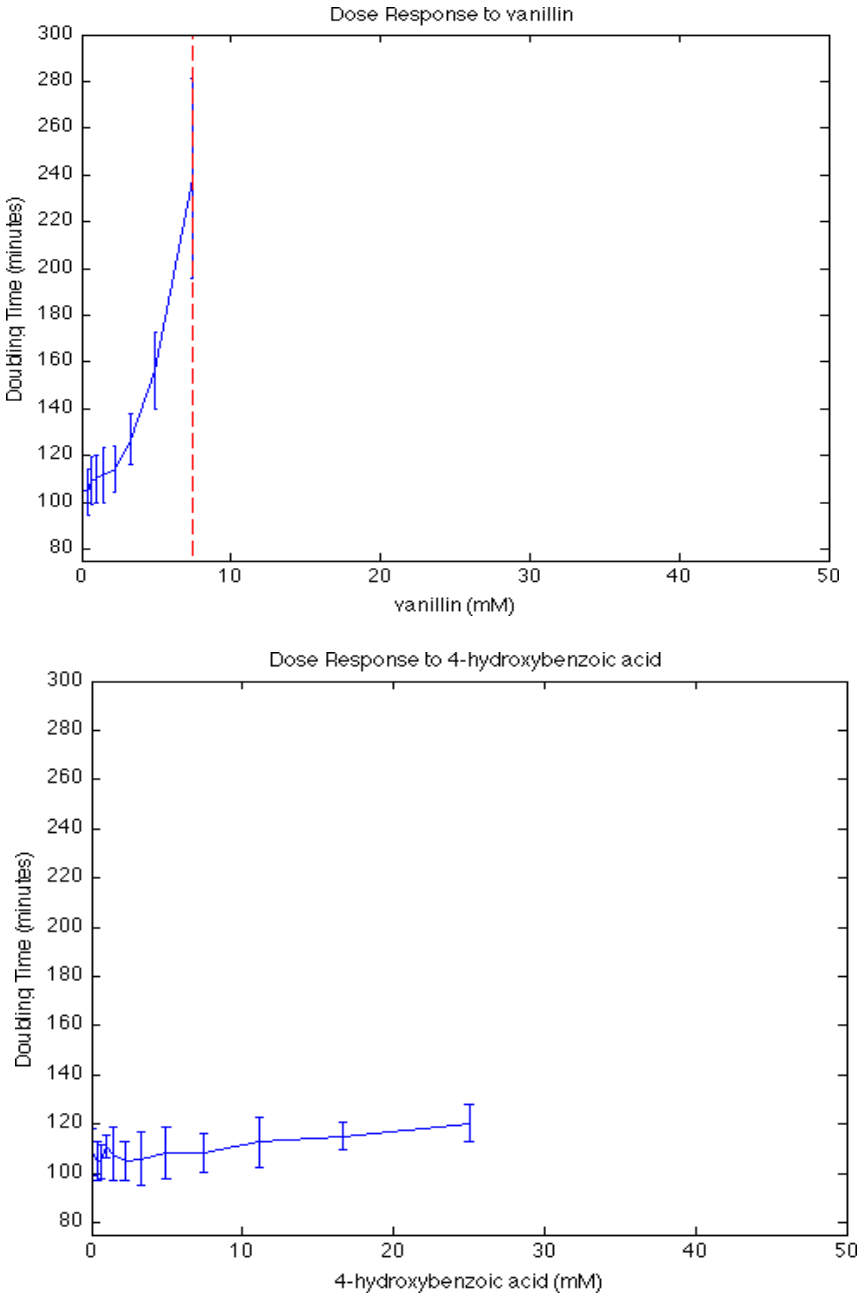


Figure 2.7 (cont'd): Growth inhibition due to various phenolic compounds. The blue lines are interpolations between doubling times measured at the intersections with the error bars. Errors were calculated between three biological replicates measured on separate days. The red dotted line indicates the concentration above which no measurable growth was seen over 16 hours.

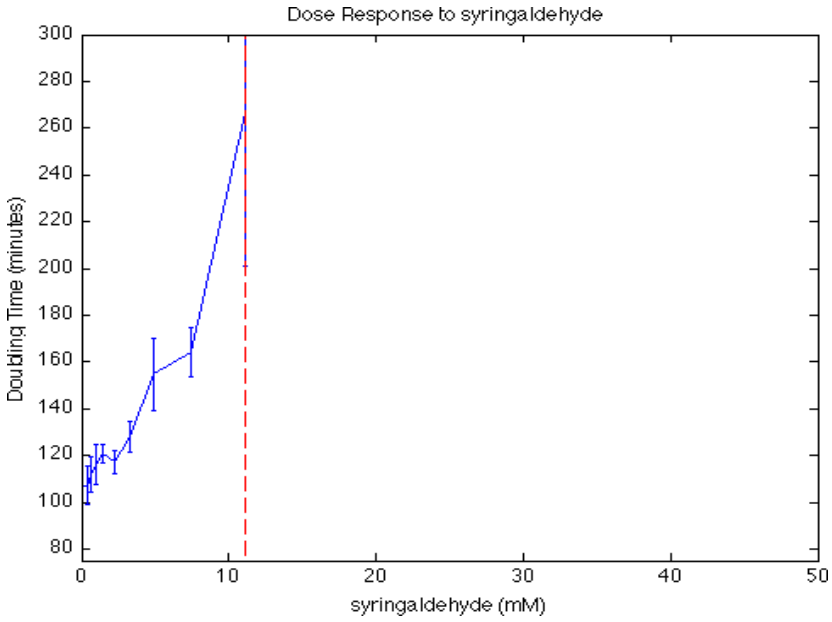


Figure 2.7 (cont'd): Growth inhibition due to various phenolic compounds. The blue lines are interpolations between doubling times measured at the intersections with the error bars. Errors were calculated between three biological replicates measured on separate days. The red dotted line indicates the concentration above which no measurable growth was seen over 16 hours.

Finding Mechanisms of Tolerance

To find potential targets for use in an engineered strain, a series of single-gene deletion mutants were screened for exacerbated growth defects in the presence of various lignocellulose- derived growth inhibitors.

In yeast, furfural and 5-HMF tolerance is associated with increased expression of three endogenous NAD(P)-linked oxidoreductases, *ADH6*, *ADH7*, and *GRE3* (19). These three deletion mutants were included in the knockout screen. Two strains were deletions of dubious ORFs that do not encode a protein, and were used for comparison as a putative G418-resistant wild-type. Twenty pumps were chosen via searches in the Saccharomyces Genome Database (yeastgenome.org) for genes annotated as multi-drug resistance pumps. The panel of twenty-five deletion mutants is described in Table 2.3.

ORF	Name	Type
YAL066W		Non-coding ORF, for use as wild-type comparison
YAR047C		Non-coding ORF, for use as wild-type comparison
YMR034C		Putative SLC10 transporter
YOR328W	PDR10	ABC multi-drug transporter
YDR338C		Putative MATE/MOP multi-drug transporter
YDR406W	PDR15	Putative ABC multi-drug transporter
YHK048W	YHK8	Putative MFS/DHA1 multi-drug transporter
YHR104W	GRE3	Multi-substrate aldose reductase
YOR153W	PDR5	ABC multi-drug transporter
YPL058C	PDR12	ABC multi-drug transporter
YPR198W	SGE1	MFS multi-drug transporter / extrusion permease
YCR105W	ADH7	NADPH-dependent cinnamyl alcohol dehydrogenase
YIL013C	PDR11	ABC multi-drug transporter
YGR281W	YOR1	ABC multi-drug transporter
YNR070W	PDR18	Putative ABC multi-drug transporter (mitochondrial)
YHR032W	ERC1	MATE/MOP multi-drug transporter
YMR318C	ADH6	NADPH-dependent cinnamyl alcohol dehydrogenase
YML116W	ATR1	MFS multi-drug transporter
YDR011W	SNQ2	ABC multi-drug transporter
YBR008C	FLR1	MFS multi-drug transporter
YBR043C	QDR3	MFS multi-drug transporter
YIL120W	QDR1	MFS multi-drug transporter
YIL121W	QDR2	MFS multi-drug transporter
YNL065W	AQR1	MFS multi-drug transporter
YNR055C	HOL1	Putative MFS/DHA1 multi-drug transporter

Table 2.3: Collection of deletion mutants used in this study, with their annotation from the literature as reported by the Saccharomyces Genome Database at yeastgenome.org

The deletion mutants in the BY4741 Mat-a haploid *S. cerevisiae* background were from the Saccharomyces Genome Deletion Project as a commercial library available from Applied Biosystems, and were kindly provided by Nathan Pierce of the Deshaies lab. Each deletion mutant was generated via homologous recombination and chromosomal integration of a geneticin resistance cassette kanMX, replacing the target gene with a selection marker.

These twenty-five deletion mutants were streaked out on YPD-agar with 100 µg/mL of the selective agent G418 and grown for two days. A loopful of each mutant was resuspended in YPD and then diluted to an OD_{600} of 0.5, and then serially diluted in a 1:10 ratio three successive times in separate tubes. 10 µL of the original resuspension and each of the serial dilutions were then pipetted in duplicate onto a large plate of YPD-agar with 100 µg/mL of the selective agent G418 and grown for two days (Figure 2.8).

Figure 2.8: All of the deletion mutants grew at a qualitatively identical rate on YPD agar with no lignocellulosic growth inhibitor. The left half of the plate is a duplicate of the right half of the plate.

This serves as a baseline measurement showing that none of the deletion mutants show dramatic growth defects when grown on rich solid media. Next, YPD-agar plates were prepared with a variety of concentrations of the six previously assayed lignocellulosic

hydrolysate derivatives which showed substantial growth inhibition in the microtiter growth assay (Table 2.4). The concentrations were chosen to be near the point where the doubling time had increased by 50% in the previous growth assays. The same procedure of serial dilution and spotting was followed for each deletion mutant, and plates were grown from 2-5 days until the control strains in the top row appeared to have grown to a similar degree as they had after 2 days on the YPD-agar only plate.

2-furoic acid	32 mM, 16 mM
5-hydroxymethylfurfural	16 mM, 8 mM
2-furaldehyde	40 mM, 20 mM
4-hydroxybenzaldehyde	8 mM, 12 mM
vanillin	7 mM, 10 mM
syringaldehyde	10 mM

Table 2.4: List of growth inhibitors and concentrations used in knockout screen.

Several deletion mutants showed dramatic growth defects on just one or two growth inhibitors, suggesting that the deleted gene might serve as a natural endogenous response in yeast for those growth inhibitors. The first example found was on the plate for 5-hydroxymethylfurfural at 8 mM (Figure 9). The Δ*YMR318C* strain was noticeably slower to grow than all other strains on the plate. YMR318C is the open reading frame for ADH6, a cinnamyl alcohol dehydrogenase previous noted in the literature to participate in 5-hydroxymethylfurfural detoxification (19).

Figure 2.9: Image of a YPD-agar plate with 8 mM 5-hydroxymethylfurfural added. Deletion mutants have been spotted on the surface of the plate. The dimmer rows, shown in duplicate on the right and left, are indicative of a growth defect due to the deletion of the ADH6 gene.

This re-discovery serves as helpful validation for this screening technique. Two other deletions were found to have a profound growth defect on specific substrates using this technique.

Efflux Pump YHK8

The next growth defect discovered was with the deletion mutant Δ*YHK048W* on the vanillin and 4-hydroxybenzaldehyde plates. YHK048W encodes a 515 amino acid putative DHA1 proton antiporter named YHK8. There are twelve known in yeast and

YHK8 was the only one with no known function as of 2011 when this study was conducted.

Efflux Pump PDR12

Another growth defect discovered was with the deletion mutant *ΔYPL058C* on the vanillin and 2-furoic acid plates. YPL058C encodes a 1512 amino acid ATP-binding cassette (ABC) transporter named PDR12. PDR12 is known to be a medium-chain alcohol and moderately lipophilic carboxylate transporter and is regulated via the weak-acid response War1p transcriptional factor (23, 24, 25).

Efflux Pump PDR18

After incubating an agar plate of deletion mutants on 12mM 4-hydroxybenzaldehyde, one strain Δ*YNR070W* survived when none of the other strains, including the wild-type strains, grew. The surviving knockout strain was for PDR18, a putative ATP-binding cassette transporter with mitrochondrial localization suggested by preparative ultracentrifugation followed by mass spectrometry (26).

It is unclear why this strain might survive. 4-hydroxybenzaldehyde is a close analog of 4-hydroxybenzoate, a molecule that is condensed with a polyisoprenoid in the ubiquinone biosynthesis pathway. If this knockout is deficient in the import of 4-hydroxybenzoate, respiratory defects might be measurable. Additionally, if 4-hydroxybenzaldehyde is acting as a competitive inhibitor for 4-hydroxybenzoic acid either at a transport or reaction step, addition of 4-hydroxybenzoate to the media may help alleviate some of the effects of 4- hydroxybenzaldehyde toxicity.

Improving tolerance via overexpression

The next step towards developing tolerant strains of yeast is the overexpression of the efflux pumps PDR12 and YHK8 and showing improved tolerance to their putative substrate growth inhibitors. As the *kanMX* cassette had already been used to generate the knockout strains, and auxotrophic selections are not possible in rich media, there were no readily available commercial vectors for this purpose. Several yeast expression shuttle vectors were thus created via Gibson Assembly, combining alternate yeast antibiotic selections, an ampicillin bacterial antibiotic selection, an *E. coli* pUC19- derived origin of replication, a *S. cerevisiae* 2μ derived origin of replication, the strong glyceraldehyde-3-phosphate dehydrogenase promoter, a yeast codon optimized red fluorescent protein (mCherry), and a terminator (Figure 2.10) (27, 28, 29, 30). By expressing mCherry by default, subsequent cloning steps should allow for visible screening for non-fluorescent inserts due to the natural red coloration imparted to the yeast by high expression of the original insert. Varied restriction sites and alternate yeast antibiotic selective markers were generated using further Gibson assembly steps with pJTM027 as a template. The alternate yeast antibiotic selective markers were derived from the pAG series of deletion marker vectors, provided by Tri Vu of the Varshavsky lab.

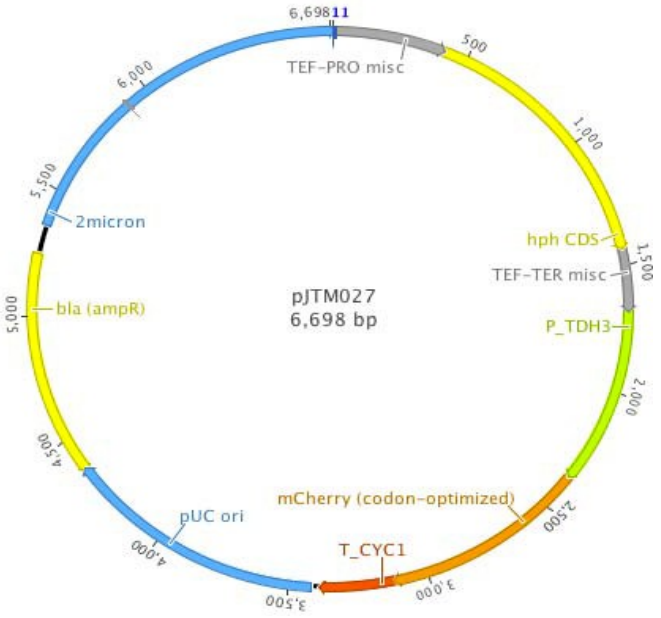


Figure 2.10: Plasmid map of the first vector successfully assembled

Plasmid Name	Promoter Restriction Site	Upstream Restriction Site	Downstream Restriction Site	Antibiotic Resistance	Constructed and sequenced
pJTM027	Sal1	Nhe1	Sal1	hygromycin	yes
pJTM031	Sal1	Nhe1	BamH1	hygromycin	yes
pJTM033	Sal1	Nhe1	Not1	hygromycin	yes
pJTM051	Sal1	Nhe1	BamH1	nourseothricin	yes
pJTM052	Sal1	Nhe1	Not1	nourseothricin	yes
pJTM055	Sal1	Nhe1	BamH1	geneticin	paused
pJTM056	Sal1	Nhe1	Not1	geneticin	paused

Table 2.5: Yeast expression shuttle vectors constructed for this portion of the study with varied cloning sites and antibiotic resistances.

YHK8 was subcloned from the yeast genome and placed in the pJTM031 background and named pJTM032. The plate growth assay was performed as previously described

with the addition of 50 µg/mL of hygromycin B as a selective agent in place of G418. By comparison with an mCherry-expressing negative control strain in the wild-type BY4741 background, the YHK8 overexpression does not appear to help alleviate vanillin toxicity (Figure 2.11). Indeed, at all vanillin concentrations the YHK8 overexpression strain appears to grow more slowly. This may be due to toxic effects from protein overexpression.

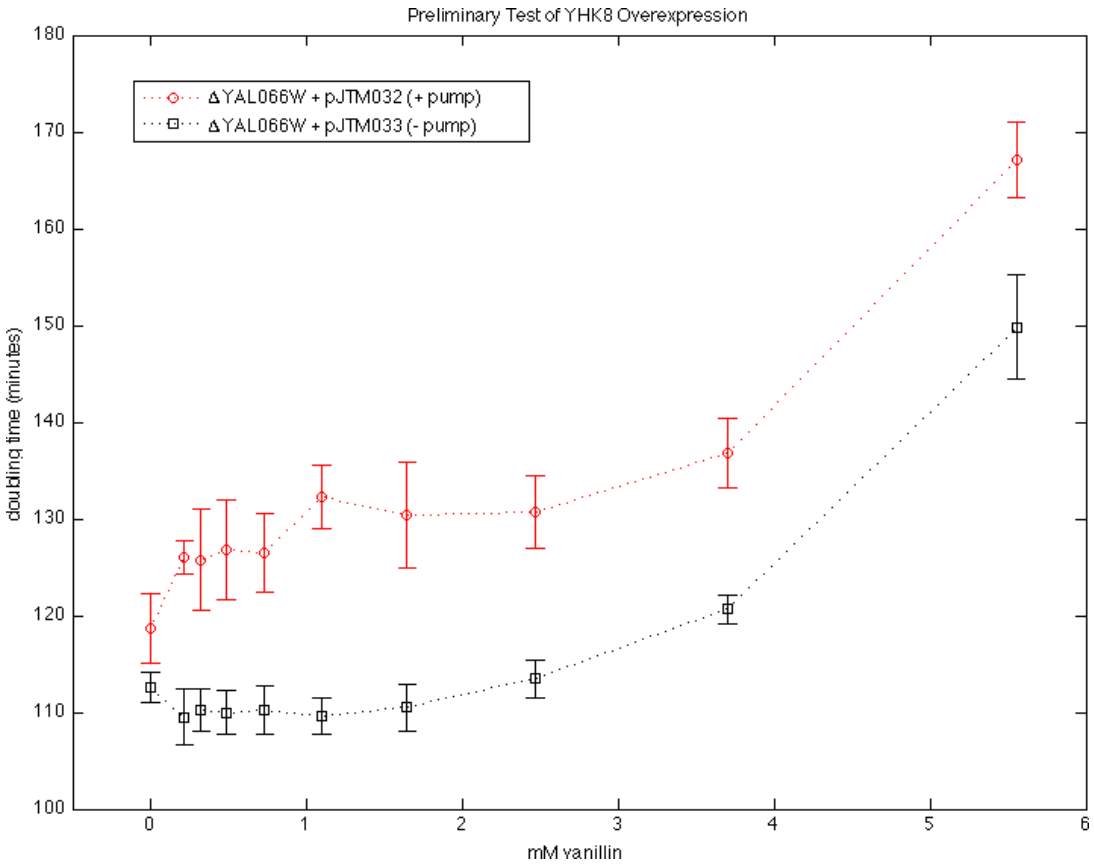


Figure 2.11: YHK8 overexpression with the TDH3 promoter does not improve vanillin tolerance.

This finding that extreme overexpression of YHK8 produces a general growth defect, and deletion of YHK8 produces a vanillin-dependent growth defect, suggested a middle ground wherein increased expression of YHK8 will improve growth of the organism

relative to wild-type in the presence of vanillin. This result motivates the next chapter of this thesis, in which a vanillin sensor was engineered to aid in the feedback control of expression.

Acknowledgements

The author is supported by the Department of Defense through a National Defense Science and Engineering Graduate Fellowship and by the National Science Foundation through a Graduate Research Fellowship. This work has been supported by a Rosen Project Grant for "Systems Engineering of Microbial Stress Response", by the U.S. Army Research Office's Institute for Collaborative Biotechnologies under contract W911NF-09-D-0001, and by Grant Number 068664 from the National Institute of General Medical Sciences.

I would like to extend special thanks to Tri Vu for the pAG series of yeast selection markers, to Nathan Pierce for the BY4741 knockout strains, to Paul Tarr and Ophelia Venturelli for helpful suggestions, to Emzo de los Santos for his help with the redesign of the qacR transcriptional factor, and to Devin Trudeau for his assistance in cloning and characterizing ADH6 and other enzymatic degradation mechanisms.

Citations:

1) F. H. Arnold, "The Race for New Biofuels", *Engineering & Science*, 2: 12-19, (2008)

2) Z. L. Liu, "Molecular mechanisms of yeast tolerance and in situ detoxification of lignocellulose hydrolysates", *Applied Microbiology and Biotechnology*, 90, 3:809-825 (2011)

3) Klinke et. al. “Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.” *Applied Microbiology and Biotechnology*, (2004)

4) Jun-ichi Yamagishi, Hiroaki Yoshida, Michiko Yamayoshi and Shinichi Nakamura “Nalidixic acid-resistant mutations of the gyrB gene of *Escherichia coli*”, *Molecular and General Genetics*, 204, 3:367-373 (1986)

5) Gritz et. al., “Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*”, *Gene* (1983)

6) Krügel et. al. “Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene nat1 from *Streptomyces noursei*”, *Gene* (1993)

7) WV Shaw, “Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria.”, *Methods in Enzymology* (1975)

8) Heffron et. al., “Origin of the TEM-beta-lactamase gene found on plasmids”, *J. Bacteriology* (1975)

9) H. Nikaido, “Antibiotic Resistance Caused by Gram-Negative Multidrug Efflux Pumps”, *Clinical Infectious Diseases* (1998)

10) Saier, M. H., Jr., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R. A., Nikaido, H. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *FASEB J.* 12, 265–274 (1998)

11) V. Burdett, “Purification and characterization of Tet(M), a protein that renders ribosomes resistant to tetracycline”, *J. Biol. Chem*. (1991)

12) Wu et. al., “Tetracycline-Resistant Clinical *Helicobacter pylori* Isolates with and without Mutations in 16S rRNA-Encoding Genes”, *Anti. Ag. Chemo*. (2005)

13) Bacher et al. “Selection and Characterization of *Escherichia coli* Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue” *J. Bateriol*. (2001)

14) Eom et al. “Enhancement of the Efficiency of Secretion of Heterologous Lipase in *Escherichia coli* by Directed Evolution of the ABC Transporter System”, *Appl. & Env. Microbiol.*, (2005)

15) Bokma et al. “Directed evolution of a bacterial efflux pump: adaptation of the *E. coli* TolC exit duct to the *Pseudomonas* MexAB translocase.” *FEBS Letters*, (2006)

16) Rem et al. “An evolved xylose transporter from *Zymomonas mobilis* enhances sugar transport in *Escherichia coli*”, *Microbial Cell Factories*, (2009)

17) Burd et al., Integrated Genomics Inc. "Methods and Compositions for Butanol Production" Patent US 7,659,105 B2, (2010)

18) Kensy et. al, "Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in *Escherichia coli* and *Hansenula polymorpha* fermentations", *Microbial Cell Factories* (2009)

19) Liu et. al. "Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*" *Appl Microbiol Biotechnol*, (2008)

20) Hirokawa T., Boon-Chieng S., and Mitaku S., "SOSUI: classification and secondary structure prediction system for membrane proteins." *Bioinformatics*, 14 378-9 (1998)

21) Käll et. al., "Reliability of transmembrane predictions in whole-genome data", *FEBS Letters* (2002)

22) Sigrist et. al. "PROSITE, a protein domain database for functional characterization and annotation." *Nucleic Acids Res*. 38:161-6 (2010)

23) Hatzixanthis K, et al. "Moderately lipophilic carboxylate compounds are the selective inducers of the *Saccharomyces cerevisiae* Pdr12p ATP-binding cassette transporter". *Yeast* 20(7):575-85 (2003)

24) Kren A, et al. "War1p, a novel transcription factor controlling weak acid stress response in yeast." *Mol Cell Biol* 23(5):1775-85 (2003)

25) Piper et. al., "The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast." *EMBO J* 17(15):4257-65 (1998)

26) Reinders J, et al. "Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics." *J Proteome Res* 5(7):1543-54 (2006)

27) Gibson et. al., "Enzymatic assembly of DNA molecules up to several hundred kilobases", *Nature Methods* (2009)

28) Mumberg D, et al. "Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds" *Gene* 156(1):119-22 (1995)

29) Goldstein, A.L., and McCusker, J.H.: "Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*." *Yeast* 15, 1541-1553 (1999)

30) Shaner et. al., "Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp*. red fluorescent protein", *Nature Biotech*. (2004)

31) Paulsen, I., "Multidrug efflux pumps and resistance: regulation and evolution", *Current Opinion in Microbiology* (2003)

32) Schumacher et. al., “Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR”, *EMBO* (2002)

33) Murray et. al., “Crystal Structures of QacR-Diamidine Complexes Reveal Additional Multidrug-binding Modes and a Novel Mechanism of Drug Charge Neutralization”, *J. Biol. Chem.* (2004)

Chapter 3

ENGINEERING TRANSCRIPTIONAL REGULATOR EFFECTOR SPECIFICITY USING COMPUTATIONAL DESIGN AND *IN VITRO* RAPID PROTOTYPING: DEVELOPING A VANILLIN SENSOR

The expression and activity membrane proteins is often tightly regulated, and overexpression of a transporter can be toxic for the cell (31, Chapter 2). If we were to overexpress the efflux pumps described above without regard for the intracellular concentration of the growth inhibitors, the cell would bear the burden of expressing the pump even in the absence of the growth inhibitor. The YHK8 over-expression strain in the previous chapter has a longer doubling time compared to the wild-type across a wide range of vanillin concentrations.

In this chapter, we describe the design of a new transcriptional factor responsive to vanillin. In the presence of vanillin, it will bind vanillin and unbind from a promoter region on a stretch of DNA, de-repressing a downstream gene and allowing expression of an engineered efflux pumps. Vanillin was chosen as the target because both YHK8 and PDR12 have data suggesting they use vanillin as a substrate as shown in Chapter 2.

Combining sensing and actuation would close the stress response loop and create a negative feedback system in which increased efflux of a growth inhibitor will reduce the concentration of the growth inhibitor. Feedback-controlled response would reduce the metabolic cost of expressing an efflux pump by reducing expression when it is not needed and increasing expression when it is needed. By connecting engineered sensing with engineered response, a new fully synthetic microbial stress response will be created.

We were successful here; excerpted below is the resulting publication in ACS Synthetic Biology.

Adapted with permission from the American Chemical Society, Emmanuel L. C. de los Santos, Joseph T. Meyerowitz, Stephen L. Mayo, and Richard M. Murray. *ACS Synthetic Biology* 2016 5 (4), 287-295. DOI: 10.1021/acssynbio.5b00090

Abstract

The pursuit of circuits and metabolic pathways of increasing complexity and robustness in synthetic biology will require engineering new regulatory tools. Feedback control based on relevant molecules, including toxic intermediates and environmental signals, would enable genetic circuits to react appropriately to changing conditions. In this work, variants of qacR, a tetR family repressor, were generated by computational protein design and screened in a cell-free transcription-translation (TX-TL) system for responsiveness to a new targeted effector. The modified repressors target vanillin, a growth-inhibiting small molecule found in lignocellulosic hydrolysates and other industrial processes. Promising candidates from the *in vitro* screen were further characterized *in vitro* and *in vivo* in a gene circuit. The screen yielded two qacR mutants that respond to vanillin both *in vitro* and *in vivo*. While the mutants exhibit some toxicity to cells, presumably due to off-target effects, they are prime starting points for directed evolution towards vanillin sensors with the specifications required for use in a dynamic control loop. We believe this process, a combination of the generation of variants coupled with *in vitro* screening, can serve as a framework for designing new sensors for other target compounds.

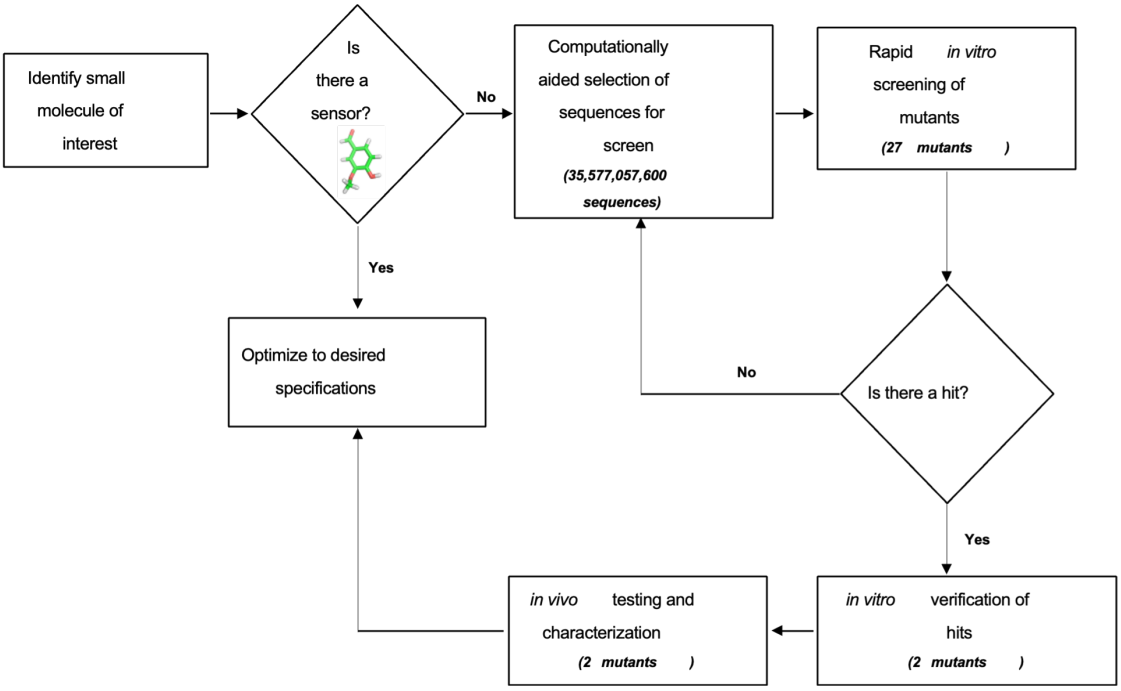


Figure 3.1: Workflow for generating novel sensors. The *in vitro* TX-TL platform allows for the rapid screening of sequences selected with the help of computational protein design. Hits from the *in vitro* screen are then verified by further *in vitro* testing. *In vivo* testing and characterization can then be performed to see if they meet the desired specifications. Further refinement of the hits through directed evolution or further computational design can be performed until specifications necessary are achieved. Numbers in parentheses are the number of sequences considered by the computational algorithm, or the number of mutants assayed at the specified step for vanillin.

Cells are able to dynamically respond to their environment. They execute genetic programs that can sense the available resources around them and alter the specific protein production based on their surroundings. These functions are mediated by a set of sensors and regulators with exact mechanisms and functions that are still being elucidated. These components form the basis for the genetic controllers that execute programs and allow for dynamic cellular response. In natural circuits, these controllers sense metabolites or external signals related to the processes they control: expression of genes in the tryptophan biosynthesis operon is controlled by a sensor for tryptophan, genes that

control antibiotic resistance genes are controlled by sensors for that antibiotic. However, synthetic circuits often use these components out of context in order to implement non-native programs in their hosts. As a result, the cell does not "know" what the program is trying to accomplish since there is no regulatory machinery in place that gives the cell information about the synthetic circuit being executed. This is problematic in the case when the synthetic circuit contains toxic intermediates, or requires the careful balancing of different proteins. To improve synthetic gene circuit performance it would be desirable to establish a framework for transmitting information about the synthetic gene circuit to the cell, analogous to the regulatory mechanisms of natural gene circuits. Feedback control based on relevant molecules, including toxic intermediates and environmental signals, would enable genetic circuits to react appropriately to changing conditions. This requires us to be able to transmit the levels of the relevant molecules to existing transcriptional control machinery.

In this work, we develop a framework to use a combination of sequence generation by computational protein design (CPD) and rapid prototyping using a cell-free transcription translation (TX-TL) system to switch effector specificity of existing transcriptional regulators to respond to targeted small molecules of interest (Figure 3.1). This approach allows us to explore a larger part of design space than traditional directed evolution approaches. In order to establish this as a general framework, we chose the transcription factor and small molecule inducer independently. The transcription factor we selected, qacR, was chosen due to the existence of crystal structures of the DNA-bound and multiple ligand-bound states and a clear mechanism for ligand sensitivity. The target small molecule inducer was vanillin, a byproduct from the lignin degradation.

Transcription factor engineering has primarily focused on modifying DNA-binding specificity (*1*) or altering the effector specificity to similar molecules from the original effector (*2*). More recently, a combination of computational protein library design and flow cytometry was used to obtain a transcription factor with altered effector specificity (*3*). However, the target small molecule effector was structurally similar to the native inducer and sequences that were already known to have sensitivity to the new effector were included in the design. To the best of our knowledge, the results presented here are the first instance of effector specificity engineering where the target small molecule effector was selected independently of the transcriptional repressor and was an unrelated, dissimilar small molecule to its natural effectors. Once sensitivity to a small molecule is established, the sensor can be further tuned using directed evolution or computational protein design until the required specifications regarding dynamic range, specificity, leakiness, and toxicity, are achieved.

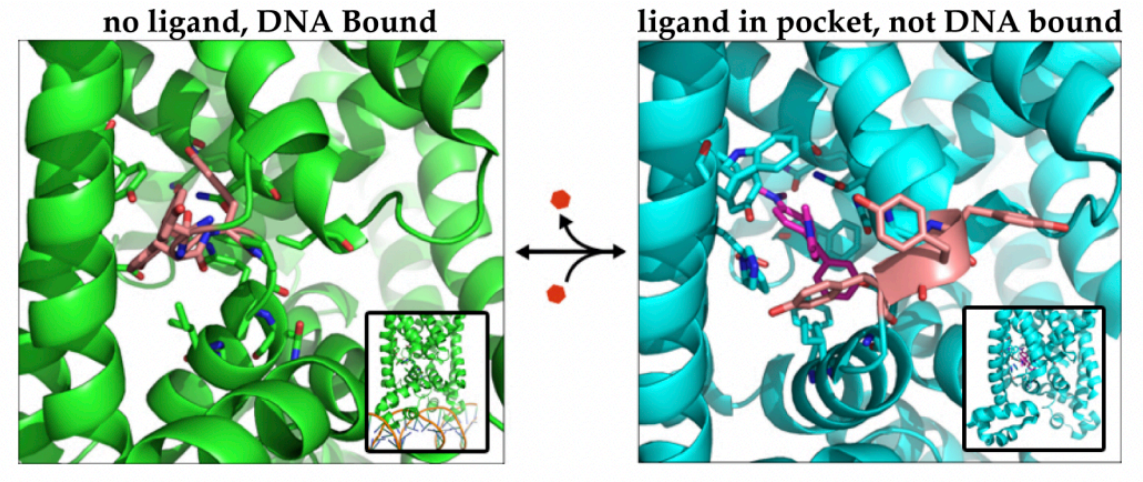
The tetR family is a large family of transcriptional regulators found in bacteria. It is named after the tetR repressor, which controls the expression of tetA, an efflux pump for tetracycline (*4*). TetR family repressors contain two domains, a helical-bundle ligand binding domain and a helix-turn-helix DNA-binding domain. In the absence of their inducing molecule, tetR repressors bind to DNA, preventing the transcription of downstream genes. Inducer binding to the ligand-binding domain causes a conformational change in the DNA-binding domain resulting in dissociation from the DNA, allowing the transcription of downstream genes. The tetR transcriptional regulation machinery has been widely and reliably used in the design of synthetic circuits, including the repressilator (*5*) and the toggle switch (*6*).

QacR is a tetR-family repressor found in *S. aureus* that controls the transcription of qacA, an efflux pump that confers resistance to a large number of quaternary anionic compounds. The protein has been studied because it is induced by a broad range of structurally dissimilar compounds (*7*). Structural examination of qacR in complex with different small molecules has shown that qacR has two different binding regions inside a large binding pocket. While qacR has multiple binding modes for various inducers, in all cases for which there are structures, binding of the inducer causes a tyrosine expulsion that moves one of the helices and alters the conformation of the DNA-binding domain, rendering qacR unable to bind DNA (*8–10*). Solved crystal structures of inducer-bound forms of qacR and the qacR-DNA complex combined with a definitive structural mechanism for qacR induction make it the ideal starting point for CPD of new transcriptional regulators.

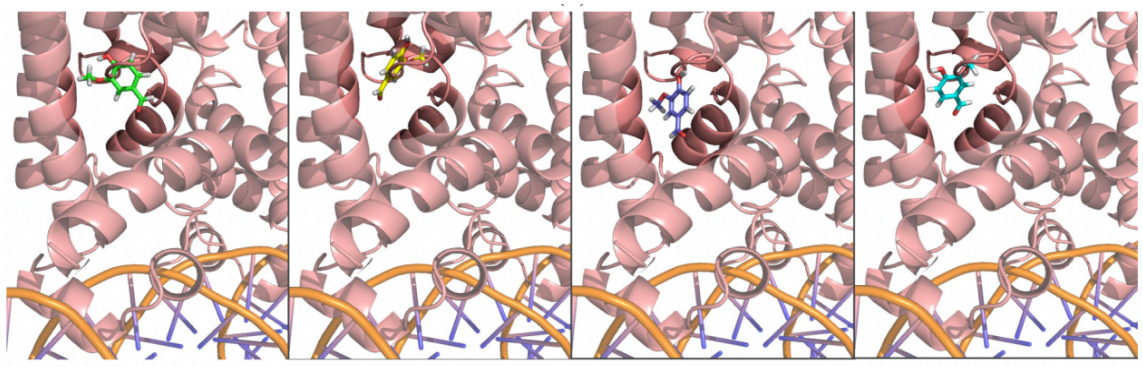
In this work, we describe our efforts to apply our framework to engineer qacR to sense vanillin, a phenolic growth inhibitor that is a byproduct of lignin degradation performed during the processing of biomass into intermediate feedstock in biofuel production (*11*).

Results and Discussion

Computationally Aided Selection of QacR Mutant Sequences



(a)



(b)

Figure 3.2: Computationally aided selection of qacR mutants. **(a)** The ligand binding pocket of the non-ligand bound (green, PDB ID: 1JT0) and ligand bound (cyan, PDB ID: 3BQZ) conformations of qacR. A conformational shift in the binding pocket occurs upon entry of the small molecule causing the protein to dissociate from DNA. The binding of the ligand in magenta causes the displacement of three tyrosine shown as sticks in pink. **(b)** Computational model for potential vanillin binding locations overlayed with the DNAbound state of qacR. The computational design algorithm looked for locations close to the tyrosine displacement to cause steric clashes between the DNA-bound state of the protein and vanillin. The computational design algorithm was asked to

suggest mutations for amino acids close to the potential binding sites to support the placement of vanillin in these sites.

We created a computational model of vanillin to place into a crystal structure of qacR.

Protein	Amino Acid Position																		
	50	54	57	58	61	86	89	90	93	96	99	102	116	119	120	126	154	157	161
qacR-wt	F	L	E	E	W	S	T	E	Y	Q	I	F	M	L	E	A	N	N	T
mutant 1 (15)	A	W	Q	L	Y	S	T	Q	Y	M	Q	S	Q	Y	Q	A	M	Q	M
mutant 2 (12)	A	W	Q	L	Y	S	T	Q	Y	Q	Q	F	Q	Y	Q	A	M	L	M
mutant 3 (13)	A	W	Q	L	Y	S	T	Q	Y	Q	I	S	Q	Y	Q	A	M	L	M
mutant 4 (12)	A	W	Q	L	Y	S	T	Q	Y	M	Q	Q	M	Y	Q	A	M	Q	M
mutant 5 (10)	A	W	E	E	Y	S	T	Q	Y	M	Q	S	Q	Y	Q	A	N	N	T

mutant 6 (8)	A	W	E	E	Y	S	T	Q	Y	Q	Q	F	Q	Y	Q	A	N	N	T
mutant 7 (8)	A	W	E	E	Y	S	T	Q	Y	Q	I	S	Q	Y	Q	A	N	N	T
mutant 8 (9)	A	W	E	E	Y	S	T	Q	Y	M	Q	Q	M	Y	Q	A	N	N	T
mutant 9 (14)	A	W	Q	L	W	S	T	Q	Y	M	Q	S	Q	Y	Q	A	M	Q	M
mutant 10 (5)	A	W	E	E	W	S	T	Q	Y	Q	I	F	M	Y	Q	A	N	N	T

Table 3.1: Initial qacR mutants chosen through computationally guided design
List of the amino acid mutations by position of the initial set of qacR mutants selected. Low energy sequences from different optimization runs were analyzed and a set of 10 mutants was selected for testing. The number of mutations from wild-type for each mutant is shown in parentheses.

Phoenix Match (*12*), a computational protein design algorithm was used to find potential vanillin binding sites close to the location of the tyrosine expulsion in the binding pocket of qacR (Figure 3.2-A) while being in the proximity of amino acid positions that allowed for favorable pi-stacking and hydrogen bonding interactions. We used targeted ligand placement (*12*) to find potential binding positions for vanillin by defining an idealized binding site for the molecule. The algorithm yielded four potential binding positions for vanillin (Figure 3.2-B). Computational protein sequence design was then used to select

amino acid residues at positions around the potential vanillin binding sites. Suggested amino acid mutations can cause steric clashes within the protein which would lock the protein in either the DNA bound or ligand bound state. In order to minimize the possibility of such steric clashes, we performed calculations that considered both the states using a multi-state design algorithm (*13*). Finally, we also ran calculations that included an energy bias to favor the wild-type residue. The lowest energy sequences from these four calculations (single-state biased, single state non-biased, multi-state biased, and multi-state non-biased) were analyzed and used as a guide to compile a set of ten mutants for *in vitro* testing. Table 3.1 shows the ten mutants tested and the number of mutations for each mutant. The number of mutations for each of the variants spanned from 5–15 mutations. A more detailed description of the computational methods used can be found in the Materials and Methods section.

In Vitro Screening of Generated Sequences

We first decided to validate function of the wild-type protein. This was done by placing green fluorescent protein (GFP) downstream of the qacA promoter sequence (P_{qacA}) to create a qacR sensitive reporter. We observed a hundred-fold repression in the GFP signal in cells containing plasmids encoding the wild-type qacR gene with the reporter but the addition of the native qacR inducer berberine yielded no observable difference (Figure S1). We hypothesized that the inducer was not getting into the cells due to the differences in cell wall permeability between gram-positive and gram-negative bacteria. Because of this, we decided to use an *in vitro* transcription-translation (TX-TL) system to test the mutants (*14*).

The TX-TL system contains whole cell lysate from BL21 *E. coli* Rosetta 2 with no endogenous mRNA or DNA. A TX-TL reaction is typically done in a 10μL reaction volume and consisting of the cell extract, an energy solution and DNA. It contains the transcriptional and translational machinery of *E. coli* necessary for protein expression from plasmid or linear DNA. Protein concentration is modulated directly by varying the amount of DNA placed in the reaction. You can execute genetic networks in a TX-TL reaction by adding plasmids that contain proteins that interact (*15*, *16*). The TX-TL prototyping provides advantages over *in vivo* circuit testing as it allows us to control protein levels without worrying about promoter and ribosomal binding site strength. Cell wall permeability and protein toxicity are also not issues in the TX-TL system.

We first tested the wild-type protein in our TX-TL system. We observed an increase in GFP fluorescence as we increased the concentration of plasmid encoding P_{QacA}–GFP from 2 nM to 8 nM (Figure 3.3-A). The addition of plasmid encoding the qacR repressor to the system resulted in a decrease in fluorescence. Because of the high autofluorescence of berberine, we used dequalinium, a colorless native qacR inducer. The addition of dequalinium resulted in an increase in fluorescence (Figure 3.3-B). These results demonstrated a functional wild-type qacR repressor in TX-TL. After validating the function of wild-type protein in TX-TL, we used the system to look at the functionality of the qacR mutants.

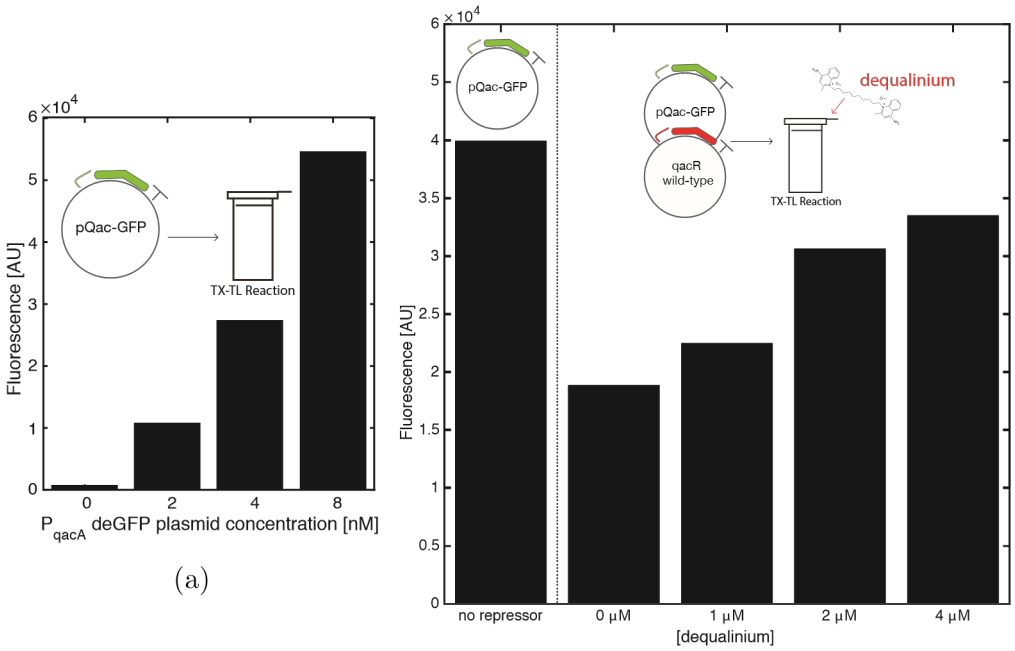


Figure 3.3: Validation of TX-TL screening. (a) GFP signal after three hours of a TX-TL reaction. Plasmid encoding GFP downstream of the native qacA promoter was added to the TX-TL platform. Higher concentrations of plasmid yielded more GFP signal. (b) Response of wild-type qacR to dequalinium. DNA encoding GFP and wild-type qacR was added to the TX-TL system. Increasing fluorescent signal is observed with increasing concentrations of dequalinium. The highest fluorescent signal is observed when there is no repressor in the system, demonstrating the ability of TX-TL to test for qacR repression and de-repression.

None of the initial mutants showed any repression of GFP fluorescence. We analyzed the ligand bound and DNA-bound computational models of mutant10, a qacR mutant that contained only three amino acid substitutions from a mutant that was previously shown to be functional by Peters *et al.* (10). The computational model showed the potential for some mutations to cause steric clashes in the DNA bound state (Figure S2). We created a

second library of 17 mutants reverting either the 50th and 54th positions (A50F/W54L) or the 119th position (Y119L) to their wild-type identity (Table S1).

In order to determine if any of the mutants of our library warranted further characterization, we performed a rapid screen of 17 qacR mutants in TX-TL (Figure S3). Plasmids containing DNA that encoded each of the qacR variants or the wild-type qacR sequence were placed into a TX-TL reaction containing either water, dequalinium or vanillin. QacR repression was monitored using the reporter plasmid (P_{qacA}-GFP). Two of the mutants, qacR2 and qacR5, displayed an increase in fluorescence in the presence of vanillin and dequalinium over water (Figure 3.4). While qacR1 also displayed an increase in fluorescence in the presence of vanillin, we did not characterize it due to low overall signal (Figure S3). We focused on these two mutants for further *in vitro* and *in vivo* characterization.

Further In Vitro Testing of QacR2 and QacR5

More extensive TX-TL tests on the mutants were conducted in order to verify the response of qacR2 and qacR5 to vanillin. TX-TL reactions were set up with a constant amount of reporter plasmid and either no repressor (water), or plasmids encoding wild-type qacR, qacR2, or qacR5. Reactions were incubated for 85 minutes at 29° C to produce the repressor protein. This bulk reaction was then added to solution containing dequalinium, vanillin, or water. We monitored the rate of GFP production between the first and third hours of the reaction, where the rate of protein production was linear.

We make the assumption that maximum GFP fluorescence for any specific inducer condition occurred when there is no repressor present. This normalizes the potential effect of

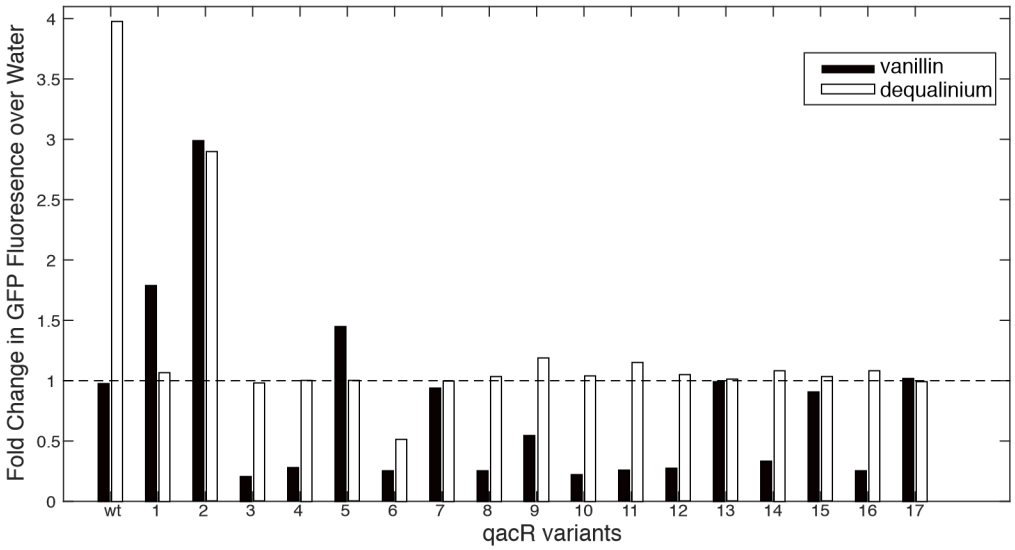


Figure 3.4: *In vitro* TX-TL screen of qacR mutants found potential candidates for further testing. Fold change in maximum fluorescence between water and inducer for qacR mutants. Seventeen qacR mutants were screened using TX-TL. Plasmids containing DNA encoding each of the qacR variants were placed into the system along with water, dequalinium (native qacR inducer) and vanillin. To monitor qacR response, a plasmid encoding GFP downstream of the native qacA promoter was also added to the system. qacR2 and qacR5 were selected for further characterization. qacR1 was not selected due to low signal (Figure S3).

toxicity of the inducer to the TX-TL reaction, as well as potential competition for resources due to the addition of input DNA encoding the repressors. Previous work has shown that competition for resources in TX-TL can inhibit protein production in otherwise orthogonal genes (*17*). The addition of the repressor-encoding DNA could utilize the resources used to produce the reporter, resulting in a decrease in fluorescence when compared to the no repressor case preventing the reaction from reaching the maximum amount of fluorescence. Conversely, repressor sensitivity to an inducer would

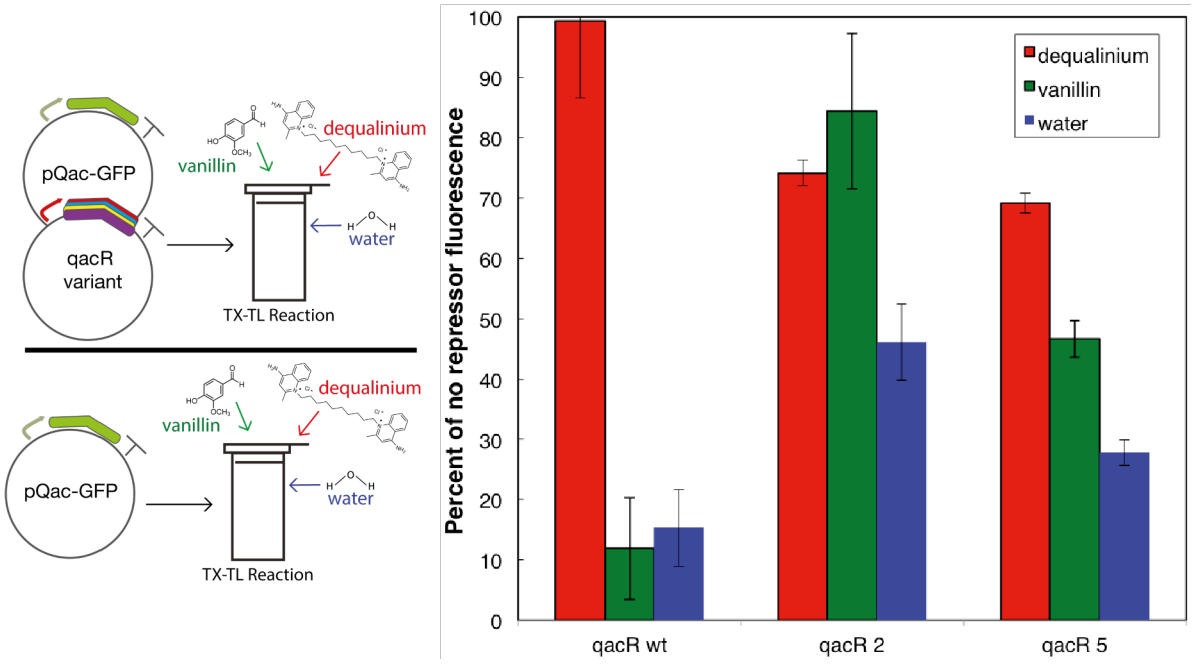
allow the production of more reporter and have a positive effect on the reaction reaching its maximum fluorescence level.

Figure 3.5 shows the response of each repressor to different inducers normalized to the no repressor case. The wild-type qacR inhibits the production of fluorescence to approximately 15% of its maximum value. The mutants are less efficient at repressing the production of GFP. Three times and four times more repressor DNA was added to the reactions of qacR2 and qacR5 respectively. Despite the additional repressor DNA, we did not observe the same level of repression as that of the wild-type protein. Wild-type qacR reacts well to the native inducer, and we observed full de-repression at the dequalinium concentration used. Induction of qacR2 and qacR5 with dequalinium is also observed, although to a lesser degree than the wild-type. QacR2 and qacR5 display a response to vanillin at the concentration we tested, while no response to vanillin was detected for the wild-type protein. The mutants show a reduced ability to repress DNA. This could be due to protein instability, or due to a weaker protein-DNA interaction. However, these mutations also increased the sensitivity of the mutants to vanillin, allowing their response to be detectable in our *in vitro* platform.

In Vivo Testing of QacR2 and QacR5

In order to further characterize the qacR mutants, and to see if vanillin could be detected in a more complex system, we tested the *in vivo* response of the qacR variants to vanillin. Plasmids containing genes that encode the wild-type qacR sequence, qacR2 or qacR5 downstream of P_{tet} and the P_{qacA}−GFP reporter were cloned into DH5αZ1 cells (Figure 3.6).

For each of the qacR variants, we compared differences in fluorescence signal across increasing vanillin concentrations from 0–1 mM. We tested different repressor concentrations by varying the amount of anhydrous tetracycline (aTc) in the system. Similar



(b)

Figure 3.5: *In vitro* testing of qacR2 and qacR5. (a) TX-TL reactions were set up with the reporter plasmid and plasmids containing the qacR variants under different inducer conditions. To account for inducer toxicity and resource limitations we normalized each condition to reactions that only contained the reporter plasmid. (b) Ratio of the rate of GFP production between TX-TL reactions with and without repressor DNA. 10 μM of dequalinium and 5 mM of vanillin was used to induce the production of GFP for each of the qacR variants tested.

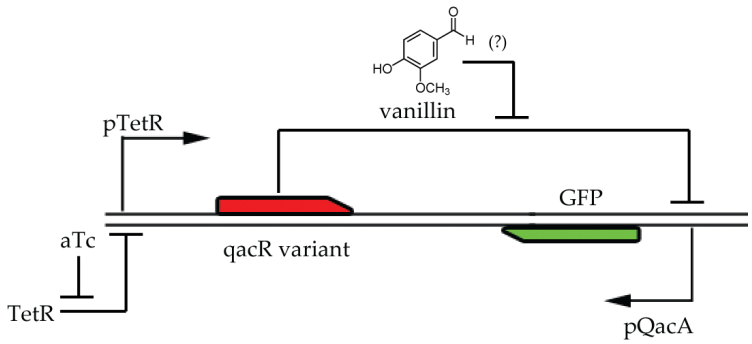


Figure 3.6: Circuit layout for *in vivo* tests. Genes encoding GFP under the control of the native qacA promoter, and our QacR designs under the control of a tet-inducible promoter were placed in a single plasmid and transformed into DH5αZ1 cells. qacR levels were controlled using aTc for varying vanillin concentrations. Candidate designs that are responsive to vanillin should show an increase in fluorescence with increasing vanillin concentrations.

to the *in vitro* experiments, and in order to get an idea for the maximum fluorescence the system could achieve, we grew cells that only contained the reporter without any repressor. The reporter only strain was used as a control for normalization of signal. Cells that were grown in higher aTc concentrations had a lower measured optical density (OD), indicating a slower doubling time. We hypothesize that this is due to the toxicity of the qacR repressor to the *E. coli* strain. Since qacR is not a native protein, it is possible that qacR is binding to locations in the *E. coli* genome. Interestingly, the differences in optical density measurements become less pronounced with increasing vanillin concentration, suggesting that vanillin may provide a mitigating effect to this toxicity. Figure 3.7 shows the different optical densities for cells at different aTc and vanillin concentrations. In order to account for differences in OD, fluorescence measurements were normalized to OD.

The lowest OD measurements were observed for cells encoding the wild-type qacR at 12 ng/mL aTc where very little growth was observed for cells expressing the wild-type

protein. At this aTc concentration, all of the cells expressing repressor exhibited lower optical densities when compared to cells that were only expressing fluorescent protein.

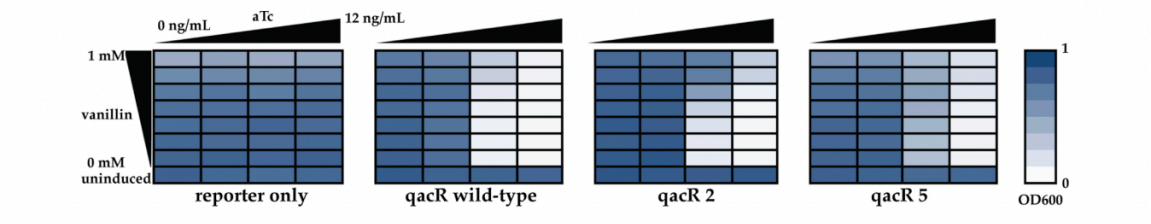
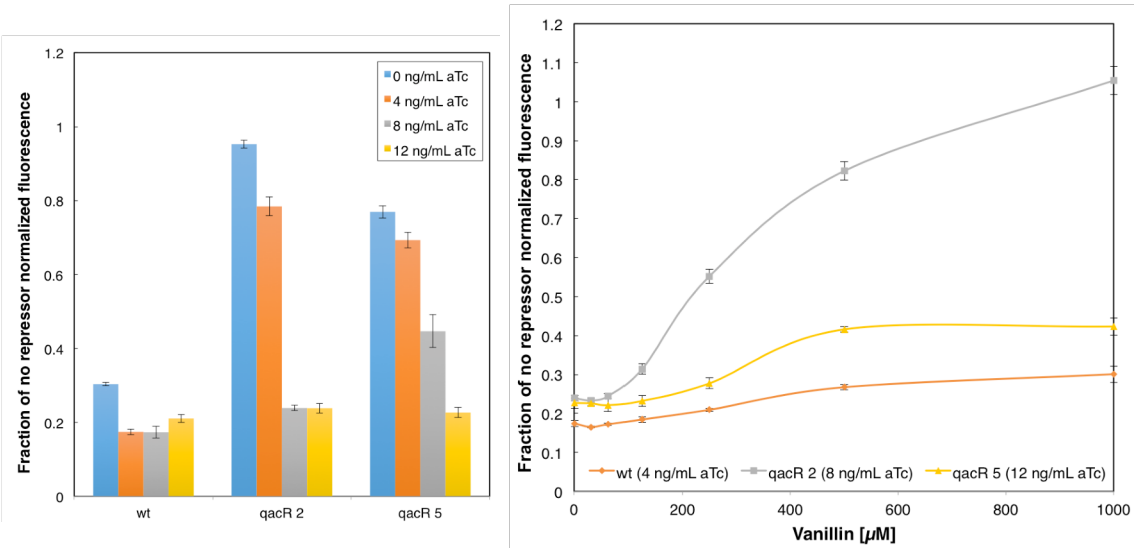


Figure 3.7: Effect of different vanillin and aTc concentrations on cell growth. Heatmap of the average OD600 of each cell strain under different aTc and vanillin concentrations. Both vanillin and the expression of the protein have a negative effect on the optical density of the cells. The growth defect is less pronounced at higher vanillin concentrations suggesting that vanillin may have a mitigating effect on protein toxicity.

The differences in optical density are less pronounced at lower aTc concentrations. When no aTc is present in the system, cells at the higher vanillin concentrations had lower ODs. At higher aTc concentrations, cells at higher vanillin concentrations had higher ODs. This implies that both the vanillin concentration and the expression of the repressor have an effect on cellular growth. Figure 3.8 summarizes the different effects of both inducer and protein production on the growth rate of the cells. The optical densities for the cells at different aTc and vanillin concentrations are shown in Tables S2-S5.

Figure 3.8-A shows the effect of increasing the aTc concentration on the fluorescence of cells in the absence of vanillin. Similar to the *in vitro* tests, fluorescence was normalized to the no repressor case. Increasing the aTc concentration decreased the fluorescence of cells in the absence of vanillin, confirming that the qacR mutants are able to repress the expression of GFP at higher protein concentrations.

The response of wild-type qacR, qacR2, and qacR5 to increasing vanillin concentrations is shown in Figure 3.8-B. The response curves for each protein are plotted for the minimum aTc concentration that high GFP repression is observed. This corresponds to aTc concentrations of 4, 8, and 12 ng/mL for wild-type qacR, qacR2, and qacR5 respectively. This is consistent with the *in vitro* data that more qacR2 and qacR5 DNA was required to repress the expression of GFP. Similar to the *in vitro* tests, we expected the ability of the cell to reach the maximum fluorescence level to be dependent on its response to inducer, and toxicity from vanillin and qacR. Indeed, cells expressing the qacR mutants exhibited an



(a) (b)

Figure 3.8: *In vivo* response of qacR to vanillin**.** Cells expressing GFP without any repressor were used as a control to normalize for differences in fluorescence due to aTc and vanillin levels. (a) All of the proteins are able to repress the expression of GFP. The wild-type protein is able to inhibit the expression of GFP at lower aTc concentrations, while higher aTc concentrations are necessary for the mutants to achieve a similar level of repression. (b) QacR mutants respond to vanillin in a concentration dependent manner.

increase in fluorescence with increasing vanillin levels demonstrating that they are capable of sensing vanillin. While all three proteins appear to be sensitive to vanillin, the mutants exhibit a marked increase in sensitivity to vanillin. QacR2 displays a response that goes from approximately 20 % of the maximum to matching the fluorescence of the non-repressed control cells at 1 mM vanillin. QacR5 saturates at around 40 % of the maximum. This correlates with the *in vitro* data that show qacR2 achieving close to the non-repressed fluorescence, with qacR5 less sensitive to vanillin (Figure 3.8). Figures S4 and S5 show the vanillin dosage response of qacR wild-type, qacR2, and qacR5 for different aTc concentrations tested.

Framework Enables Engineering of Sensors through Rational Reduction of Design Space

The framework developed— a combination of sequence generation using computationally aided design, preliminary screening with TX-TL, and *in vitro* and *in vivo* validation—can be applied to other small molecule targets and be used to facilitate the design of customized sensors for synthetic circuits. While it is possible that the computational model of vanillin binding was inaccurate, the computational design provided value in drastically reducing the pool of candidates to test into an experimentally tractable number. Without the computational design to reduce the size of the design space, we would not have had starting points to even attempt the engineering of a vanillin sensor.

The use of the *in vitro* cell-free system in a preliminary screen provides many advantages. It allows the screening of more mutants in a shorter amount of time and also

reduces the number of variables to consider. Complicating factors such as cell membrane permeability and cell growth do not need to be considered during this part of the screen. Repressors whose native inducers cannot enter the target organism can be used as starting points with the cell-free system. Finally, we can use this framework to target molecules that are known to be toxic to cells and measure engineering results in a cell-free context.

This approach is not without its limitations. Pre-existing crystal structures or accurate homology models of the starting repressors are necessary in order to use computational protein design to aid with the selection of mutants for *in vitro* selection. Having crystal structures for both the DNA-bound and ligand-bound states as well as a mechanism for ligand sensitivity reduced the design space in determining potential vanillin binding sites and allowed for troubleshooting the initial set of designs. These criteria limit the set of transcription factors that can be used as starting points for this workflow. The TX-TL platform also does not account for protein toxicity and did not predict the toxicity of qacR to cells. This is a factor that also must be taken into consideration when selecting the starting transcription factor to engineer. Due to the toxicity and low dynamic range of the sensors, it is likely that further refinement of the sensors is necessary via directed evolution or further design before they can be used practically in a dynamic feedback circuit.

However, the final result of mutant repressors with an increased sensitivity to vanillin demonstrates the value and potential utility of this workflow, particularly since the repressor and small molecule inducer were chosen independently. As a result of this process, we now have functional vanillin sensors that can be used as starting points for

directed evolution to design a feedback circuit that dynamically responds to vanillin concentration. Containing 11 and 8 mutations from the wild-type protein respectively, qacR2 and qacR5 would have been out of the range of access of sequence space for a typical directed evolution setup. While this work focused on *E. coli*, recent work has developed a process that facilitates the transfer of prokaryotic transcription factors into eukaryotic cells, increasing the flexibility of the molecules for use in metabolic engineering (*18*). By linking a vanillin sensor to the expression of a gene that can mitigate the toxic effect of vanillin, such as an efflux pump or an enzyme which converts vanillin to a less toxic molecule, we can design a dynamic feedback circuit and potentially improve metabolic yield.

Methods

Computationally Aided Selection of Mutant Sequences

An *in silico* model of vanillin was constructed using the Schrödinger software suite. Partial charges for vanillin were computed using Optimization in Jaguar version 7.6 (*19*) using HF/6311G** as the basis set. Vanillin rotamers were chosen by looking at the ideal angles for the carbon hybrid orbitals. A model of an idealized vanillin binding pocket was designed by looking at the protein data bank for proteins that bound small molecules similar to vanillin, specifically PDBID 2VSU. Models of vanillin in the qacR binding pocket were generated using the Phoenix Match algorithm (*12*).

Vanillin was built off a native tyrosine residue (Y123), the primary interaction considered for the algorithm was a hydrogen bonding interaction between the hydroxyl group of the tyrosine with the methoxy and hydroxyl groups of the vanillin. We modified the energy

function to include an energy bias of –100 kcal/mol for potential pi-stacking interactions between vanillin and tyrosine, phenylalanine, or tryptophan residues. We also included an energy bias hydrogen-bonding interactions with the methoxy, hydroxyl, and aldehyde groups of vanillin with serine, threonine, tyrosine, glutamine, or asparagine residues. The geometry ranges used for the energy bias function can be found in Tables S7-S10. The Phoenix Match algorithm was asked to return potential vanillin binding locations that contained interaction with the native tyrosine, at least one pi-stacking interaction, and at least two other hydrogen bonding interactions. Solutions from the algorithm were grouped together and resulted in four potential spots for vanillin. These locations were used as vanillin "rotamers" for computational protein design.

Monte Carlo with simulated annealing (*20*) and FASTER (*21*) were used to sample conformational space. A backbone independent conformer library with a 1.0 ˚A resolution was used for the designed residues (*12*). Designed residues were chosen by compiling a list of amino acid residues within 15 ˚A of vanillin and interacting residues from the Phoenix Match algorithm and visually inspecting the crystal structure. Table S2 shows the amino acid design positions, and the allowed amino acid residues for each position. Allowed amino residues for each site were selected by inspection of the qacR crystal structure with the potential vanillin binding locations. Rotamer optimization was allowed for other residues in the 15 ˚A shell for which mutations were not allowed. Computational models of qacR with vanillin present were scored using the PHOENIX forcefield (*12*) with the inclusion of an additional geometry bias term that favored pi-stacking and hydrogen bonding interactions that we used to find potential vanillin active sites. Geometry definitions for the geometry bias term are shown in Tables S7-S10. We

considered solutions that both included and excluded –20 kcal/mol wild-type bias term in the energy function.

We performed rotamer optimization calculations both using a single state of qacR which included the ligand based on PDB ID:3BQZ, and using two states, the ligand bound state and the DNA-bound state (PDB ID: 1JT0). In both cases, we asked the algorithm to return low energy solutions using the PHOENIX forcefield. In the multi-state design case, the score was the sum of the scores between the ligand-bound and DNA-bound states equally weighted. For each of the four rotamer optimization conditions (single-state wild-type biased, single-state unbiased, multi-state wild-type biased, multi-state unbiased), we analyzed the sequences of the 5 lowest energy solutions. We classified mutations some mutations as aggressive (e.g. polar to hydrophobic) and we compiled a list of ten initial mutants to test which considered the combinations of mutations from the rotamer optimization calculations and the number of aggressive mutations ensuring that variants that contained high and low numbers of aggressive mutations were included. Figure S6 shows a computational model of vanillin in the ligand-binding pocket of mutant2 and mutant5 (qacR2 and qacR5 without the A50F/W54L reversion mutation).

Cell Free In Vitro Transcription-Translation System and Reactions

The transcription-translation reaction consists of crude cytoplasmic extract from BL21 Rosetta 2 *E. coli* (*14*). Preliminary tests were done with plasmids and inducers at the specified concentrations. For the initial screen, the qacR mutants were downstream of a T7 promoter. TX-TL reactions were run with 2 nM of the plasmid encoding the qacR variant, 0.1 nM plasmid encoding T7 RNA polymerase, and 8 nM plasmid encoding

P_{qacA}–deGFP. Vanillin was added at a concentration of 2.5 mM and dequalinium was added at 10 μM.

For the *in vitro* tests to further characterize the hits, plasmids encoding qacR2 or qacR5 downstream of a tet-responsive promoter were used along with a plasmid encoding deGFP downstream of a qac-responsive promoter. Plasmids were prepared using the Macherey-Nagel NucleoBond Xtra Midi/Maxi Kit. Plasmid DNA was eluted in water and concentrated by vacufuge to the desired concentration. TX-TL reactions were set up as follows: 5 μL of buffer, 2.5 μL of cell extract and 1.5 μL repressor DNA at a specific concentration was mixed and incubated at 29° C for 75 minutes to facilitate the production of repressor DNA. This mix was then added to a mixture of 1 μL deGFP plasmid and 1 μL of an inducer stock. Measurements were made in a Biotek plate reader at 3-minute intervals using excitation/emission wavelengths set at 485/525 nm. Stock repressor plasmid concentrations were 243 ng / µL, 729 ng / µL, and 972 ng / µL for qacR wild-type, qacR2, and qacR5, respectively.

The deGFP plasmid concentration was approximately 397 nM. Inducer concentrations were 5 mM for vanillin, and 10 μM for dequalinium.

Experimental conditions were done in triplicate and the error bars are the error propagated from the standard deviation of the means.

Cell Strain and Media

The circuit was implemented in the *E.coli* cell strain DH5αZ1, a variant of DH5α that contains a chromosomal integration of the Z1 cassette (*22*). The Z1 cassette constitutively

expresses the TetR and LacI proteins. All cell culture was done in optically clear M9ca minimal media (Teknova M8010).

Genes and Plasmids

DNA encoding the qacR genes was constructed using overlap extension PCR. Plasmids used contained chloramphenicol resistance with a p15a origin of replication.

In Vivo Experiments

Cells were grown in at least two consecutive overnight cultures in M9ca minimal media. On the day of the experiment, overnight cultures were diluted 1:100 and grown for 5 hours to ensure that the cells were in log phase. Cells were then diluted 1:100 into fresh media at the specified experimental condition. Cells were grown in these conditions at 37C for 12–15 hours in Axygen 96 well plates while shaking at 1100 rpm. Endpoint fluorescence was measured by transferring the cells to clear bottomed 96-well microplates (PerkinElmer, ViewPlate, 6005182). GFP was read at 488/525 with gain 100.

Analysis of the data was done by taking fluorescence readings for each independent well. Experimental conditions of the qacR proteins were done in triplicate and repeats were averaged. Error bars shown are the error propagated originating from the standard deviation of the mean.

Supporting Information Available

Tables containing the amino acid sequences of all of the mutants tested, the optical density values of the measured cell strains, the designed residue positions, and the geometry values used for the bias in the energy function are available as part of the

supplementary information. Different representations of the *in vivo* data are included in the supplementary information, including the normalized fluorescence and the percent of no repressor maximum for all of the aTc concentrations tested. PDB files of computational models of qacR bound to DNA with different vanillin locations, and of qacR2 and qacR5, with the L54W/F50A mutations are also available.

Acknowledgement

The authors thank Jongmin Kim and Jackson Cahn for reading the manuscript. This research was conducted with support from the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. Additional support was granted in part by the Benjamin M. Rosen Bioengineering Center, the Gordon and Betty Moore Foundation through Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative, and DARPA through the Living Foundries Program.

References

1) Desai, T., Rodionov, D., Gelfand, M., Alm, E., and Rao, C. (2009) "Engineering transcription factors with novel DNA-binding specificity using comparative genomics" *Nucleic Acids Res. 37*, 2493–2503.

2) Galvao, T., and de Lorenzo, V. (2006) "Transcriptional regulators a` la carte: engineering new effector specificities in bacterial regulatory proteins" *Curr. Opin. Biotechnol. 17*, 34 – 42.

3) Jha, R. K., Chakraborti, S., Kern, T. L., Fox, D. T., and Strauss, C. E. M. (2015) "Rosetta comparative modeling for library design: Engineering alternative inducer specificity in a transcription factor" *Proteins: Struct., Funct., Bioinf. 83*, 1327–1340.

4) Ramos, J., Martinez-Bueno, M., Molina-Henares, A., Ter´an, W., Watanabe, K., Zhang, X., Gallegos, M., Brennan, R., and Tobes, R. (2005) "The TetR family of transcriptional repressors" *Microbiol. Mol. Biol. Rev. 69*, 326–356.

5) Elowitz, M., and Leibler, S. (2000) "A synthetic oscillatory network of transcriptional regulators" *Nature 403*, 335–338.

6) Gardner, T., Cantor, C., and Collins, J. (2000) "Construction of a genetic toggle switch in *Escherichia coli*" *Nature 403*, 339–342.

7) Grkovic, S., Hardie, K., Brown, M., and Skurray, R. (2003) "Interactions of the QacR multidrug-binding protein with structurally diverse ligands: implications for the evolution of the binding pocket" *Biochemistry 42*, 15226–15236.

8) Schumacher, M., Miller, M., Grkovic, S., Brown, M., Skurray, R., and Brennan, R. (2001) "Structural mechanisms of QacR induction and multidrug recognition" *Science 294*, 2158–2163.

9) Schumacher, M., and Brennan, R. (2003) "Deciphering the molecular basis of multidrug recognition: Crystal structures of the *Staphylococcus aureus* multidrug binding transcription regulator QacR" *Res. Microbiol. 154*, 69–77.

10) Peters, K., Brooks, B., Schumacher, M., Skurray, R., Brennan, R., and Brown, M. (2011) "A single acidic residue can guide binding site selection but does not govern QacR cationic-drug affinity" *PLoS One 6*, e15974.

11) Klinke, H., Thomsen, AB, and Ahring, B. (2004) "Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass" *Appl. Microbiol. Biotechnol. 66*, 10–26.

12) Lassila, J., Privett, H., Allen, B., and Mayo, S. (2006) "Combinatorial methods for small-molecule placement in computational enzyme design" *Proc. Natl. Acad. Sci. U. S. A. 103*, 16710–16715.

13) Allen, B., and Mayo, S. (2010) "An efficient algorithm for multistate protein design based on FASTER" *J. Comput. Chem. 31*, 904–916.

14) Sun, Z., Hayes, C., Shin, J., Caschera, F., Murray, R., and Noireaux, V. (2013) "Protocols for implementing an *Escherichia coli*-based TX-TL cell-free expression system for synthetic biology" *J. Visualized Exp.* E50762.

15) Sun, Z., Yeung, E., Hayes, C., Noireaux, V., and Murray, R. (2014) "Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an *Escherichia coli* Based TX-TL Cell-Free System" *ACS Synth. Biol. 3*, 387–397.

16) Takahashi, M. K., Hayes, C. A., Chappell, J., Sun, Z. Z., Murray, R. M., Noireaux, V., and Lucks, J. B. (2015) "Characterizing and prototyping genetic networks with cell-free transcription–translation reactions" *Methods* Epub May 26, 2015 DOI:10.1016/j.ymeth.2015.05.020.

17) Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014) "Gene Circuit Performance Characterization and Resource Usage in a Cell-Free "Breadboard" " *ACS Synth. Biol. 3*, 416–425.

18) Stanton, B., Siciliano, V., Ghodasara, A., Wroblewska, L., Clancy, K., Trefzer, A., Chesnut, J., Weiss, R., and Voigt, C. (2014) "Systematic Transfer of Prokaryotic Sensors and Circuits to Mammalian Cells" *ACS Synth. Biol. 3*, 880–891, PMID: 25360681.

19) Bochevarov, A., Harder, E., Hughes, T., Greenwood, J., Braden, D., Philipp, D., Rinaldo, D., Halls, M., Zhang, J., and Friesner, R. (2013) "Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences" *Int. J. Quantum Chem. 113*, 2110–2142.

20) Kuhlman, B., Dantas, G., Ireton, G., Varani, G., Stoddard, B., and Baker, D. (2003) "Design of a novel globular protein fold with atomic-level accuracy" *Science 302*, 1364– 1368.

21) Allen, B., and Mayo, S. (2006) "Dramatic performance enhancements for the FASTER optimization algorithm" *J. Comput. Chem. 27*, 1071–1075.

22) Lutz, R., and Bujard, H. (1997) "Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements" *Nucleic Acids Res. 25*, 1203–12101210.

Chapter 4

DEVELOPMENT OF *IN VITRO* TRANSCRIPTION-TRANSLATION SYSTEMS WITH THREE WILD-TYPE *PSEUDOMONAS* SPECIES

Introduction

In Chapter 2 and Chapter 3, we described steps towards engineering a synthetic stress response to a growth inhibiting phenolic aldehyde, vanillin. However, overexpressing our identified efflux pump did not improve the engineered organism's tolerance to vanillin. As an alternative approach, in this chapter we identify non-model organisms with existing tolerance to these growth inhibitors, and develop *in vitro* transcription-translation (TX-TL) systems in these new species.

We tested the tolerance to growth inhibition with three related *Pseudomonas* bacteria isolated from soil samples. Soil contains decaying plant matter, and would have sustained an evolutionary pressure on these bacteria to build tolerance to the lignocellulosic growth inhibitors we have been working with. *Escherichia coli*, a bacterium from the gut, may have significantly less tolerance for these specific stresses.

The species *Pseudomonas synxantha* 2-79 (PSX), *Pseudomonas chlororaphis* PCL1391 (PCL), and *Pseudomonas aureofaciens* 30-84 (PAU) are culturable wild-type bacteria, kindly provided by Elin Larsson and Dianne Newman at Caltech. They have been studied separately and together as symbionts of plant roots, where they produce anti-fungal phenazines (1-3). The phenazines and these bacteria are known to control take-all, a disease in wheat plant roots, and damping-off disease, a multi-pathogen constellation of seed and seedling injuries across different plant species.

Relatives of our target species, *Pseudomonas fluorescens* and *Pseudomonas putida*, have been used for *in vitro* transcription-translation (TX-TL) before (4, 5). We reasoned that if our target *Pseudomonas* spp. were tolerant of the specific growth inhibitors we worked with, we could change the problem from making a synthetic stress response from scratch into one where we start with a working stress response already optimized by evolution in these bacteria (6). Then our challenge becomes one where we need to make engineering tractable in these bacteria, one that we may have more success with. Demonstrating transformability, genomic integration, and *in vitro* transcription-translation with these species was our goal for enabling engineering in these non-model chassis.

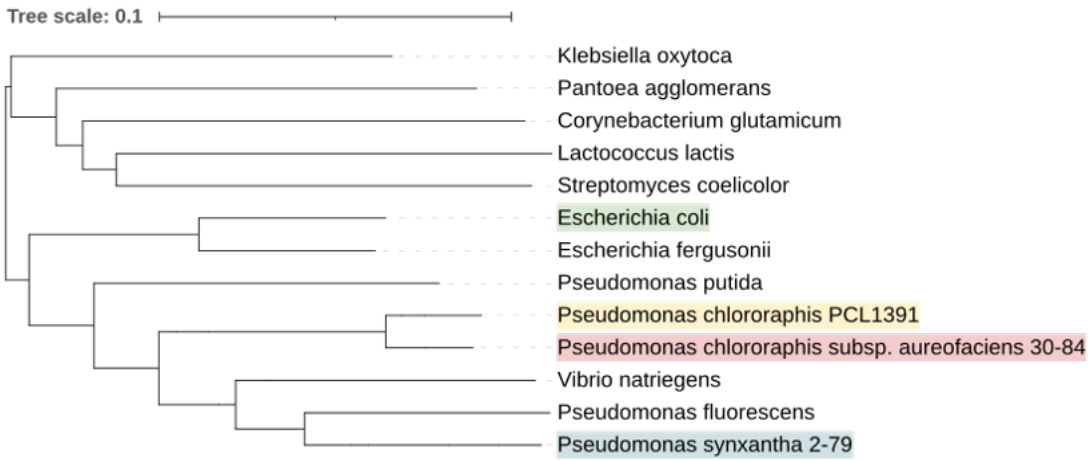


Figure 4.1 Phylogenetic tree of bacteria used for *in vitro* transcription-translation (TX-TL), including the three in this study (highlighted).

P. aureofaciens 30-84 is also called *P. chlororaphis subvar. aureofaciens* 30-84, and is closely related to *P. chlororaphis* PCL1391. In Figure 4.1 and Table 4.1 we show the phylogenetic relationships between the three target bacteria and other bacteria used for TX-TL (7). *P. synxantha* 2-79 is more distantly related, with a similar distance between

its genome, *E. coli*, PCL, and PAU as measured by digital DNA-DNA hybridization (dDDH d4) (8).

dDDH d4 in % [C.I.] (G+C difference)	*Pseudomonas synxantha* 2-79	*Pseudomonas chlororaphis* PCL1391	*Pseudomonas aureofaciens* 30-84	*Escherichia coli*
Pseudomonas synxantha 2-79		25.4 [23.0 - 27.8] (2.9%)	25.6 [23.3 - 28.1] (3.17%)	22.8 [20.5 - 25.2] (9.14%)
Pseudomonas chlororaphis PCL1391	25.4 [23.0 - 27.8] (2.9%)		58.3 [55.5 - 61.0] (0.27%)	21.7 [19.5 - 24.1] (12.18%)
Pseudomonas aureofaciens 30-84	25.6 [23.3 - 28.1] (3.17%)	58.3 [55.5 - 61.0] (0.27%)		21.8 [19.6 - 24.3] (12.31%)
Escherichia coli	22.8 [20.5 - 25.2] (9.14%)	21.7 [19.5 - 24.1] (12.18%)	21.8 [19.6 - 24.3] (12.31%)	

Table 4.1: Genome-to-genome distances as measured by digital DNA-DNA hybridization (dDDH), confidence intervals in [low - high] brackets and (G+C content difference) in parenthesis. The d_4 measure is a genome-length independent metric described in (30).

We conducted an early experiment and found PSX, PCL, and PAU had similar or improved tolerance to vanillin, 4-hydroxybenzaldehyde, and syringaldehyde compared to *E. coli* BL21(DE3). In Figure 4.2 and Supplement Section B, we show the cell density curves of each of the four species with growth inhibitor and without growth inhibitor as measured over 12 hours via optical density (600 nm).

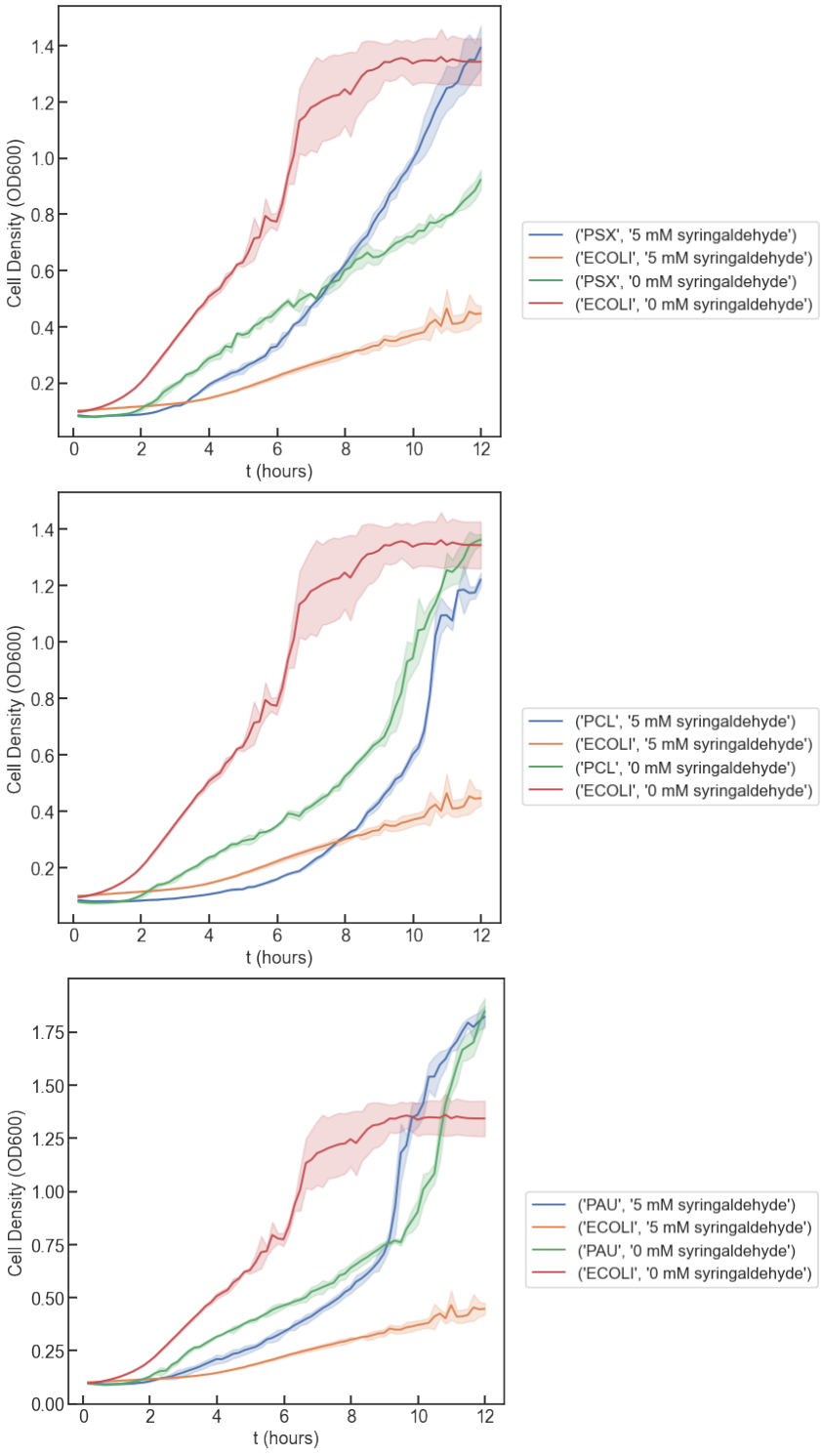


Figure 4.2: *P. synxantha* and *P. aureofaciens* show improved tolerance to syringaldehyde during growth compared to *E. coli*. See Supplementary Section B, this chapter, for all four species' responses to vanillin and 4-hydroxybenzaldehyde.

In each case some or all of the *Pseudomonas* spp. show less growth inhibition than *E. coli*, indicating a greater tolerance for the small molecule stresses tested.

Having demonstrated these bacteria have an existing tolerance to the growth inhibitors in question, we turn to the task of making these bacteria usable for engineering purposes. In this chapter we show the successful, repeatable growth of these three bacteria at high density and at moderate (>1 liter) scale. After harvesting several cell pellets, we applied the protocol used with *E. coli* to produce a clarified *Pseudomonas* lysate for *in vitro* transcription-translation (TX-TL). Varying several process parameters, we demonstrate initial working TX-TL systems in all three bacteria, and characterize the extracts of *P. synxantha* in greater detail with larger follow-up experiments showing potassium and magnesium dependence of the reaction and variations in yield due to changes in the DNA template concentrations. This is the first demonstration of working *in vitro* TX-TL in these species.

Elin Larsson created multiple strains of *P. synxantha* with genomically integrated inserts (9). The inserts contained constitutive promoters from a panel from the Registry of Standard Biological Parts, DNA parts that have been used in other synthetic biology projects (10) of a variety of strengths producing mNeonGreen, an especially bright green fluorescent protein (11). These strains were grown in different media. Fluorescence and cell density were monitored over time. The same inserts were produced as DNA template for *in vitro* TX-TL testing, and we show that the strong, medium, and weak promoter parts have similar strengths in vivo and *in vitro*.

Materials And Methods

Construction of DNA templates

DNA parts including promoters, ribosome binding sites, reporters, terminators, and backbones with antibiotic resistance were amplified by PCR in NEB Q5 2x Master Mix, purified by gel extraction and QIAGEN MinElute spin columns, and assembled using NEB Hifi 2x Master Mix. Figure 4.3 shows the plasmid map for a strong constitutive promoter expressing mNeonGreen, and Figure 4.4 shows the plasmid map for a promoter with a different strength also expressing mNeonGreen.

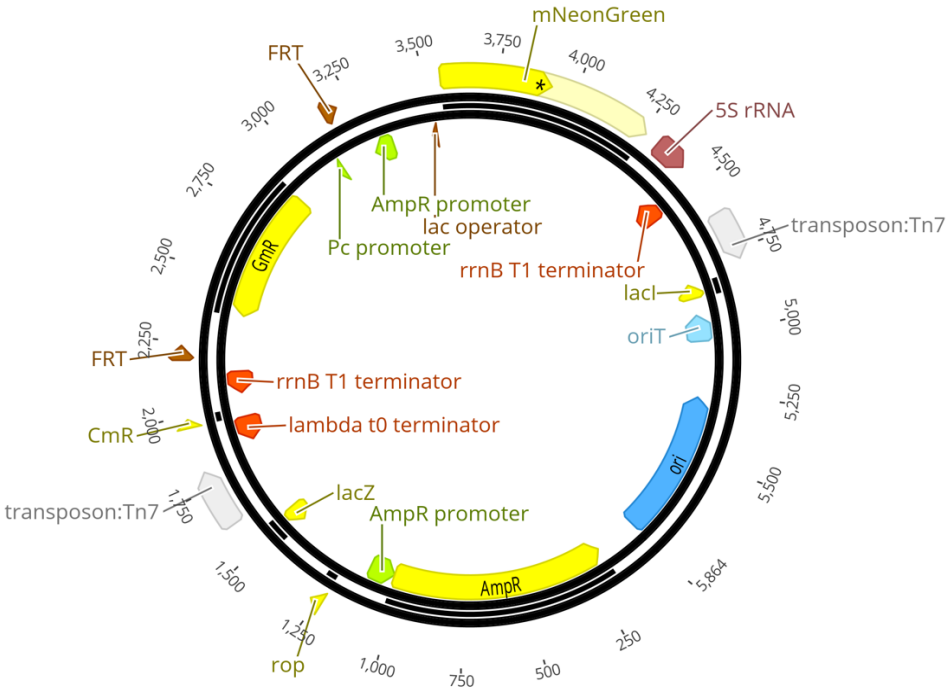


Figure 4.3: Plasmid map of Pa10403, a positive control template with a strong promoter and ribosome binding site producing the mNeonGreen fluorescent reporter.

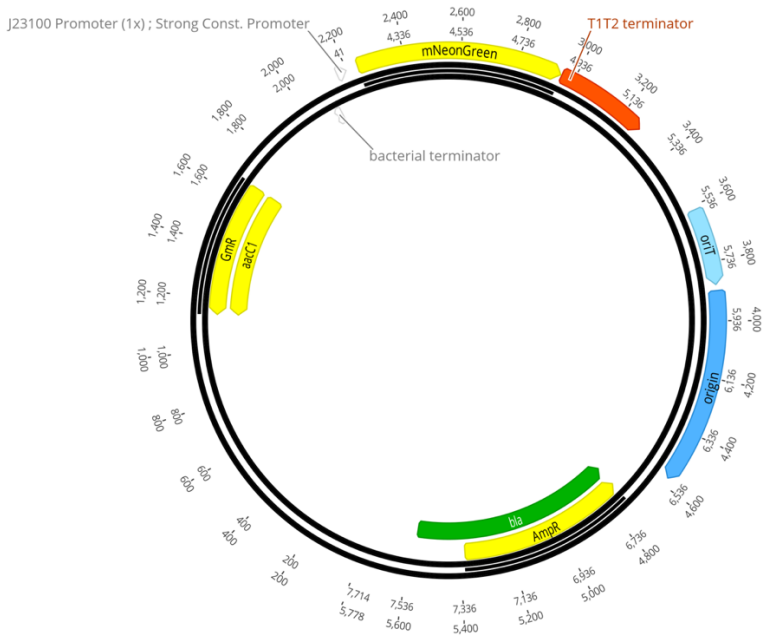


Figure 4.4: Plasmid map of P6d, one of several promoters from the Anderson constitutive promoter collection (http://parts.igem.org/Promoters/Catalog/Anderson), cloned with the same mNeonGreen reporter.

Bacterial Strains and Growth Conditions

Pseudomonas synxantha 2-79, *Pseudomonas chlororaphis* PCL1391, and *Pseudomonas aureofaciens* 30-84 were obtained from the Newman lab at Caltech. Cells were grown by streaking onto 2xYTPG (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 40 mM potassium phosphate dibasic, 22 mM potassium phosphate monobasic, 2% glucose) agar or LB (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) agar without antibiotics and incubating for 18-36 hours at 30C. Individual colonies were picked and grown in 5 mL of 2xYTPG with shaking at 200-220 rpm at 30C overnight. As needed, 50 mL of 2xYTPG in a 250 mL baffled flask or 660 mL of 2xYTPG in a 2.8L baffled flask with an adhesive AirOTop sterile seal would be inoculated 1:1000 the next day. Cell growth was monitored by optical density at 500 nm or 600 nm, with 1:1, 1:4, 1:16, 1:64, and 1:256

dilutions of culture samples prepared in fresh 2xYTPG media to stay in the Biotek H1MF plate reader linear range for absorption measurements (OD 0.005 to 0.5).

Measurement Of Protein Concentration

For each extract produced for the sonication, runoff, and salt panels, protein concentration was measured using a Bradford assay. In brief, a 1:50 dilution of lysate was prepared in Tris-buffered saline. A bovine serum albumin standard was used to prepare 8 different concentrations for a standard curve. Standard curves were prepared in triplicate on every 96-well plate used for protein measurement, and samples were also prepared in triplicate.

Production of Pseudomonas extracts

The protocol from Sun et al (12) and the protocol from Kwon et al (13) for *in vitro* transcription-translation with *E. coli* was repurposed for the *Pseudomonas* TX-TL protocol with minor modifications. Briefly, in Figure 4.4 step 1, cells are grown on 2xYTPG agar. After 18-36 hours of growth at 30C, the plates are stored at 4C for up to a month. In step 2, individual colonies are picked and used to grow 5 mL overnight cultures in 14 mL tubes. In step 3, 2.8L baffled shake flasks with 660 mL of 2xYTPG are inoculated 1:1000 the next day and grown for 8-12 hours. Note that unlike past protocols, the final culture is inoculated directly from a 5 mL overnight, without an intermediate 50 mL culture.

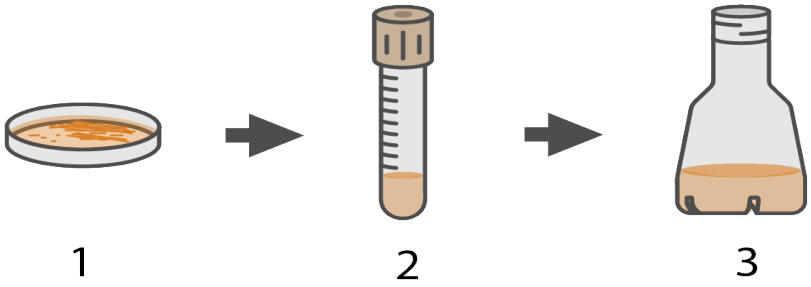


Figure 4.4: Schematic of growth procedure.

In Figure 4.5 step 1, cells are grown to OD 1.5-3.0, with monitoring of optical density performed every 60 minutes until the last doubling of the growth, then monitoring increases to once per 20 minutes. Dilutions of 1:4, 1:16, 1:64, or 1:256 in fresh 2xYTPG media are used to produce a sample with an optical density at 600 nm under 0.5. In Figure 4.5 step 2 200 mL of Buffer A (1.8 g/L Tris-acetate, 3 g/L Mg-acetate, and 12.2 g/L K-glutamate, adjusted to pH 8.2 with 2M Tris and autoclaved, then 1 mL / L of 1M DTT is added) is frozen in 1L centrifuge tubes stored at -20C at an angle for 1-2 hours.

At harvest time, the cell culture is decanted into the 1L centrifuge tube over the ice. This brings the culture temperature down from 30C to 10C within a few minutes. In Figure 4.5 Step 3, the cell culture is centrifuged at 4800g for 12 minutes at 4C. The supernatant is decanted, and the tubes are placed on ice. The cell pellet is resuspended in Buffer A, then otherwise follows the Sun et al. *E. coli* protocol with two wash steps in 1L bottles, a final wash step and transfer into two weighed 50 mL tubes, flash frozen in liquid nitrogen, and stored at -80C.

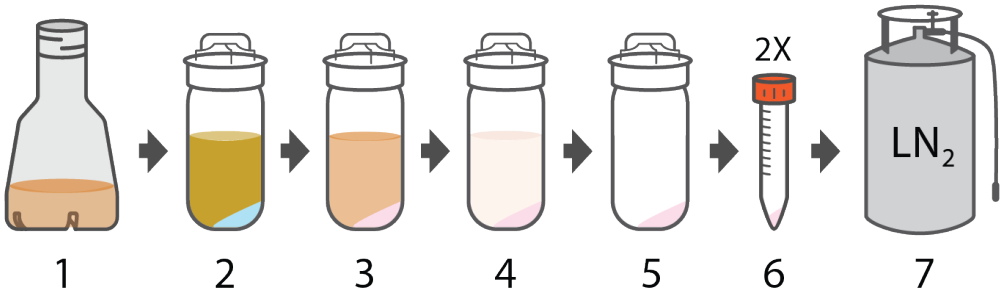


Figure 4.5: Schematic of cell harvesting, pellet washing, and freezing procedure.

At a later time, the cell pellet is removed from the freezer and placed on wet ice to thaw. In Figure 4.6 step 1, 1 mL / g of Buffer A is used to resuspend the pellet, then 4 mL of resuspended cell pellet is added to each 14 mL falcon tube. In Figure 4.6 step 2, each tube is sonicated for 120s, 5s on, 10s off, with an amplitude of 10, 25, or 50 on a Qsonica Q700. In Figure 4.6 step 3, the tubes are centrifuged at 12,000g for 10 minutes at 4C. In Figure 4.6 step 4, the supernatant is transferred to new 1.4 mL or 14 mL tubes and incubated in a "runoff" step with open lids shaking at 220 rpm at 30C for 0-60 minutes. These tubes are centrifuged again at 12,000g for 10 minutes at 4C, then the supernatant is transferred to new tubes for freezing in LN2 and storage at -80C.

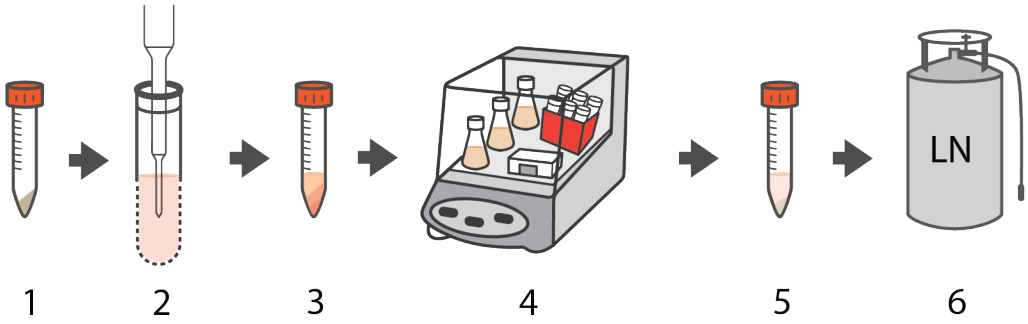


Figure 4.6: Schematic of cell pellet thawing, lysis, runoff, and extract freezing.

Reaction Conditions for In-Vitro Transcription-Translation (TX-TL)

Reaction conditions also follow the *E. coli* protocol in Sun et al (34). In brief, each 10 microliter reaction in a 384 well plate contains 29% buffer, 13% potassium and magnesium glutamate salts, 25% DNA template, and 33% processed lysate. The buffer is comprised of 26% energy solution, 57% amino acid mix, and 17% 40% PEG-8000. These are mixed on ice and hand-pipetted carefully to avoid introducing bubbles. Individual reactions are prepared in glass-bottomed 384 well plates sealed with an oxygen-permeable Breathe-Easy seal. The plate is centrifuged for 2 minutes to mix the reagents and reduce bubbles in the wells, then read at 30C in the plate reader.

Preparation of DNA Template for In-Vitro Transcription-Translation (TX-TL)

E. coli JM109 or DH10B cells with DNA constructs were grown to 100 mL (midiprep) or 400 mL (maxiprep) scale. Pellets were processed using Macherey-Nagel midiprep or maxiprep kits to produce 100s to 1000s of micrograms of purified DNA, including a "finalizer" step to re-purify and concentrate the eluate from the kit. DNA is diluted to produce a concentration of 10 nM in the final reaction unless otherwise specified.

Measurement Of Fluorescence

Biotek H1M and H1MF plate readers were used to measure fluorescence of mNeonGreen (excitation 490 nm, emission 520 nm) and optical density (500 nm and 600 nm). mNeonGreen protein was purified using a his-tag and NiNTA columns, and UV absorption was used to quantify the amount of purified protein. Triplicate dilutions at 8 concentrations from 20 nM to 0 nM were measured five times in all plate readers, and separate standard curves were created for each plate reader. Reactions were incubated at 30C.

Construction Of Constitutively Fluorescent P. synxantha

All cloning to produce *P. synxantha* genomic integration constructs was done using *E. coli* DH10B with the backbone pJM220 (https://www.ncbi.nlm.nih.gov/nuccore/KX777256). Constitutive promoters from *E. coli* (http://parts.igem.org/Promoters/Catalog/Anderson) and *P. putida* (https://doi.org/10.1021/acssynbio.5b00058) were fused to the fluorescent protein mNeongreen. The constructs were then integrated on the *P. synxantha* chromosome using transposase based insertion at the Tn7 site. The protocol used for making and transforming competent cells was modified from Choi et al (10.1038/nprot.2006.24).

Briefly, electrocompetent *P. synxantha* cells were electroporated with the construct plasmid as well as a plasmid containing the transposase and genes required for genome insertion. The cells were then recovered in rich medium (SOC) for 3 hours at 30C and plated onto LB gentamicin (20 ug/ml) plates for 24 hours before picking colonies for sequence verification.

Plate Reader Assay For In Vivo Fluorescence

In vivo fluorescence was measured using a Biotek plate reader (excitation/emission 490/520 nm), continuous orbital shaking, 30C for 24 hours. The inoculum for the final culture came from an overnight culture diluted to OD ~0.1.

Results And Discussion

Pseudomonas synxantha 2-79

The growth rate of *P. synxantha* ("PSX") was measured in 2xYTPG in triplicate at 50 mL scale. Based on the growth curve in Figure 4.7, the cells were grown to OD 3.0 +/- 10% before harvest for extract production.

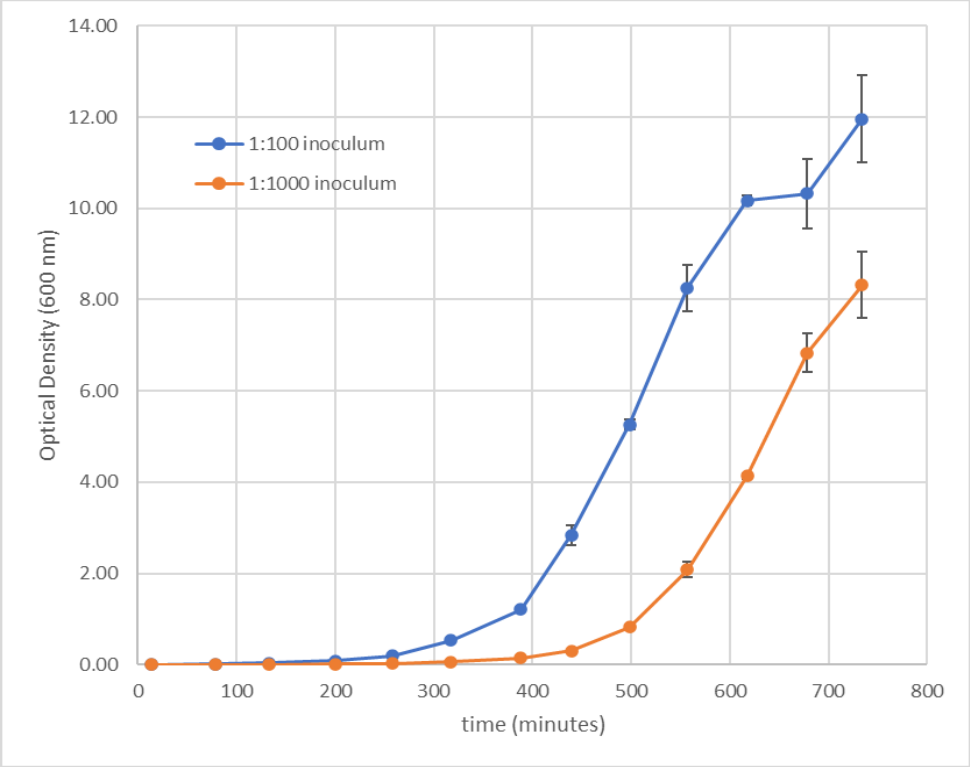


Figure 4.7: Growth of *P. synxantha 2-79* at 50 mL scale in 2xYTPG over 12 hours. Error bars are standard deviation across cultures grown in triplicate.

Extract Production

At harvest, the cells sediment into a pale pink pellet. The three wash steps take 2-2.5 hours, and the cell pellet loosens slightly by the last wash step, requiring the supernatant to be removed by pipette instead of decanting.

The first extracts tested were prepared as biological triplicates, and each replicate was comprised of extract made with three different sonication amplitudes and three runoff times. Figure 4.8 shows the protein concentration for the nine different preparation conditions and three *E. coli* extracts used as positive controls. The *E. coli* positive control extract was prepared at sonication amplitude 50 (120s, 5s on 10s off) and incubated for 60 minutes at 37C for the runoff reaction.

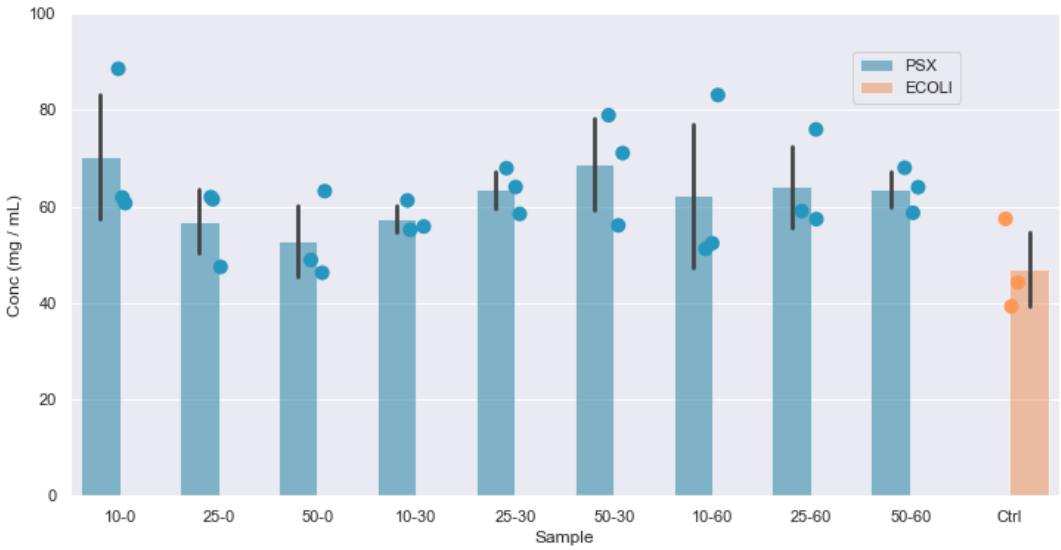


Figure 4.8: Protein concentrations as determined by Bradford assay. The sample labels at the bottom indicate "sonication amplitude" and "runoff time" (minutes). Bar heights are averages, error bars are standard deviation, and the dots are the individual measurements. These data were collected by Mark Prator.

Test reactions were prepared with the Pa10403 plasmid at 10 nM with 6 different salt conditions. The reactions were incubated at 30C and measured over 12 hours, shown in Figure 4.10. The salt conditions are potassium glutamate and magnesium glutamate in the following concentrations - 0 mM K-glutamate/0 mM Mg-glutamate, 60 mM K-glutamate/6 mM Mg-glutamate, 120 mM K-glutamate/12 mM Mg-glutamate, 180 mM K-glutamate/18 mM Mg-glutamate, 240 mM K-glutamate/24 mM Mg-glutamate, 300 mM K-glutamate/30 mM Mg-glutamate. An *E. coli* positive control with the same salt conditions is shown in Figure 4.9.

Negative controls with no DNA template were prepared at two salt concentrations, 60 mM K-glutamate/6 mM Mg-glutamate and 180 mM K-glutamate/18 mM Mg-glutamate, shown in Figure 4.11. The negative control wells show the extract fluorescence is low without a DNA template and does not change significantly over the course of the experiment.

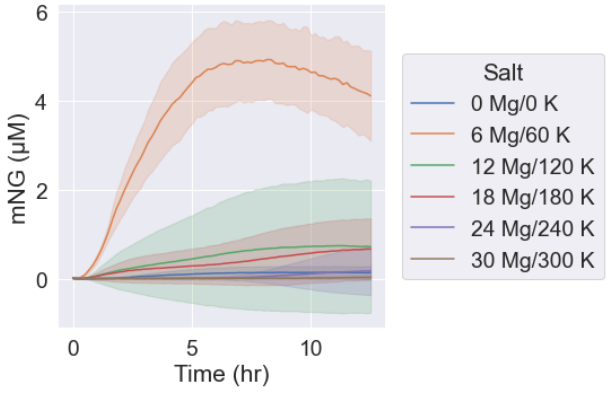


Figure 4.9: *E. coli* positive controls with Pa10403 10 nM template. Lines are means across replicates, and the bands show standard deviation.

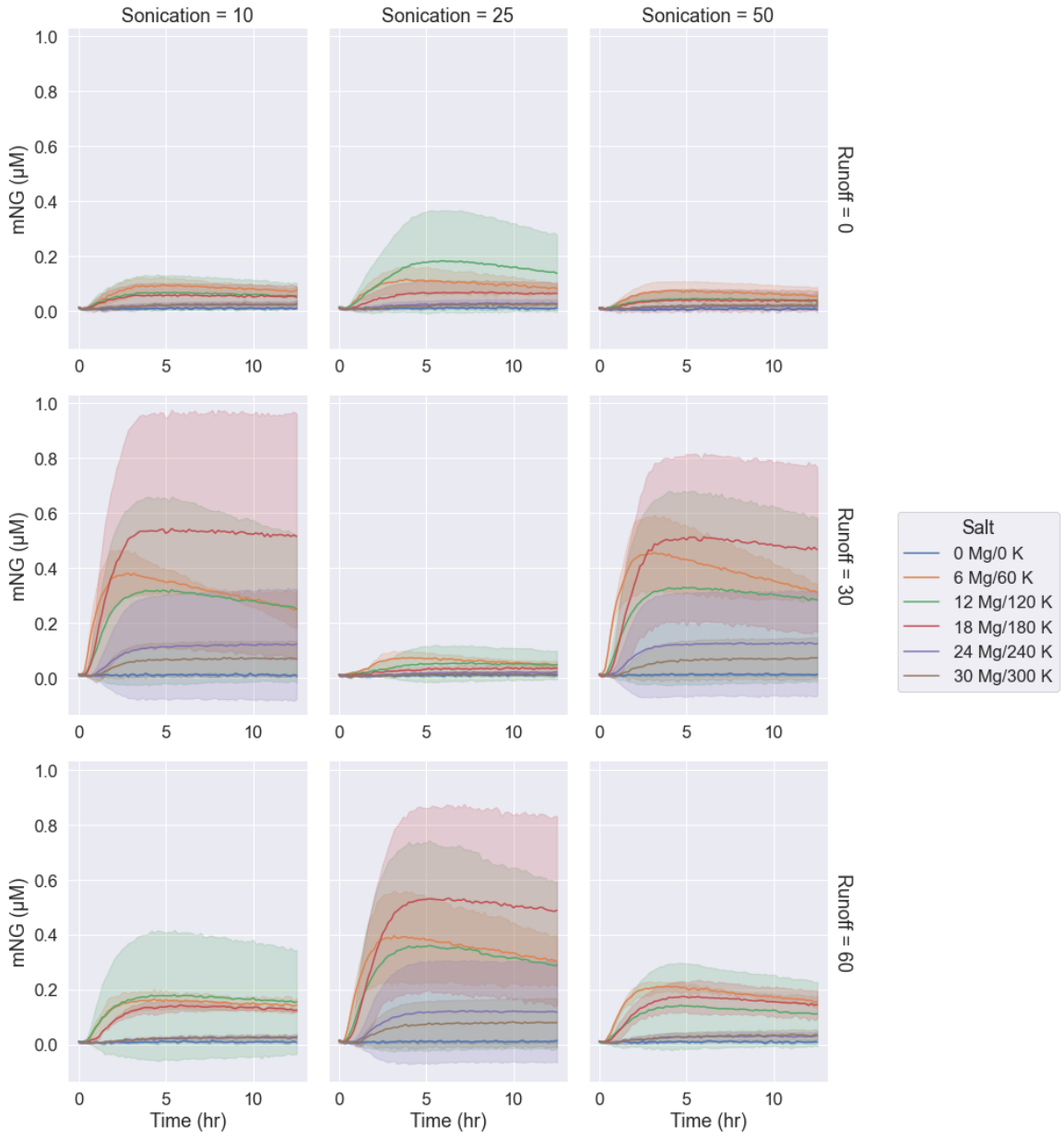


Figure 4.10: PSX extract, produced across varied sonication conditions (amplitude, 5s on 10s off, 120s total) and runoff times (minutes). Six magnesium-potassium salt concentrations were used for this test. Bands show standard deviation, and lines show means across replicates.

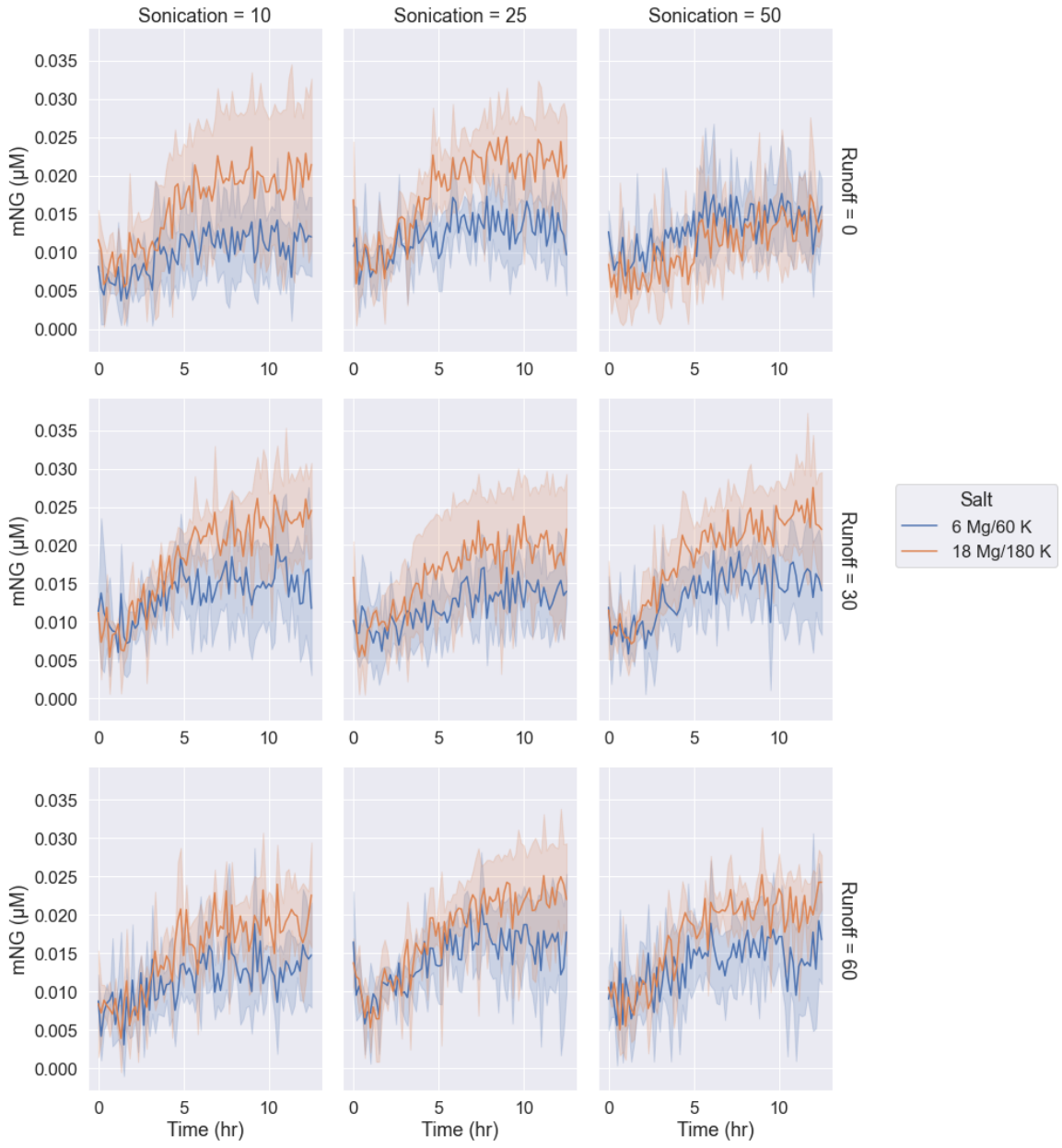


Figure 4.11: Negative controls with no DNA template added to *P. synxantha* TX-TL reactions at two different salt concentrations, showing relatively low fluorescence and small changes in fluorescence over time in the mNeonGreen band (Ex: 490 nm, Em: 520 nm).

Expanded Salt Panel

Next, three new batches of *P. synxantha* lysate were produced at a single sonication amplitude (‘50’) and a single runoff time (60 minutes). These extracts were used to test additional potassium and magnesium salt concentrations, shown below in Figure 4.12. Almost of the test conditions show a clear fluorescent signal accumulating over the first hours of the reaction for all three of the new batches. The yield and optimum salt concentration varies across the different batches in a way similar to past *E. coli* TX-TL reactions. Figure 4.12 shows batch one having the highest yield of the three at two salt concentrations, batch two having several salt concentrations where it has the highest yield, and batch three also having several different salt concentrations where it has the highest yield. This shows the extract making process was repeatable, and that there are some differences in performance between the batches despite being made all with the same conditions.

Varied Template Concentration

Next, using these three new batches of extract, we tested a varying amount of DNA template added to the reaction. Figure 4.13 shows each batch of extract produced a similar amount of the fluorescent reporter for a given template concentration, and doubling the amount of template doubles the amount of fluorescent reporter made. The 20 nM Pa10403 condition using the third extract batch reached the highest concentration of mNeonGreen among all of the PSX experiments here at approximately 2.7 μM. Qualitatively all of the extracts stop producing mNeonGreen at 3-4 hours regardless of template concentration.

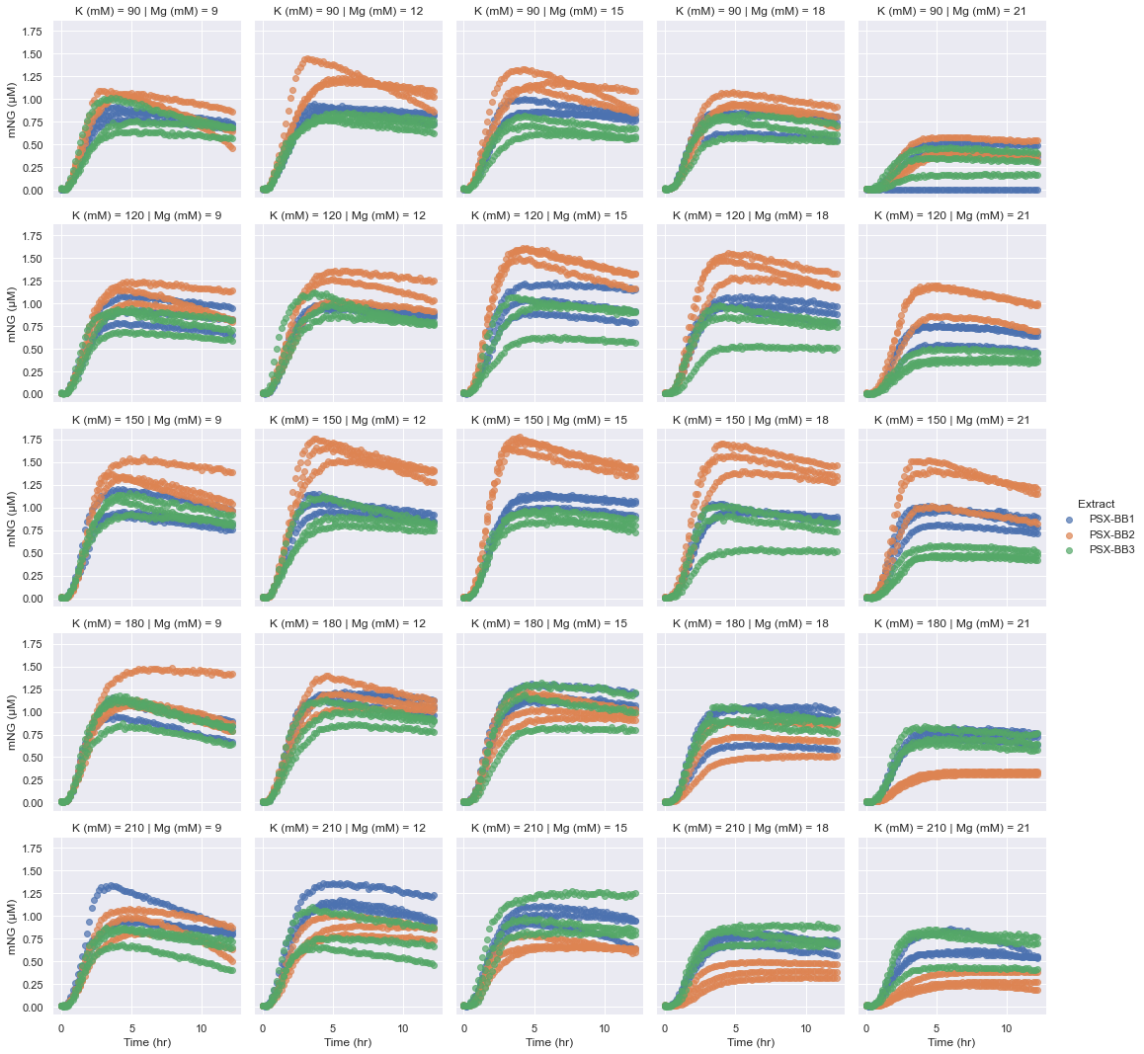
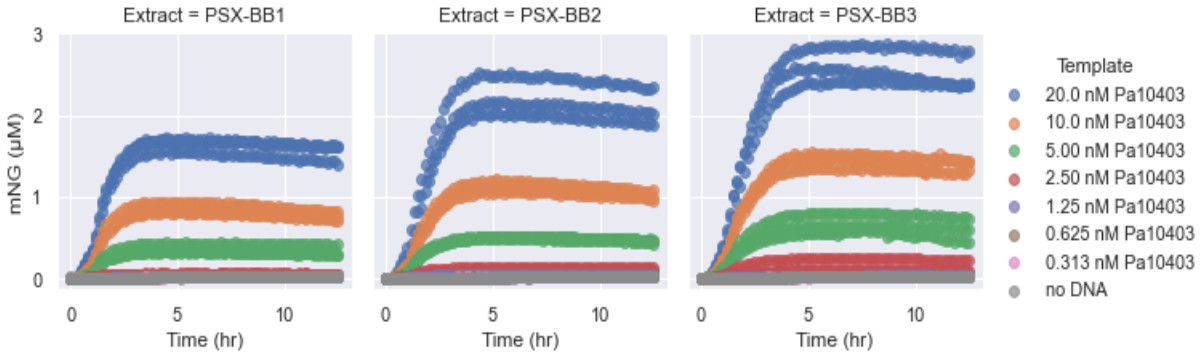


Figure 4.12: Three new batches of PSX extract were produced, and tested across five concentrations of potassium and five concentrations of magnesium salt.



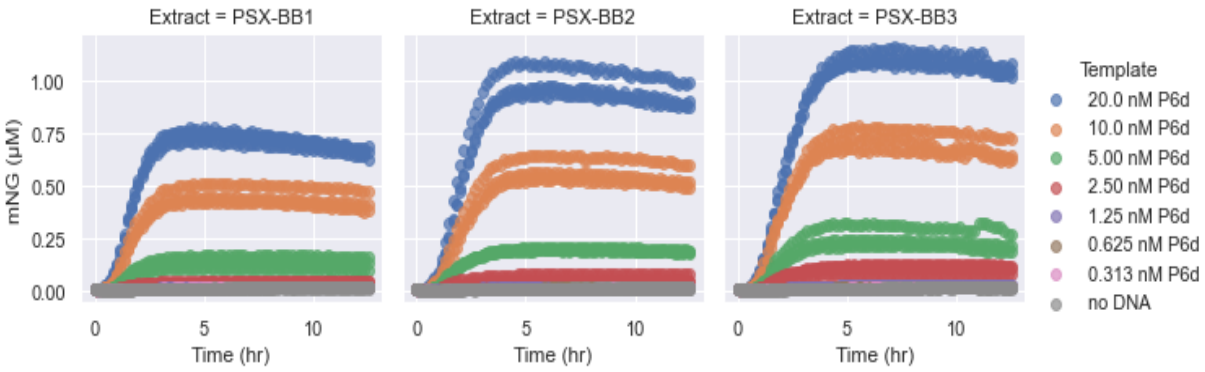


Figure 4.13: Varied concentrations of positive control plasmid Pa10403 and Anderson constitutive promoter template P6d show the reactions are in a linear range where doubling the amount of template produces approximately double the amount of reporter protein.

Comparison to in vivo expression

Strains of *P. synxantha* had mNeonGreen inserts genomically integrated with constitutive promoters from the Anderson panel and another promoter panel from the Newman lab. The fluorescence below is normalized by optical density at 500 nm. These measurements of genomically integrated strains with promoters Pa10403, P1m, P2d, P2m, P3d, P3m, P4d, P4m, P5d, P6d, Syn35, and Syn42 were performed by Elin Larsson.

To compare the results from the *in vitro* TX-TL reactions, we take the peak fluorescence from a time course at approximately 4 hours into the reaction, shown in Figure 4.16. The reactions, regardless of promoter strength, all stop producing protein at approximately the same time.

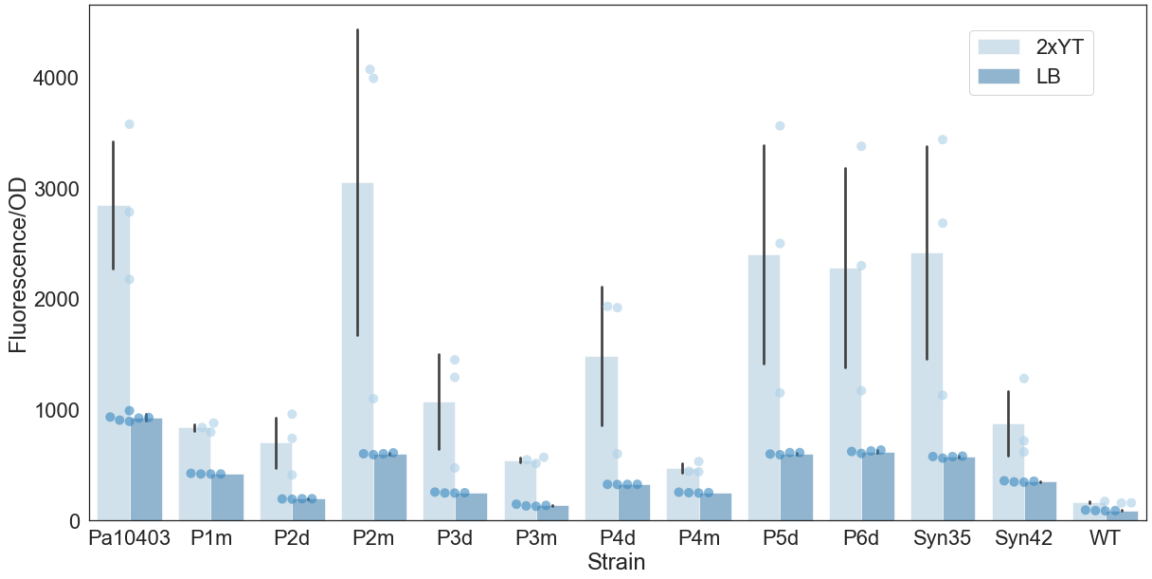


Figure 4.14: Normalized fluorescence (AFU/OD) for multiple constitutive promoters genomically integrated into *P. synxantha.* These data were collected by Elin Larsson from the Newman lab.

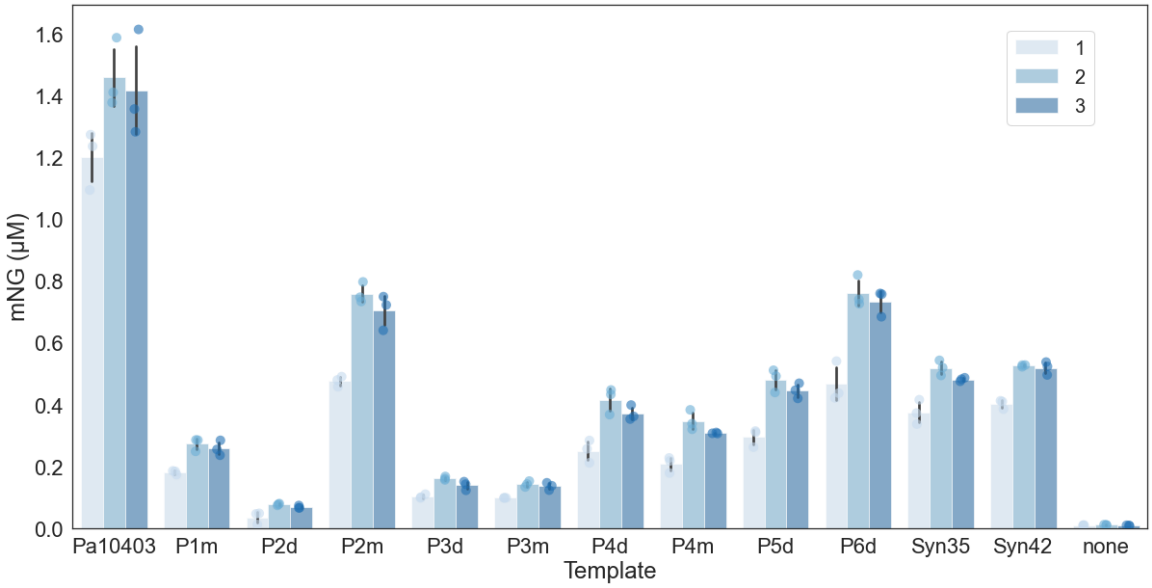


Figure 4.15: Three batches of *P. synxantha* TX-TL *in vitro*, using plasmid templates with the same promoters used *in vivo* . Bar heights show means across technical replicates, error bars show standard deviation across technical replicates, and dots show the individual data points.

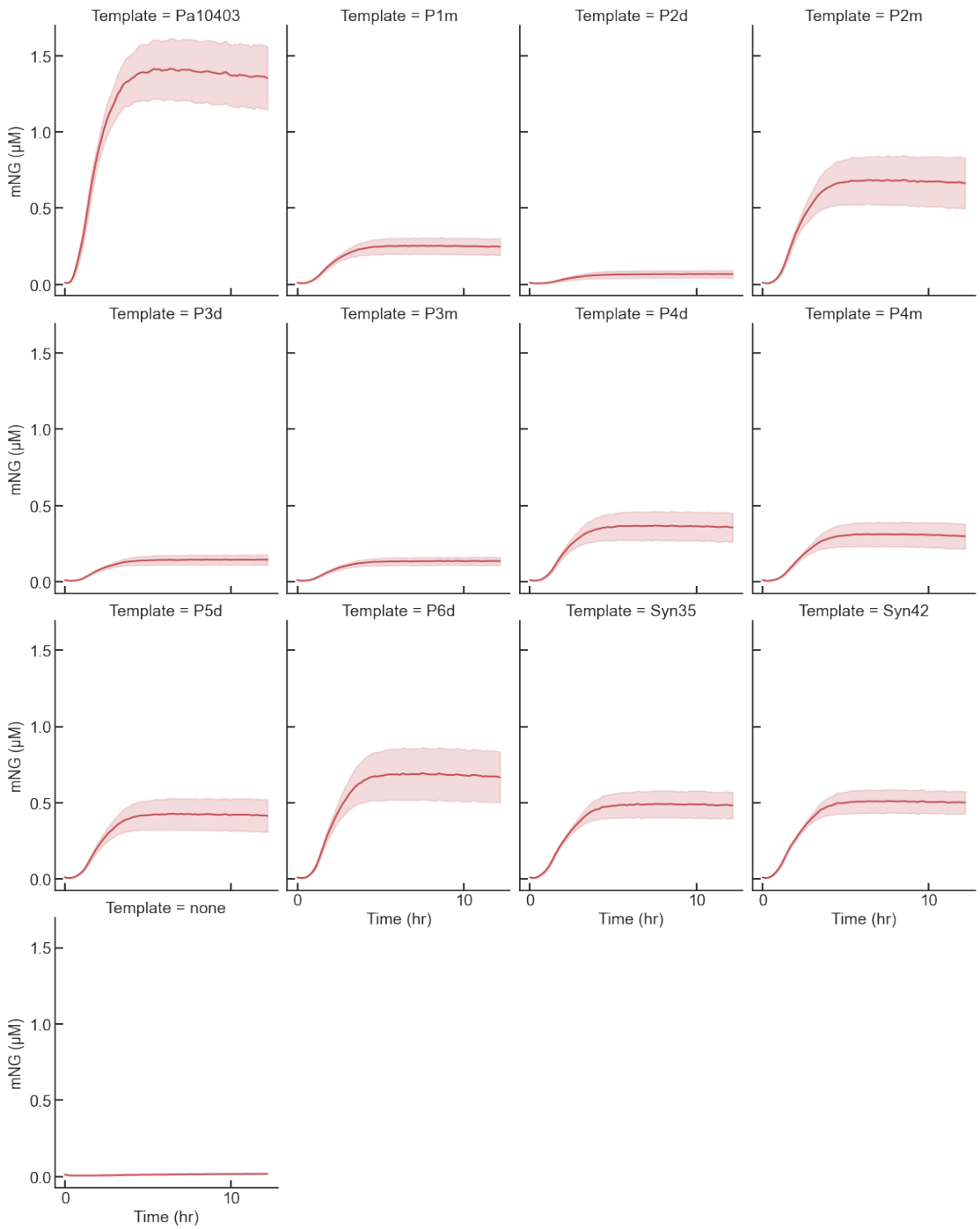


Figure 4.16: Time course data (averages of technical and biological triplicates, error band representing standard deviation) showing the reaction halting at approximately 4 hours, when the "endpoint" data for Figure 4.16 was measured. Bands show standard deviation, solid lines are means of technical replicates.

The variable template concentration measurements showed a similar exhaustion of the reaction at 3-4 hours. With some kinds of resource limitations, we would expect stronger circuits to exhaust the reaction sooner, and weaker circuits to exhaust the reaction later, which is not what we see here. This is an observation that merits later follow up.

These same promoters measured *in vivo* were transformed into plasmids with mNeonGreen and used as templates for *P. synxantha* TX-TL reactions. The fluorescence measurements below in Figure 4.16 show that the strong promoters *in vivo* are also strong *in vitro*. This shows the *Pseudomonas* extract recreates an environment that performs similarly to the whole cell.

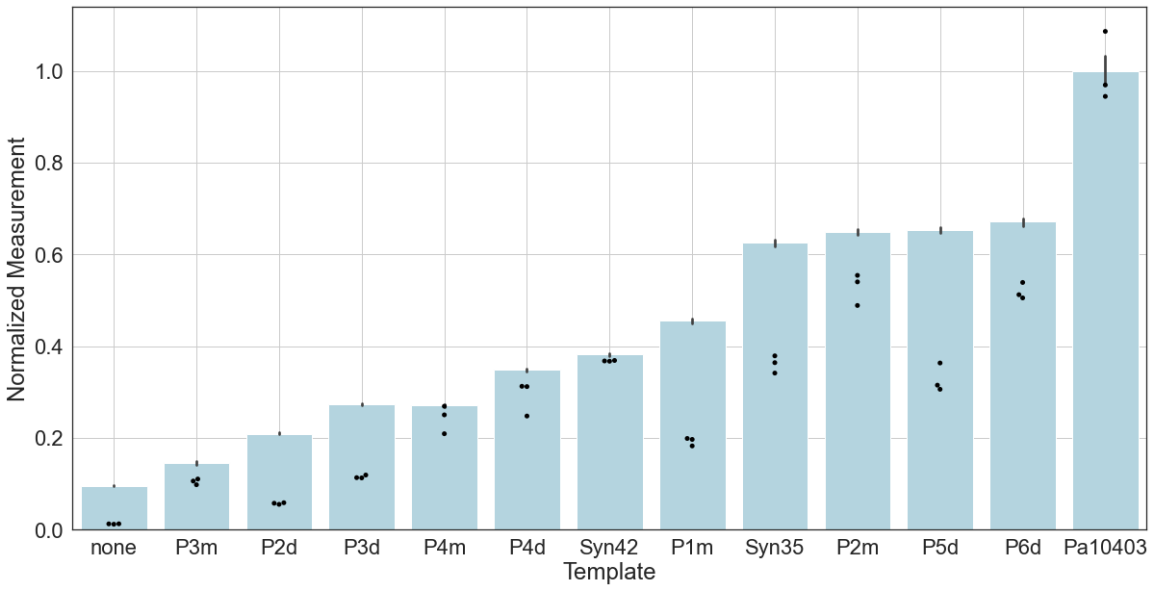


Figure 4.17: *In vivo* and *in vitro* measurements of the same promoter panel. Bar heights show means of *in vivo* fluorescence measurements normalized to the Pa10403 measurement, error bars show standard deviation. *In vitro* measurements of the same promoters are plotted as dots. Chart was prepared by Elin Larsson.

Figure 4.17 shows the templates sorted by ascending fluorescence, and normalized so the Pa10403 signal is 1 both *in vivo* and *in vitro*. The bars are the *in vivo* signal taken in LB media, and the dots are the measurements *in vitro* of the same promoter. This shows more clearly how the *in vitro* fluorescence measurements follow the *in vivo* measurements.

Triplicate PSX TX-TL reactions and *E. coli* TX-TL controls were prepared in PCR tubes and incubated at six different temperatures in six PCR machines for 12 hours. At the endpoint the reactions were transferred into a 384 well plate for an endpoint fluorescence measurement, shown in Figure 4.18 and 4.19.

PSX extract shows much less fluorescence than in previous TX-TL tests in these reactions incubated by PCR machine. These reactions are conducted in an air-impermeable PCR test tube, suggesting the oxygen permeable membrane used for the other reactions in 384 well plates may be essential for PSX extracts. *E. coli* extract does not show this same behavior, seeming to express the reporter protein at good yield. It's not clear from these tests why the PSX extracts would perform differently compared to the *E. coli* extracts. The PSX cultures are grown at 30C and the *E. coli* cultures are grown at 37C. The typical convention is to run the TX-TL reactions at the same temperature as the growth temperature, or at a temperature very close to the growth temperature, which is what has been done here.

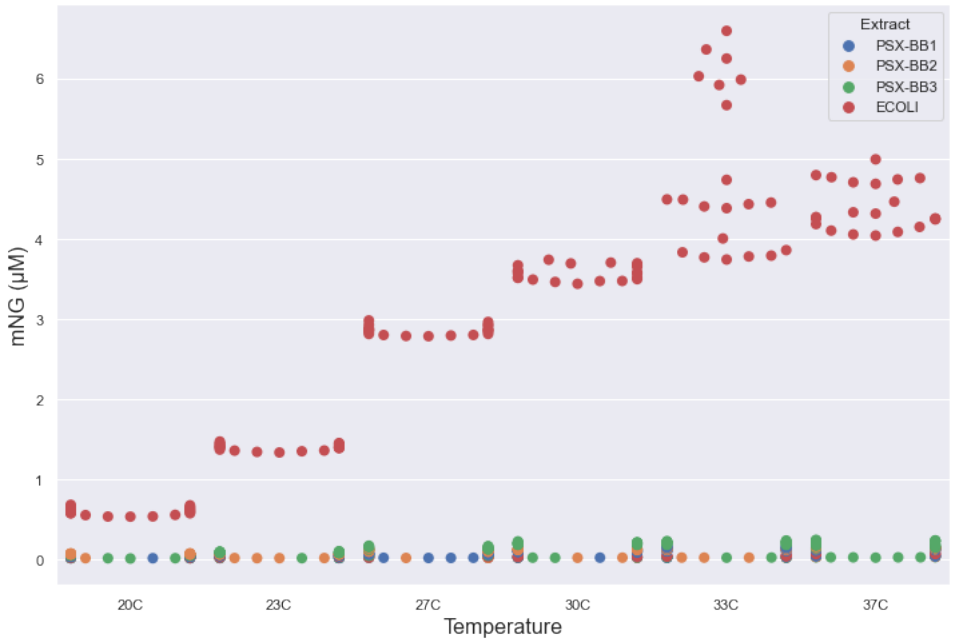


Figure 4.18: Varied temperatures at which the *in vitro* reactions were tested. These tests were performed in a PCR machine in sealed PCR tubes, then endpoint fluorescence was measured.

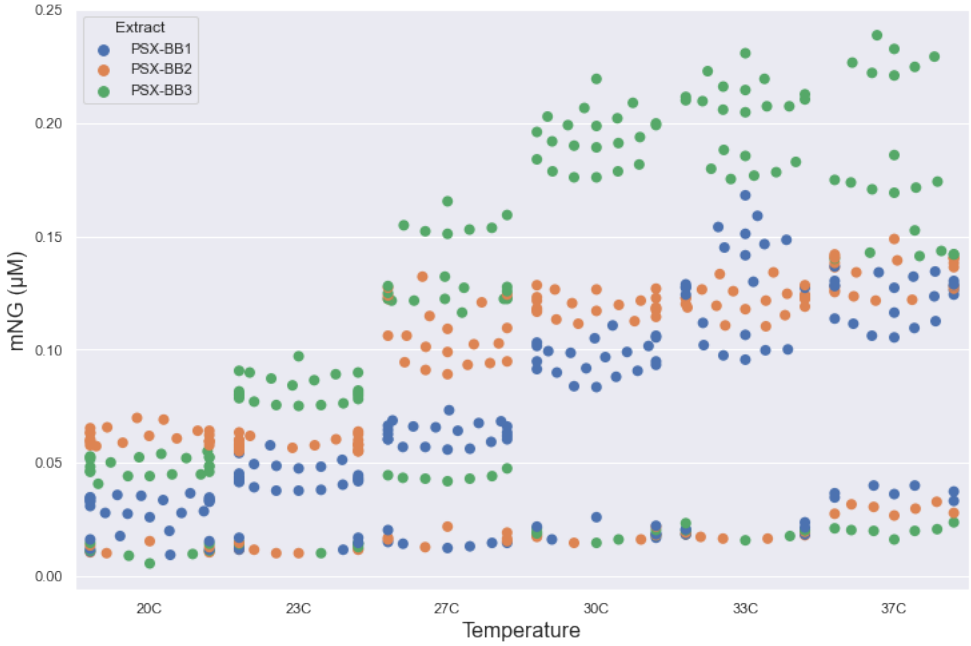


Figure 4.19: The same chart as Figure 4.18 with the *E. coli* data points removed, showing a small amount of PSX extract activity at higher temperatures.

Pseudomonas chlororaphis PCL1391

The growth rate of *P. chlororaphis* ("PCL") was measured in 2xYTPG in triplicate at 50 mL scale. Based on the growth curve in Figure 4.20, the cells were grown to OD 1.5 +/- 10% before harvest for extract production.

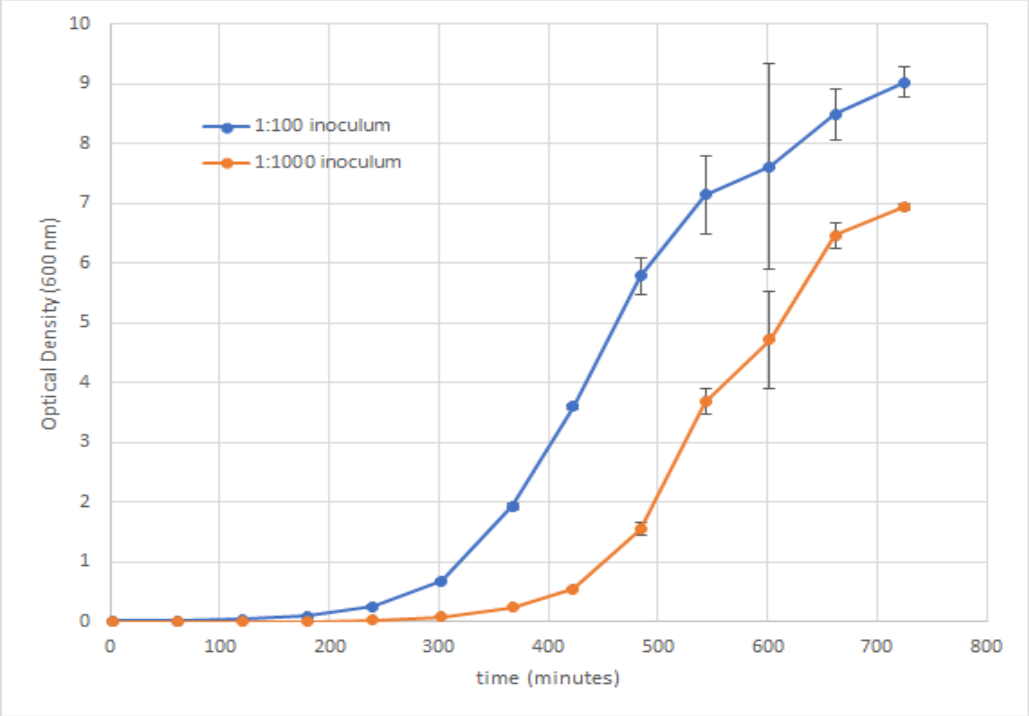


Figure 4.20: Growth of PCL at 50 mL scale in 2xYTPG.

Extract Production

At harvest, the cells sediment into a pale pink pellet. The three wash steps take 2-2.5 hours, and the cell pellet loosens slightly by the last wash step, requiring the supernatant to be removed by pipette instead of decanting.

The first extracts tested were prepared as biological triplicates, and each replicate was comprised of extract made with three different sonication amplitudes and three runoff

times. Figure 4.21 shows the protein concentration for the nine different preparation conditions and three *E. coli* extracts used as positive controls.

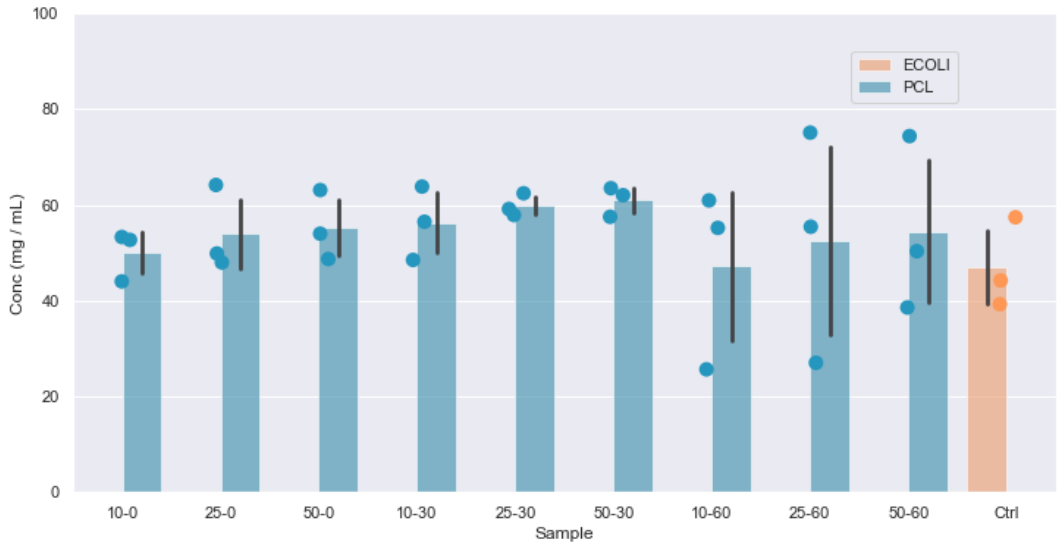


Figure 4.21: Protein concentrations as determined by Bradford assay. The sample labels at the bottom indicate "sonication amplitude" (120s process time, 5s on, 10s off) and "runoff time" (minutes). Bar heights are averages, error bars are standard deviation, and the dots are the individual measurements. These data were collected by Mark Prator.

Test reactions were prepared with the constitutive mNeonGreen Pa10403 plasmid at 10 nM and with 6 different salt conditions. The reactions were incubated at 30C and measured over 12 hours, shown in Figures 4.22 and 4.23. The salt conditions are the same as those used for the PSX TX-TL.

In Figure 4.23 we display the negative control wells, showing that the extract fluorescence is low without a DNA template and does not change significantly over the course of the experiment.

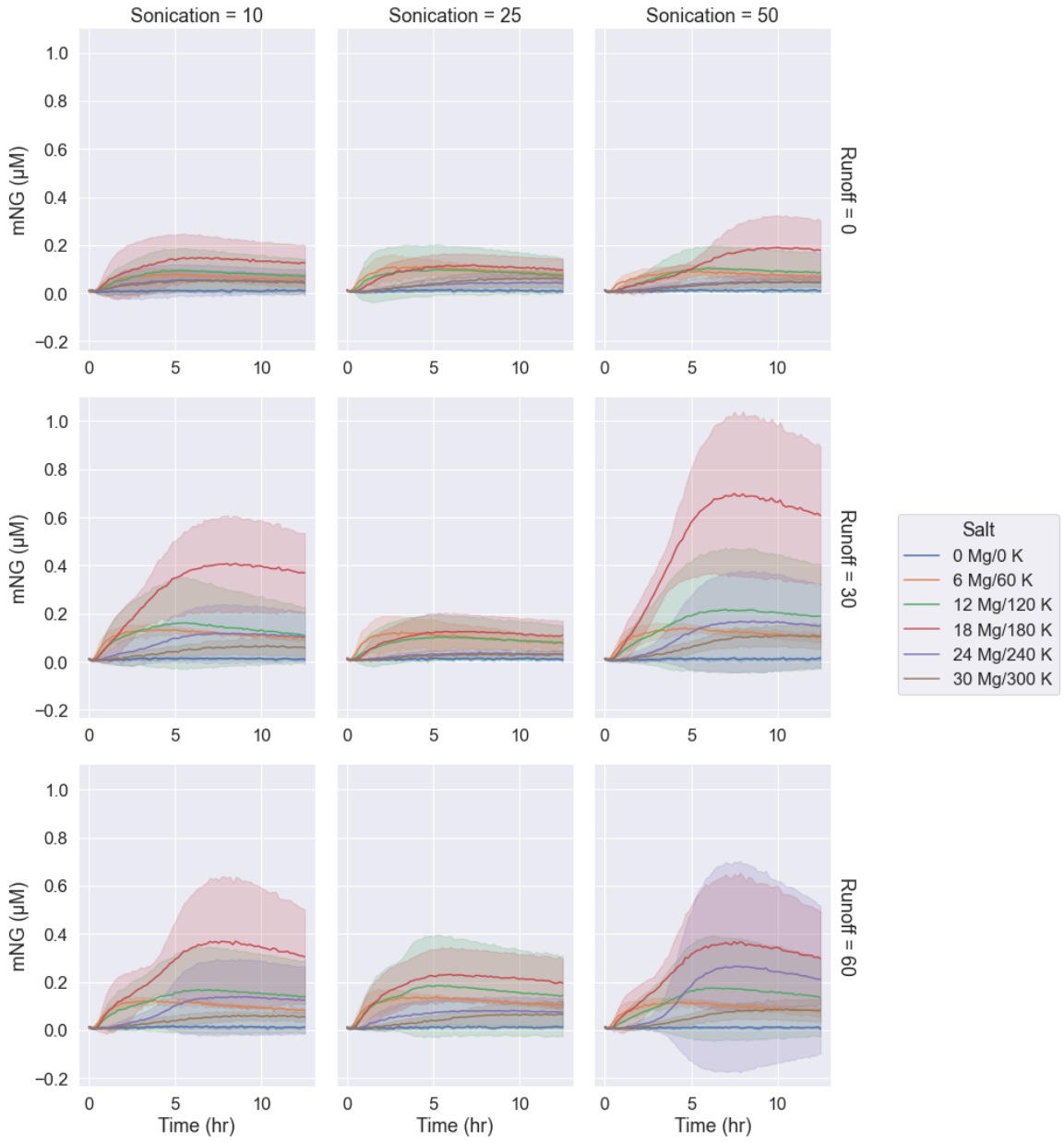


Figure 4.22: PCL extract, produced across varied sonication conditions (amplitude, 5s on 10s off, 120s total) and runoff times (minutes). Six magnesium-potassium salt concentrations were used for this test. Bands show standard deviation, and lines show means across replicates.

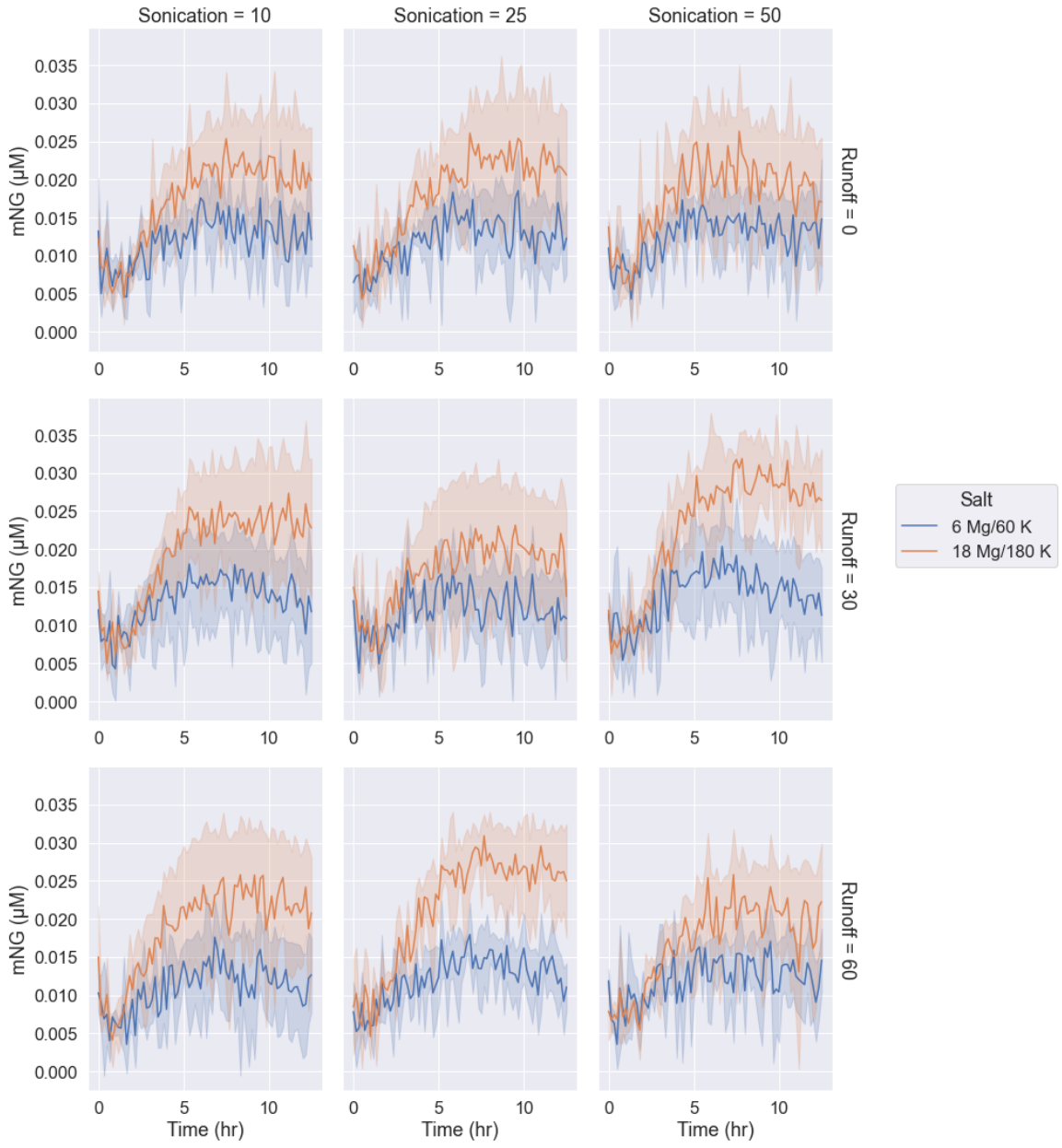


Figure 4.23: Negative controls with no DNA template added to PCL TX-TL reactions at two different salt conditions, showing relatively low baseline signal and small changes in fluorescence in the mNeonGreen band (Ex: 490 nm, Em: 520 nm).

Pseudomonas aureofaciens 30-84

The growth rate of *P. aureofaciens* ("PAU") was measured in 2xYTPG in triplicate at 50 mL scale. Based on the growth curve in Figure 4.24, the cells were grown to OD 3.0 +/- 10% before harvest.

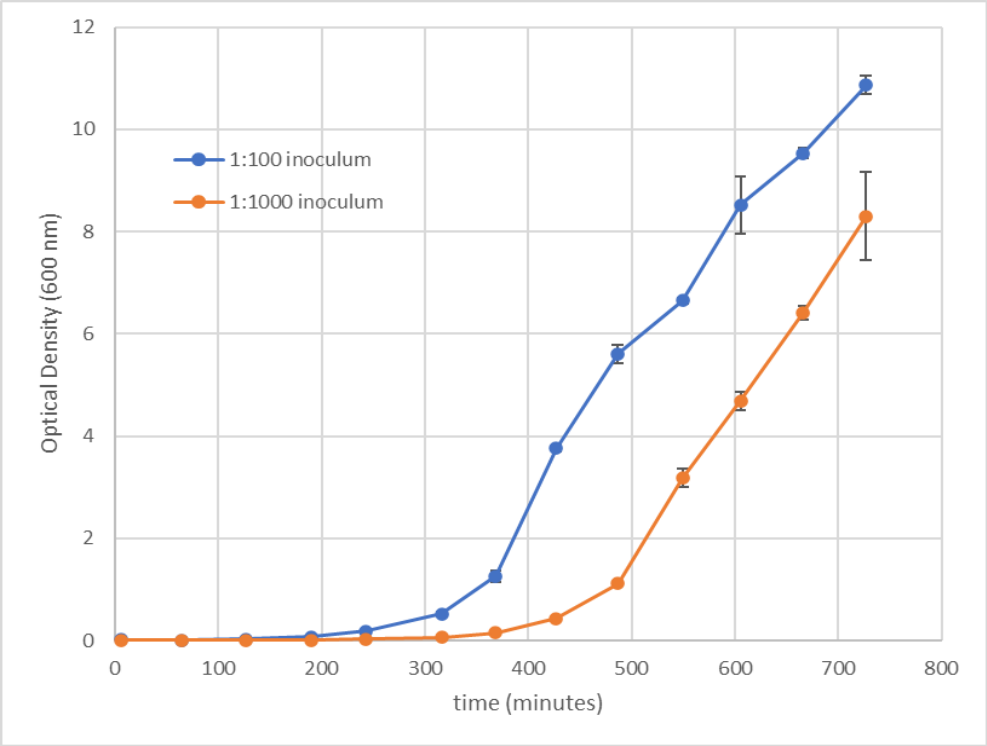


Figure 4.24: Growth of PAU at 50 mL scale in 2xYTPG media in triplicate.

Extract Production

At harvest, the cells sediment into a pale pink pellet. The three wash steps take 2-2.5 hours, and the cell pellet loosens slightly by the last wash step, requiring the supernatant to be removed by pipette instead of decanting.

The first extracts tested were prepared as biological triplicates, and each replicate was comprised of extract made with three different sonication amplitudes and three runoff

times. Figure 4.25 shows the protein concentration for the nine different preparation conditions and three *E. coli* extracts used as positive controls. The *E. coli* positive control extract was prepared at sonication amplitude 50 (120s, 5s on 10s off) and incubated for 60 minutes at 37C for the runoff reaction.

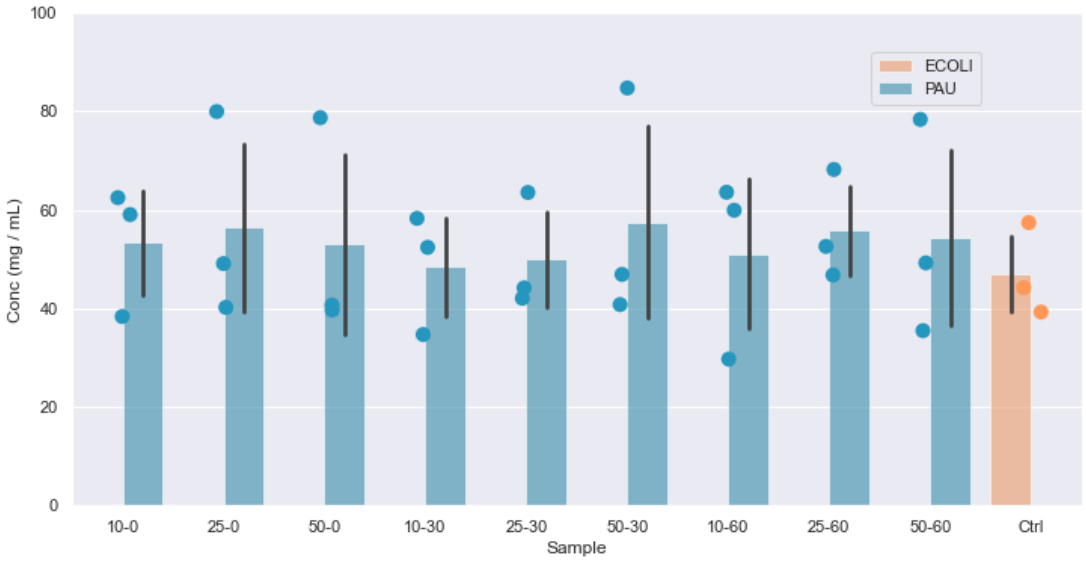


Figure 4.25: Protein concentrations as determined by Bradford assay. The sample labels at the bottom indicate "sonication amplitude" (120s process time, 5s on, 10s off) and "runoff time" (minutes). Bar heights are averages, error bars are standard deviation, and the dots are the individual measurements. These data were collected by Mark Prator.

Test reactions were prepared with the constitutive mNeonGreen Pa10403 plasmid at 10 nM and with 6 different salt conditions. The reactions were incubated at 30C and measured over 12 hours, shown in Figures 4.26 and 4.27. The salt conditions are the same as those used for the PSX TX-TL and the PCL TX-TL.

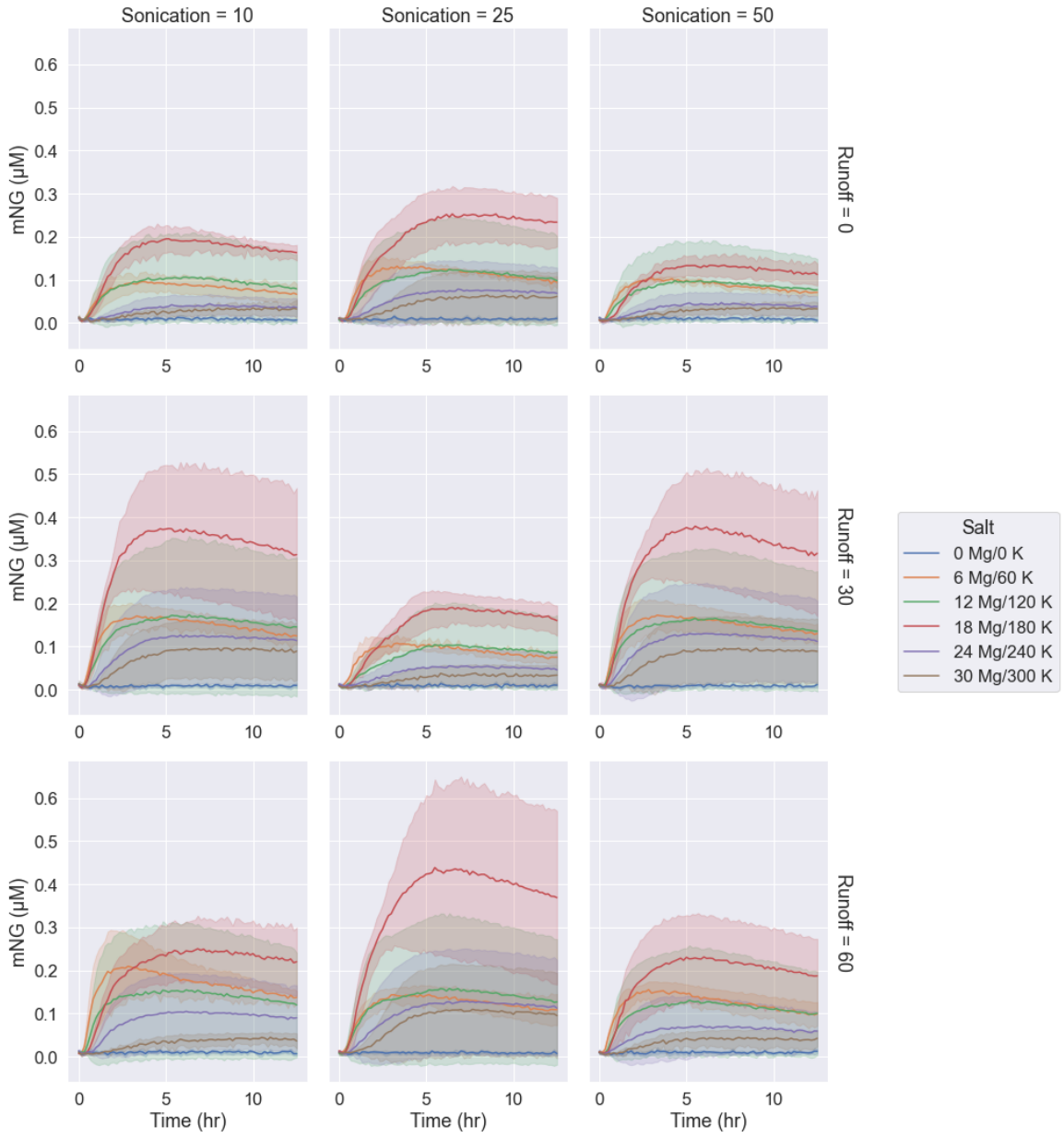


Figure 4.26: PAU extract, produced across varied sonication conditions (amplitude, 5s on 10s off, 120s total) and runoff times (minutes). Six magnesium-potassium salt concentrations were used for this test. Bands show standard deviation, and lines show means across replicates.

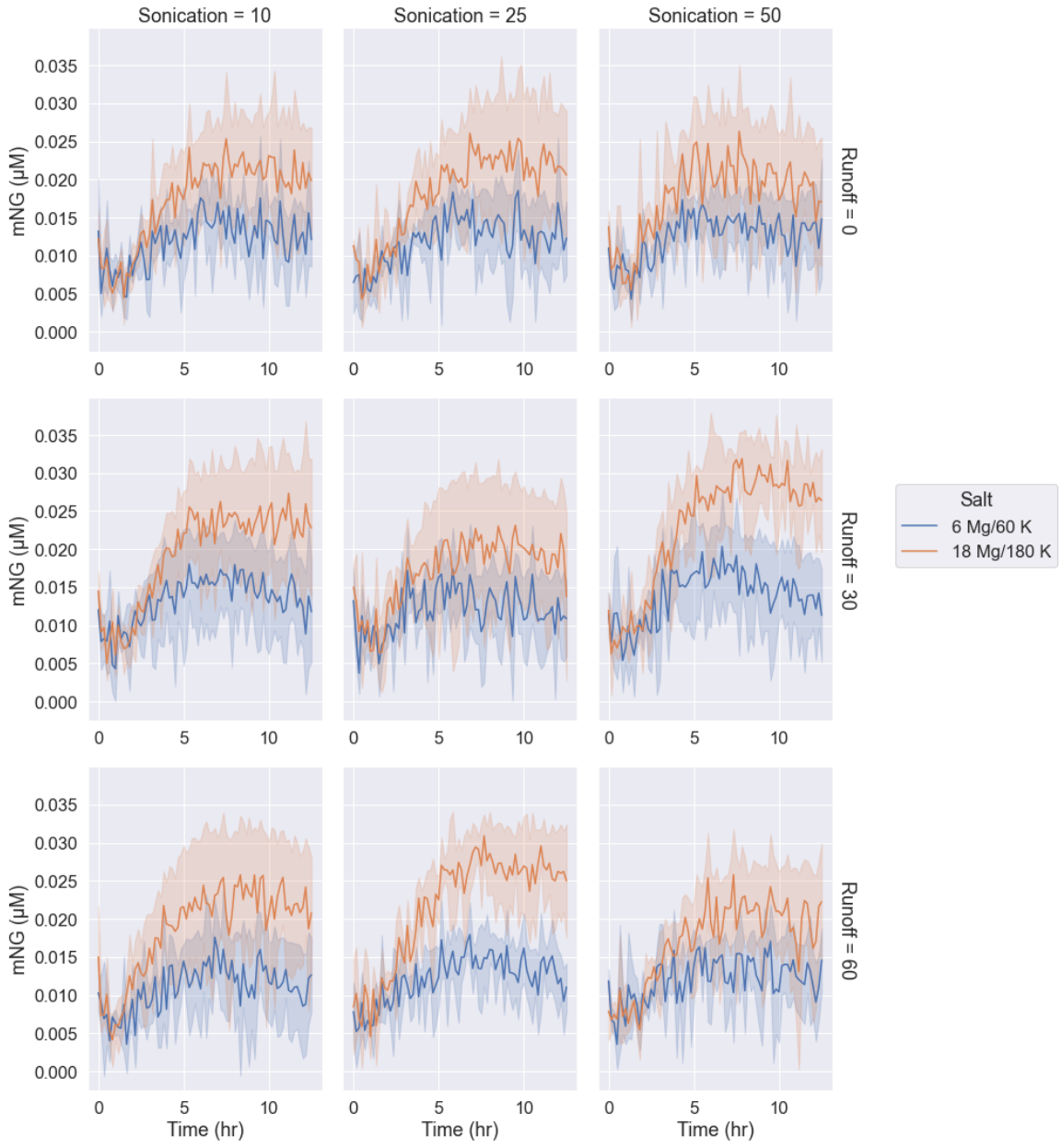


Figure 4.27: Negative controls with no DNA template added to PAU TX-TL reactions at two different salt conditions, showing relatively low baseline signal and small changes in fluorescence in the mNeonGreen band (Ex: 490 nm, Em: 520 nm).

In Figure 4.27 we display the negative control wells, showing that the extract fluorescence is low without a DNA template and does not change significantly over the course of the experiment.

Discussion and Conclusion

Together these experiments show that our target *Pseudomonas* spp. have pre-existing tolerance to the lignocellulosic growth inhibitors we tested, and these species are tractable for further engineering work. The TX-TL measurements of the promoter panel match the *in vivo* measurements of the same promoters. This shows the TX-TL system can be used for characterizing parts for later use *in vivo*. The ability to perform TX-TL reactions with these species is a key component of engineering and rapid prototyping with new non-model organisms. A working TX-TL system can enable a faster design-build-test cycle, and make characterization of new parts practical.

Acknowledgements

With thanks to Elin Larsson for performing cloning, genomic integrations in *P. synxantha*, *in vivo* measurements of the Anderson promoter panel, and writing parts of this manuscript, to Mark Prator for performing the Bradford assays of the lysates, to Zoila Jurado for purifying the mNeonGreen fluorescence standard, to Prof. Dianne Newman for the *Pseudomonas* strains and the Pa10403 plasmid, to Dr. R. Sidney Cox III for assistance with data visualization, to Ayush Pandey for assistance with data processing, to Dr. Dmitri Mavrodi for assistance with the genomic integration protocol for *P. synxantha*, and to Hasnat Rabi for assistance in producing illustrations. This research is supported by the Institute for Collaborative Biotechnologies through contract W911NF-19-D-0001, cooperative agreement W911NF-19-2-0026, and grant W911NF-09-0001 from the U.S. Army Research Office, the National Science Foundation through grant CBET-1903477, and the International Human Frontiers Science Program Organization

through HFSP-TX-TL. J.T.M. was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program and by a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745301. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. The content of this chapter does not necessarily reflect the position or the policy of the U.S. Government, and no official endorsement should be inferred.

References

1) Bilal, Muhammad, Shuqi Guo, Hafiz M. N. Iqbal, Hongbo Hu, Wei Wang, and Xuehong Zhang. 2017. "Engineering *Pseudomonas* for Phenazine Biosynthesis, Regulation, and Biotechnological Applications: A Review." *World Journal of Microbiology & Biotechnology* 33 (10): 191.

2) Haas, Dieter, and Geneviève Défago. 2005. "Biological Control of Soil-Borne Pathogens by Fluorescent Pseudomonads." *Nature Reviews. Microbiology* 3 (4): 307–19.

3) Nesemann, Kai, Susanna A. Braus-Stromeyer, Andrea Thuermer, Rolf Daniel, Dmitri V. Mavrodi, Linda S. Thomashow, David M. Weller, and Gerhard H. Braus. 2015. "Draft Genome Sequence of the Phenazine-Producing *Pseudomonas Fluorescens* Strain 2-79." *Genome Announcements* 3 (2). https://doi.org/10.1128/genomeA.00130-15.

4) Nakashima, Nobutaka, and Tomohiro Tamura. 2004. "Cell-Free Protein Synthesis Using Cell Extract of *Pseudomonas Fluorescens* and CspA Promoter." *Biochemical and Biophysical Research Communications* 319 (2): 671–76.

5) Wang, He, Jian Li, and Michael C. Jewett. 2018. "Development of a *Pseudomonas Putida* Cell-Free Protein Synthesis Platform for Rapid Screening of Gene Regulatory Elements." *Synthetic Biology* 3 (1): ysy003.

6) Silver, Pamela A., Jeffrey C. Way, Frances H. Arnold, and Joseph T. Meyerowitz. 2014. "Synthetic Biology: Engineering Explored." *Nature* 509 (7499): 166–67.

7) Letunic, Ivica, and Peer Bork. 2021. “Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation.” *Nucleic Acids Research* 49 (W1): W293–96.

8) Meier-Kolthoff, Jan P., and Markus Göker. 2019. “TYGS Is an Automated High-Throughput Platform for State-of-the-Art Genome-Based Taxonomy.” *Nature Communications* 10 (1): 2182.

9) Choi, Kyoung-Hee, and Herbert P. Schweizer. 2006. “Mini-Tn7 Insertion in Bacteria with Single attTn7 Sites: Example *Pseudomonas Aeruginosa*.” *Nature Protocols* 1 (1): 153–61.

10) Liow, Lu Ting, Maybelle Darlene Kho Go, and Wen Shan Yew. 2019. “Characterisation of Constitutive Promoters from the Anderson Library in *Chromobacterium Violaceum* ATCC 12472.” *Engineering Biology* 3 (3): 57–66.

11) Shaner, Nathan C., Gerard G. Lambert, Andrew Chammas, Yuhui Ni, Paula J. Cranfill, Michelle A. Baird, Brittney R. Sell, et al. 2013. “A Bright Monomeric Green Fluorescent Protein Derived from *Branchiostoma Lanceolatum*.” *Nature Methods* 10 (5): 407–9.

12) Sun, Zachary Z., Clarmyra A. Hayes, Jonghyeon Shin, Filippo Caschera, Richard M. Murray, and Vincent Noireaux. 2013. “Protocols for Implementing an *Escherichia Coli* Based TX-TL Cell-Free Expression System for Synthetic Biology.” *Journal of Visualized Experiments: JoVE*, no. 79 (September): e50762.

13) Kwon, Yong-Chan, and Michael C. Jewett. 2015. “High-Throughput Preparation Methods of Crude Extract for Robust Cell-Free Protein Synthesis.” *Scientific Reports* 5 (March): 8663.

Supplementary Information

SECTION A - Minimal growth medium recipe

CaCl2 - 0.68 mM

MgSO4 - 0.41 mM

NH4Cl - 16 mM

KH2PO4 - 4.7 mM

K2HPO4 - 2.3 mM – ***pH 7***

Glucose - 55 mM

MOPS - 25 mM

Aquil trace metal mix - 1 mL/L

Aquil trace metal mix recipe

Metal	Final molar concentration
$FeCl3^{+}$	8.41e-6 (Use 10 μM for Pseudomonas)
MnCl2.4H20	1.21e-7
CuSO4.5H20	1.96e-8
CoCl2.4H20	5.03e-8
Na2MoO4.2H20	1.0e-7
Na2SeO3	1.0e-8
ZnSO4.7H20	7.97e-8

Na2EDTA.2H20	100 uM

Section B - Figure 4.2 Expanded Detail

These measurements were conducted in a 96 well plate with cultures at 200 µL scale, and were grown at 30C.

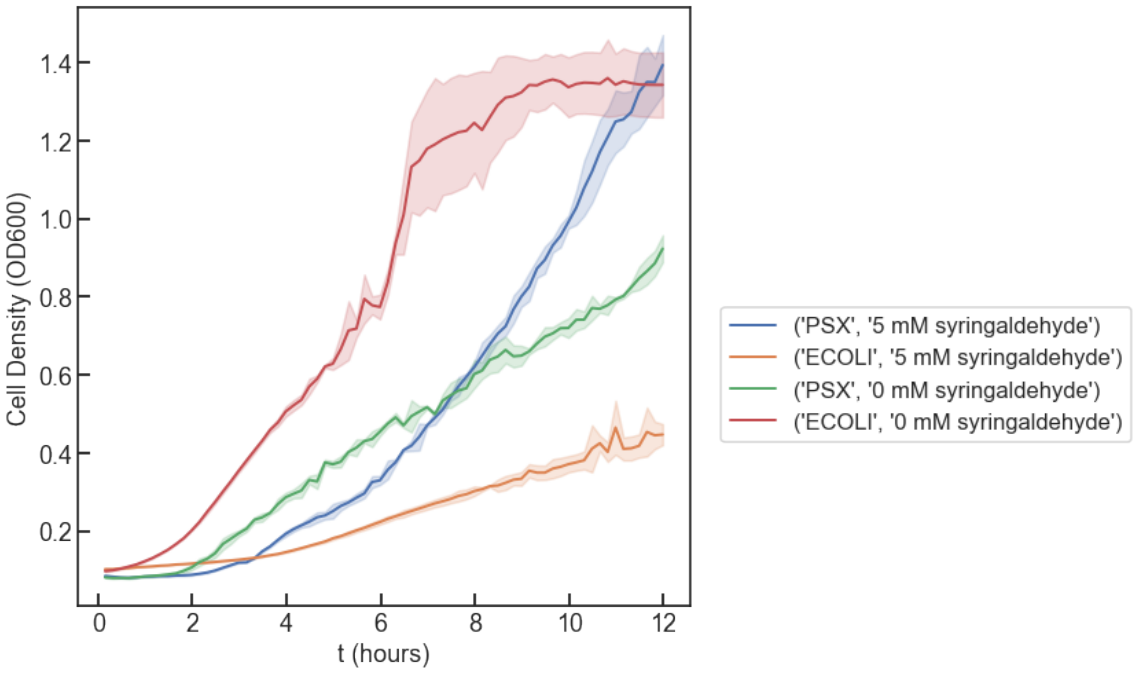


Figure 4.2-cont'd: PSX growth and *E. coli* growth in syringaldehyde

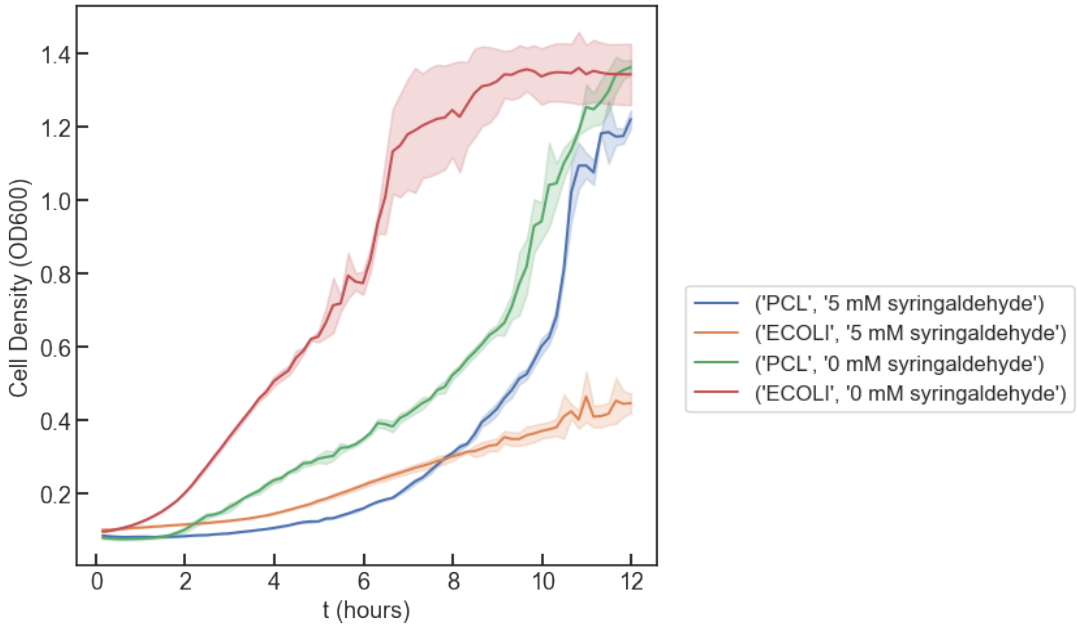


Figure 4.2-cont'd: PCL growth and *E. coli* growth in syringaldehyde

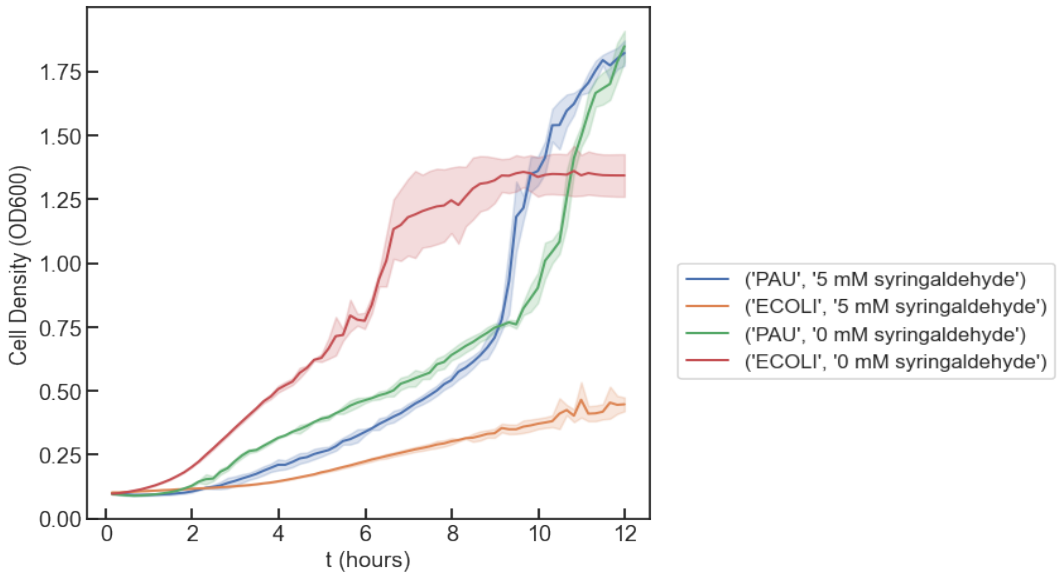


Figure 4.2-cont'd: PAU growth and *E. coli* growth in syringaldehyde

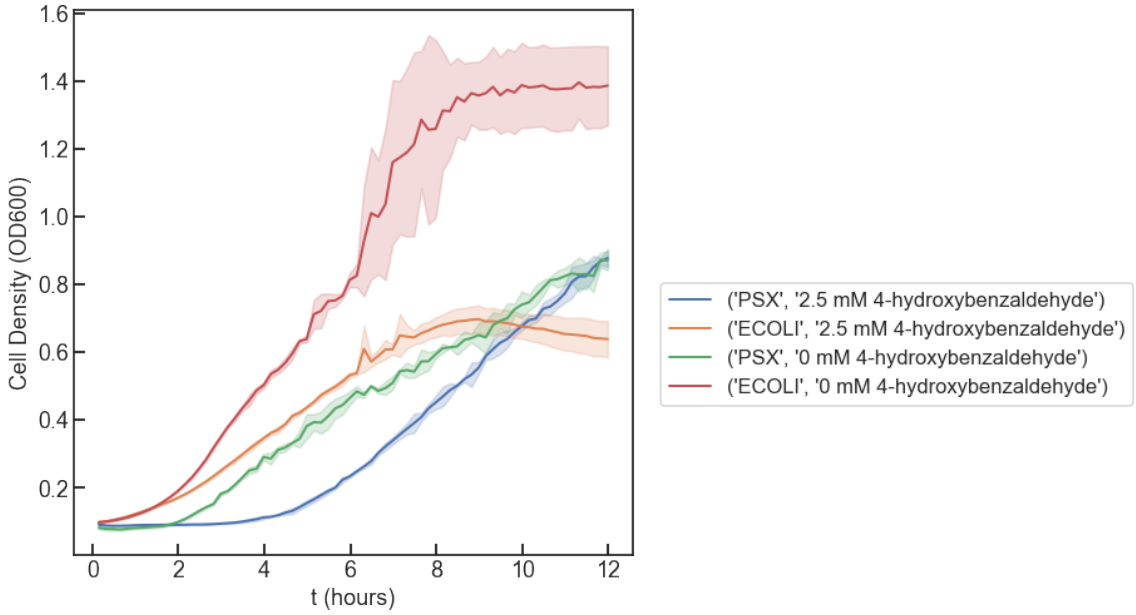


Figure 4.2-cont'd: PSX growth in 4-hydroxybenzaldehyde

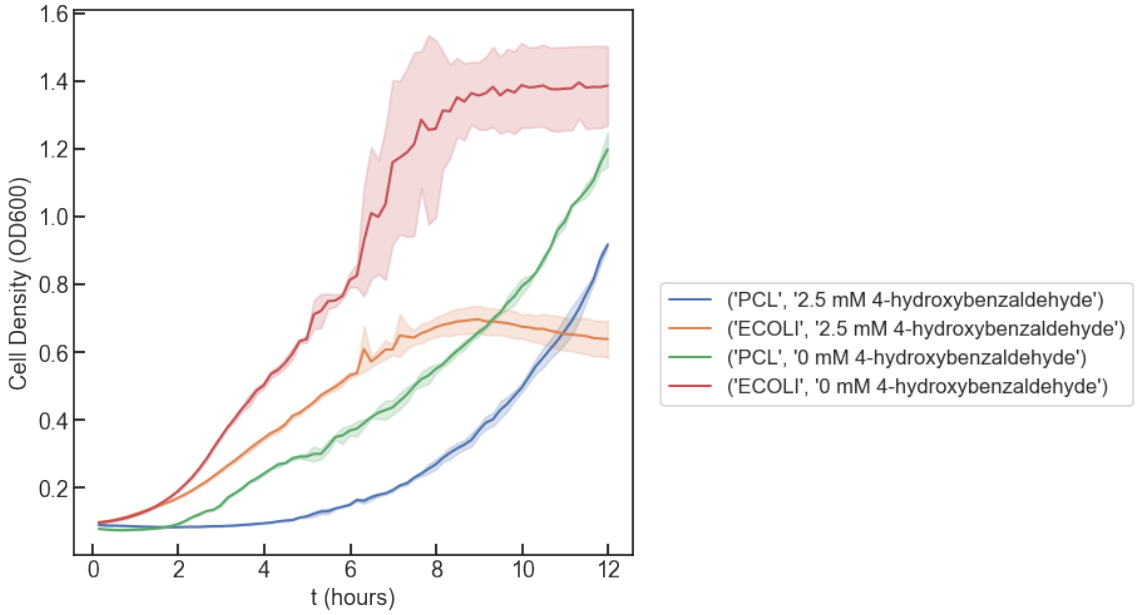


Figure 4.2-cont'd: PCL growth in 4-hydroxybenzaldehyde

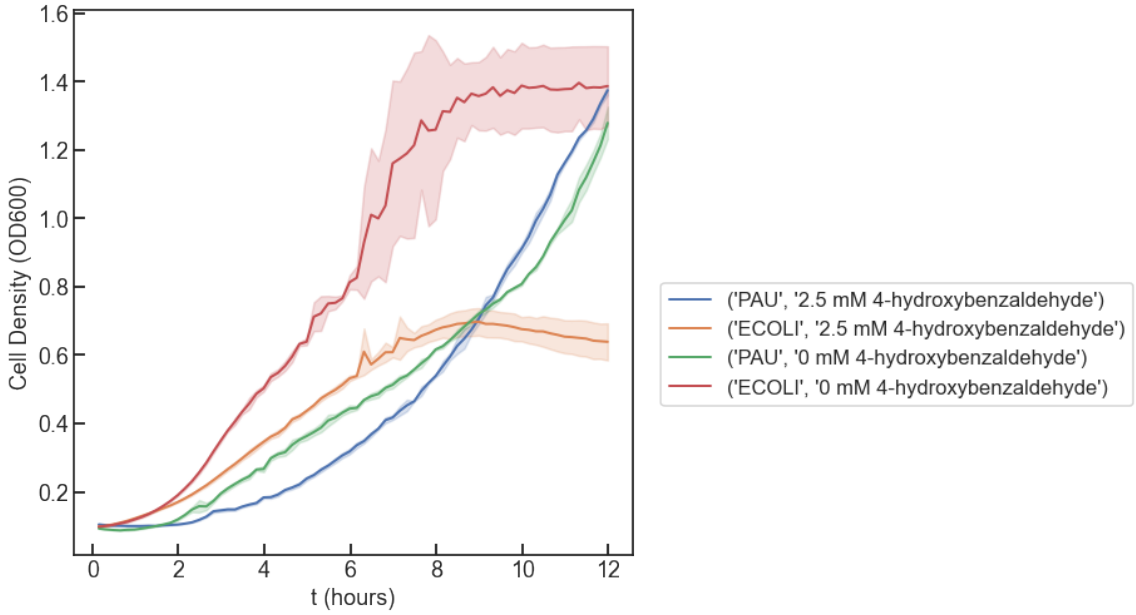


Figure 4.2-cont'd: PAU growth in 4-hydroxybenzaldehyde

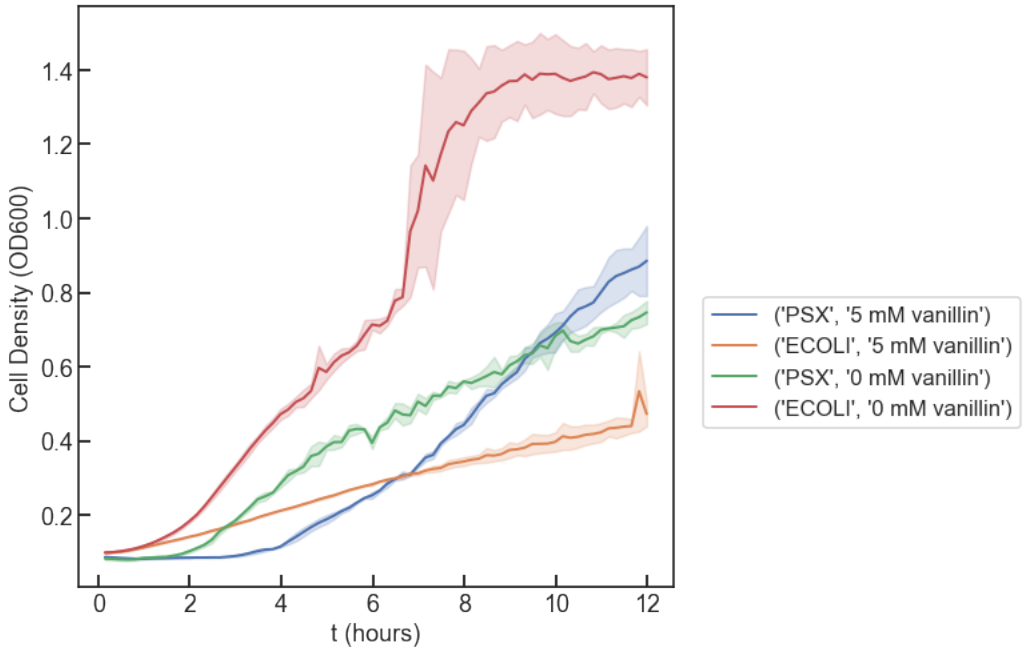


Figure 4.2-cont'd: PSX growth in vanillin

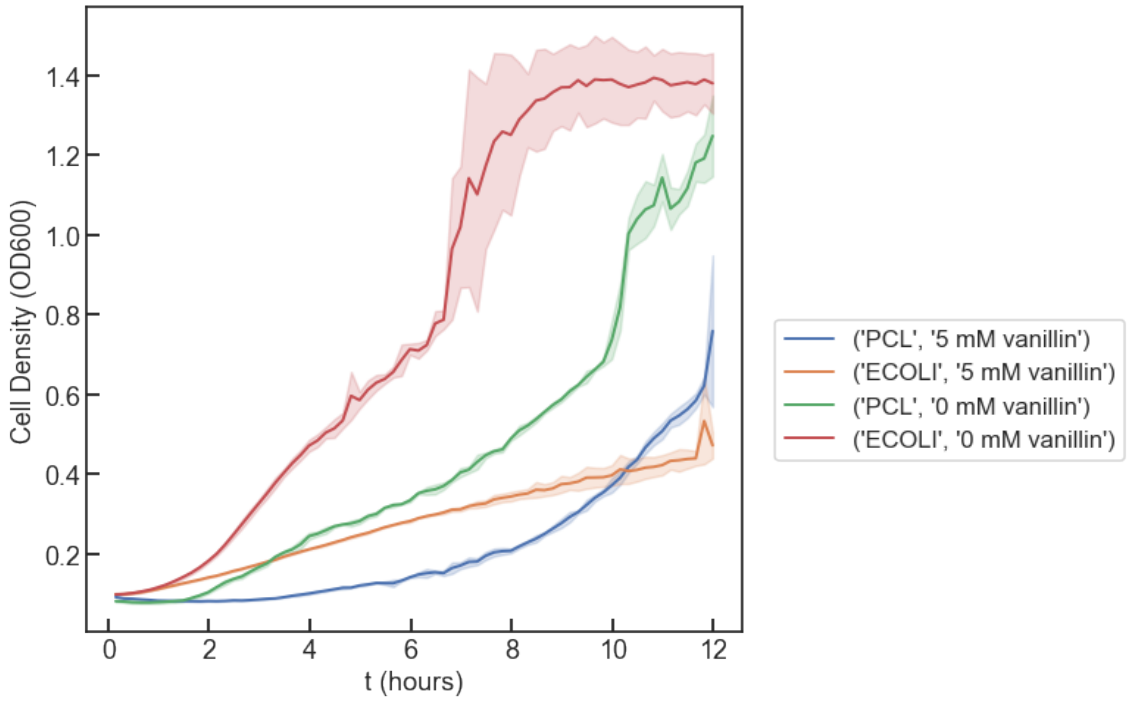


Figure 4.2-cont'd: PCL growth in vanillin

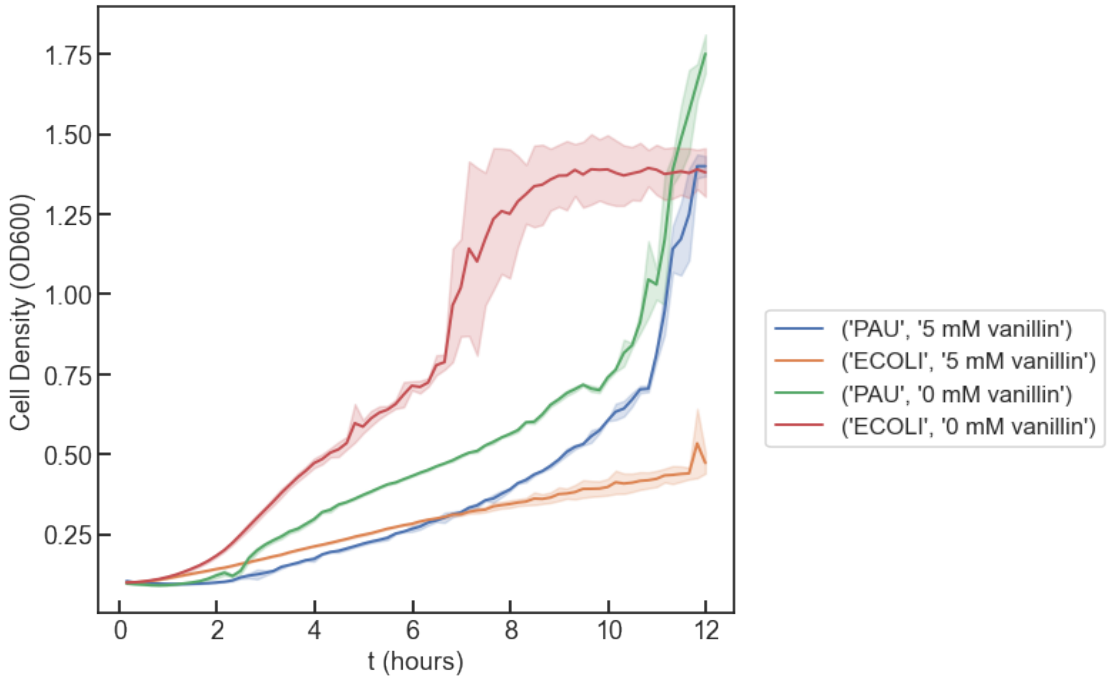


Figure 4.2-cont'd: PAU growth in vanillin

Chapter 5

CONCLUSION

If we can build something intentionally, then we understand how it works. The goal in the previous chapters is to show we successfully built working synthetic biological systems using existing parts and strains from previous studies. Richard Feynman described a bottom-up engineering approach to biology in his talk "There's Plenty of Room at the Bottom", where careful control of the different components of biology would allow for molecule by molecule construction of biological systems. He asked "what would happen if we could arrange the atoms one by one the way we want them?" (1). Biology can produce controlled patterns of DNA, RNA, and protein. Our understanding of the biological parts is a limiting factor in getting these components to perform a given task. Arranging atoms any way we want requires us to exercise some control over biology, and techniques like TX-TL allow us to operate the machinery of biology for a functional design-build-test cycle.

While traditionally *E. coli* has been one of the most used bacteria for engineering, and *E. coli* TX-TL has been one of the most useful methods for rapid prototyping, this bacterium has limits. It doesn't grow everywhere – if we are trying to arrange atoms in soil, the gut-derived *E. coli* is not evolved to grow well in that environment. There is a tremendous diversity of different bacteria, each adapted to different environments and ecosystems (2). Even within the human microbiome there may be ecological niches like the mouth that might grow *Streptococcus mutans* more readily than *E. coli* (3). There are any number of circuits that may be better expressed in different species of bacteria. The tools for engineering in *E. coli* such as cloning and TX-TL are not always available for other

species, so our work here shows the successful transfer of the *E. coli* TX-TL method to non-model species.

Species Name	TX-TL Progress	Reason for Developing
Deinococcus radiodurans	Cells grown, several extract attempts made	Radiophile
Deinococcus geothermalis	Cells grown, several extract attempts made	Radiophile, moderate thermophile
Halomonas elongata	Cells grown, several extract attempts made	Halophile
Geobacillus stearothermophilus	Cells grown, one extract attempt made	Thermophile
Geobacillus LC300	Cells not grown to high density	Thermophile
Bacillus firmus OF4	Cells grown, one extract attempt made	Alkaliphile
Acetobacter aceti	Cells grown, one extract attempt made	Acidophile
Azotobacter vinelandii DJ	Cells not grown to high density	Nitrogen-fixation, soil microbiome
Paenibacillus polymyxa	Cells not grown to high density	Nitrogen-fixation, soil microbiome
Kocuria rhizophila	Cells not grown to high density	Soil microbiome
Marinobacter vinifirmus	Unable to obtain original strain from source lab	Perchlorate tolerance
Streptococcus mutans	Cells not grown to high density	Human commensal
Citrobacter freundii	Cells not grown to high density	Human commensal

Table 5.1: Bacterial species used in other TX-TL attempts.

The *Pseudomonas spp* were the only extracts we got working out of multiple species we tried to use for new TX-TL systems. We spent multiple years attempting to grow various extremophiles well enough to attempt making lysates for TX-TL and running TX-TL reactions with new species. Native promoters were used for the DNA templates where possible. Enzymes like T7 RNA polymerase won't work at high temperature, for example. The list of other species we tried to use for TX-TL, and did not succeed with, is in Table 5.1.

In each case the organisms above were chosen because they had a fast growth rate in the prior literature, a known aerobic growth medium, and a source for the strain. Finding a protocol to grow these organisms to high density was a challenge for many of the species above. For example, *B. firmus OF4* is an alkaliphile that can grow at pH of 10, and is supposed to have a non-neutral cytoplasmic pH (4). Several media were tested, and eventually we found that it would only grow to a high OD at high pH if a trace mineral solution was added to the culture. The carbonate buffer used to hold the high pH for OF4 had a tendency to precipitate if the media was left for long enough at room temperature, but it grew fast enough to avoid problems with contamination.

Acetobacter aceti was grown on "A4" media, the fourth iteration of finding a good medium for growth to high density. A4 was 15 g/L proteose peptone and 15 g/L yeast extract, and *A. aceti* would only grow to a high density if magnesium sulfate was also added to the media. *A. aceti* grew slow enough that contamination was an issue, especially in a rich media without antibiotics like A4, and our tools for monitoring contamination are limited. We can go through the process of growing the culture,

washing it, freezing the pellet, and then get sequencing back. Sequencing or restreaking the culture out typically takes a day, much slower than the 6-8 hour culture growth, so if the culture is contaminated we would not find out in time to cancel the centrifuging and wash steps at the end. *D. radiodurans* also grew slowly, and produces a red pellet. When we made a pellet with streaks of other colors like white we could tell that *D. radiodurans* culture was contaminated, but this technique is limited. White cell pellets don't show white colored contaminants. The best approach we found was monitoring the optical density carefully – if the culture started growing at a much faster doubling time than usual then it was presumed to be contaminated. Microscopy may have worked, but imaging alone was insufficient to identify all contaminated cultures.

Once the first extracts were made, T7 RNA polymerase promoters or native promoters were used for testing and no protein expression was found with either. Varying the harvest conditions and reaction conditions didn't change the result. We don't have a straightforward answer for why some species work for TX-TL and others don't yet, but our limited experience was that gram-positive bacteria did not produce lysates that worked for TX-TL, and all of the TX-TL successes were with gram-negative bacteria.

Working with these non-model organisms offers the promise of engineered bacteria with capabilities beyond those available from *E. coli*. Some of these organisms such as *Acetobacter aceti* have a proteome adapted to an extreme environment at low pH and high acetate levels. These adaptations may enable the organism (or its lysate) to perform tasks that *E. coli* cannot do (5). *E. coli* can be engineered for improved tolerance to certain growth inhibitors (6), but has limits to what it can express (7). Developing

working engineering tools and protocols for these non-model organisms involves building some meaningful understanding of the organism and its parts, and building that understanding enables these non-model chassis to be used in the future for new applications.

Citations:

1) Feynman, Richard P. 1960. “There’s Plenty of Room at the Bottom.” *Engineering and Science* 23 (5): 22–36.

2) Hug, Laura A., Brett J. Baker, Karthik Anantharaman, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, et al. 2016. “A New View of the Tree of Life.” *Nature Microbiology* 1 (April): 16048.

3) Lemos, J. A., S. R. Palmer, L. Zeng, Z. T. Wen, J. K. Kajfasz, I. A. Freires, J. Abranches, and L. J. Brady. 2019. “The Biology of *Streptococcus Mutans*.” *Microbiology Spectrum* 7 (1). https://doi.org/10.1128/microbiolspec.GPP3-0051-2018.

4) Sturr, M. G., A. A. Guffanti, and T. A. Krulwich. 1994. “Growth and Bioenergetics of Alkaliphilic *Bacillus Firmus OF4* in Continuous Culture at High pH.” *Journal of Bacteriology* 176 (11): 3111–16.

5) Menzel, Ulrike, and Gerhard Gottschalk. 1985. “The Internal pH of *Acetobacterium Wieringae* and *Acetobacter Aceti* during Growth and Production of Acetic Acid.” *Archives of Microbiology* 143 (1): 47-51

6) Dunlop, Mary J., Zain Y. Dossani, Heather L. Szmidt, Hou Cheng Chu, Taek Soon Lee, Jay D. Keasling, Masood Z. Hadi, and Aindrila Mukhopadhyay. 2011. “Engineering Microbial Biofuel Tolerance and Export Using Efflux Pumps.” *Molecular Systems Biology* 7 (May): 487.

7) Thomas, Jeffrey G., and François Baneyx. 1996. “Protein Misfolding and Inclusion Body Formation in Recombinant *Escherichia Coli* Cells Overexpressing Heat-Shock Proteins (∗).” *The Journal of Biological Chemistry* 271 (19): 11141–47.