© Copyright 2018

Jason Thomas Stevens
Controlling Enzyme Expression Dynamics to Improve Production from Engineered Biosynthetic Pathways

Jason Thomas Stevens

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:
James Carothers, Chair
Wendy Thomas
Paul Wiggins

Program Authorized to Offer Degree:
Bioengineering
Metabolic engineering promises to reduce our reliance on non-renewable chemical synthesis methods by harnessing microbial metabolisms to convert simple renewable resources, such as sugars, into useful chemicals, such as biofuels. Most metabolic engineering efforts require expressing heterologous enzymes in a microbial host, such as *E. coli*. Both the process of producing enzymes and the chemical metabolites that are produced can have deleterious impacts on host...
fitness, however. Here we detail efforts to use experimental and computational methods to better understand how to mitigate the deleterious impact of a metabolic pathway on its host. Firstly, we sought to computationally predict how metabolic pathway titers could be improved, and product toxicity reduced, by implementing genetic feedback networks that can sense pathway metabolite concentrations in a cell and respond by up- or down-regulating enzyme levels. To this end, we developed a computational methodology for modeling and simulating large sets of genetic feedback networks acting on a metabolic pathway that produces 4-aminocinnamic acid, or 4-ACA. Our analysis revealed genetic feedback network architectures and implementations that promise to improve pathway titers. Secondly, we experimentally analyzed the burden imposed by expressing a Phenylalanine Ammonia Lyase enzyme, PAL2 from Arabidopsis thaliana, which converts 4-aminophenylalanine (4-AF) to 4-ACA in Escherichia coli cells. We uncovered how the timing and level of induction of a PAL2-superfolder GFP fusion (PAL2-sfGFP) influences cell growth, expression of PAL2-sfGFP, and in vivo cellular catalysis of 4-AF to 4-ACA conversion. We then identified stationary phase promoters that can be used to autonomously express high levels of PAL2-sfGFP, increase conversion, and reduce deleterious impacts. Lastly, we produced 4-AF from glucose in E. coli using a three-enzyme operon, PapABC, from Pseudomonas fluorescens, achieving titers of around 100 mg/L, forming the basis for a pathway to produce 4-ACA. Together, this work serves to improve our understanding of how to mitigate the deleterious impacts of a metabolic pathway on its host by using
computational model analysis and by controlling in vivo enzyme expression dynamics.
# Table of Contents

Table of Contents 1

List of Figures 3

List of Tables 4

Chapter 1 | Introduction 7
  Abstract 7
  1.1 | Selected Topics in Metabolic Engineering 7
    1.1.1 | Introduction 7
    1.1.2 | Producing Aromatic Compounds with E. coli 8
    1.1.3 | Toxicity and Burden in Metabolic Engineering 11
    1.1.4 | Programming Gene Expression of Metabolic Pathway Enzymes in E. coli 14
      1.1.4.1 | Promoters 14
      1.1.4.2 | Ribosome Binding Sites 16
      1.1.4.3 | mRNA Transcript Stability 17
      1.1.4.4 | Dynamic Control of Gene Expression 18
    1.1.5 | Kinetic Models of Metabolic Pathway Gene Expression 21
  1.2 | Dissertation Organization 24

Chapter 2 | Computational Analysis of Genetic Control Systems for Metabolic Pathway Engineering 26
  Abstract 26
  2.1 | Introduction 27
  2.2 | Methods 33
  2.3 | Results and Discussion 36
  2.4 | Supplementary Information 49
  2.5 | Modular Modeling with PyMDC 54
  2.6 | Conclusions and Future Work 56

Chapter 3 | Programming Enzyme Expression Dynamics to Reduce Cellular Burden and Improve Expression and In Vivo Catalysis 60
  Abstract 60
  3.1 | Introduction 61
  3.2 | Materials and Methods 66
  3.3 | Results 70
# List of Figures

2.1 | Engineered p-aminostyrene (p-AS) pathway and control system design space 32

2.2 | Mechanistic model and genetic control mechanisms 38

2.3 | Relative production yields across the entire design space 41

2.4 | Performance of core topologies 43

2.5 | Relative production yields for Core Topology 12 (CT 12) 46

2.6 | Dependence of relationship between sensor input and node output on aptazyme ligand-activation ratio 48

2.7 | Base module hierarchy of PyMDC 57

3.1 | Construct and experimental designs 71

3.2 | Measurements of burden and expression of PAL2-sfGFP and sfGFP 72

3.3 | Flow cytometry measurements of population distributions 74

3.4 | Induction times relative to ininduced culture growth rate 77

3.5 | *In vivo* conversion of 4-AF to 4-ACA at 24 hours post-inoculation 78

3.6 | Comparison of tet and σ38 promoters 79

4.1 | Biosynthesis pathway for chorismate from D-erythrose-4-phosphate (E4P) 100

4.2 | Biosynthesis of 4-ACA from chorismate 102

4.3 | 4-AF-HCl effect on expression from 5' UTR aRED constructs 114

4.4 | HCl effect on expression from 5' UTR aRED constructs 115
List of Tables

4.1 | Shikimate production data 107
4.2 | Highest observed 4-AF titers for the constructs shown 111
4.3 | Impact of growth temperature on 4-AF titer 112
4.4 | Impact of induction time on pathway titer 113
Acknowledgments

This work is the synthesis of innumerable insights and ideas, most of which were not my own. It would not have been possible without the care, resources, and time of many others, some of whom I would like to acknowledge here. I would first like to thank James Carothers, my adviser, for providing me with an amply supplied and sunny lab in which to do my work and the freedom to pursue the questions that interested me in a manner that allowed me to maintain a healthy lab-life balance. I would like to thank my labmates for creating a positive and helpful environment and for assisting me with navigating the challenges of doing rigorous lab work. I would especially like to thank Willy Voje, for our many conversations on lab and life that continually bolstered and strengthened me over these five and a half years and for encouraging my adventures in music that made these years so much more rich and exciting. I would also like to thank my fellow BioEs Andrew Wang, Blake Bluestein, Bob Lamm, David Younger, and Jon Tsui for being the most tremendous set of friends. Grad school was infinitely more fun and manageable with you all. I would also like to thank the many other people I’ve had the immense good fortune to call friends: Anna, Arjun, Caitlin, Claire, Eric, Jessica, Jasmin, Joe, Leandra, Leli, Stephanie, Sunny, Surya, and others. I would like to thank my family for their constant and unwavering support. Lastly, I would like to thank Christine Texeira for being the most thoughtful and supportive and funny person/girlfriend. The time we've spent together was the greatest reward to come out of grad school.
Dedicated to my parents
Chapter 1 | Introduction

Part of this work (located in Section 1.1.4) was published as and reprinted with permission from: Programming Gene Expression by Engineering Transcript Stability Control and Processing in Bacteria. Stevens JT, Carothers JM. Synthetic Biology: Parts, Devices, Applications. 2018 May. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.

Abstract

What follows is a brief overview of several topics of relevance to the work presented in this dissertation. Topics include: the production of aromatic compounds, which is the class of compounds discussed in this work; toxicity and burden to cellular hosts that arises during metabolic engineering efforts; programming gene expression in microbes; and a discussion of kinetic modeling of genetic and metabolic networks. This is followed by an overview of how this dissertation is organized.

1.1 | Selected Topics in Metabolic Engineering

1.1.1 | Introduction

Metabolic engineering seeks to harness the capacity for living systems — often microbes — to convert renewable materials, such as sugars, into valuable compounds, such as drugs or biofuels, via metabolic processes [1]. This conversion is done by a series of enzymes expressed by microbes that act like an assembly line,
converting one compound into the next until the final product is made. Microbes natively produce hundreds of enzymes and compounds, and the reportoire of possible metabolic products can be expanded by introducing new enzymes from other microbes, plants, or mammalian systems [2]. Many thousands of potential metabolic pathways have now been identified as feasible means of production in microbes. The challenge lies in improving the titer, yield, and rate of production until a bio-based manufacturing process is economically competitive with traditional, often petroleum-based, synthesis.

The last fifteen years have seen an explosion of interest and research in metabolic engineering and related areas. The time it takes to optimize a metabolic pathway has been greatly reduced due to advances in gene assembly, automation, gene expression engineering, high-throughput analysis platforms, predictive modeling, and fundamental advances in understanding the biology underpinning metabolic pathways. Yet, it can still require many person-years and millions of dollars to fully develop a novel metabolic pathway. There is still much to understand about microbial behavior, how it foils engineering efforts, and how it can be exploited or overcome.

1.1.2 Producing Aromatic Compounds with E. coli

Aromatic monomers are attractive targets for microbial synthesis, as many aromatic polymers possess desirable properties such as high thermal resistance and mechanical strength, making them potentially useful in a wide variety of
advanced applications [3]. *E. coli* natively produces three aromatic amino acids: phenylalanine, tyrosine, and tryptophan, making it a good host for microbial production of aromatics. The metabolic pathway is largely the same for all three compounds, with a branch point at chorismate, a precursor for numerous other metabolic products in *E. coli*. This pathway is well studied and often the starting point for biosynthesis of novel aromatic compounds in *E. coli*.

Native regulation exists that needs to be overcome in order to produce high levels of aromatics, however. Expression of several enzymes in the pathway is subject to transcriptional negative feedback, whereby a transcriptional regulator binds to a product then downregulates enzyme expression. Three enzymes, which catalyze the same reaction, are also subject to feedback inhibition, whereby a downstream product inhibits the enzyme from creating more product. Both transcriptional negative feedback and feedback inhibition are used to control levels of compounds in a cell, which is helpful in natural environments but can thwart attempts to produce high levels of a compound in an engineering context. In the case of aromatic biosynthesis, the regulators TyrR and TrpR bind aromatic amino acids and tryptophan, respectively, then downregulate two different points in chorismate biosynthesis, aroG and aroL. aroG, aroF, and aroH (all DAHP synthases) are feedback inhibited by phenylalanine, making the reaction they catalyze a key regulatory point for the pathway. Thus, attempts to overproduce aromatics often include expressing a feedback-resistant aroG mutant [4], defunctionalizing or knocking out tyrR and trpR, and/or overexpressing the enzymes that are
downregulated, aroG and aroL [5]. These measures eliminate most of the negative feedback and thereby greatly increase the limit of aromatic production.

The intermediate in the E. coli aromatic pathway that has attracted perhaps the most attention is shikimate, a precursor to many important compounds, including Tamiflu. Using, amongst other means, the aforementioned strategies to undo native pathway regulation, yields on glucose of around 30% have been reported [6]. This work is the foundation for producing most aromatics in E. coli.

Of interest to our lab is a pathway for microbial production of 4-aminostyrene (4-AS), which has yet to be produced in microbes, for two principal reasons. The first is that a precursor to 4-AS, 4-aminocinnamic acid (4-ACA), is highly toxic to E. coli, making this pathway a potential testbed for studying the mitigation of product toxicity to host cells. The second is that an RNA aptamer exists for another precursor in the pathway, 4-aminophenylalanine (4-AF), which could be used to develop both in vitro RNA ribosensors and in vivo RNA-based genetic controllers [7], which could be used to develop genetic control systems to mitigate product toxicity. Similar pathways to the 4-AS pathway, namely, the 4-hydroxystyrene [8] and styrene [9] pathways, also note product toxicity as a hinderance to achieving higher titers.

In 2003 [10] and 2004 [11], two pathways were used to produce 4-aminophenylalanine (4-AF) in E. coli, albeit without the goal of optimal production. This work was later built on to develop a 4-AF pathway with a 17% yield on glucose [12]. PAL4, a phenylalanine ammonia lyase from A. thaliana, has been
used to convert 4-AF to 4-ACA using resting cells [13,14], and produced in growing cells to a level of 3 mg/L [15]. Thus, there is still much work to be done to create high levels of 4-ACA and 4-AS in cells.

1.1.3 | Toxicity and Burden in Metabolic Engineering

While toxicity and burden can result in similar, unwanted phenotypes in engineered cells, such as poor growth and low productivity [16–18], they stem from different sources. Toxicity arises from deleterious interactions of a substance, such as a compound or protein, with microbial physiology, as an antibiotic would. Many molecules under study for metabolic pathway engineering exhibit toxicity, often at concentrations much lower than desired titers [19]. Burden arises from the load that is placed on a cell by extraneous processes, such as producing heterologous enzymes or genetic controller proteins. Both toxicity and burden can prevent cells from achieving their optimal potential as cellular factories; thus, these phenomena, and how to mitigate their effects, are active areas of study [19–21].

Much of toxicity mitigation research relates to methods for extracting the offending substance from microbes themselves or from the culture. Efflux pumps are used by microbes to expel unwanted molecules, such as antibiotics. Studies to find [22] or engineer [18,23] efflux pumps to expel toxic metabolic products-of-interest have shown promise. To remove toxic metabolites from culture, several means have been developed for chemical extraction and recovery, such as solvent layers that sequester metabolites [24,25].
Much of the focus of burden research is on the zero-sum game of resource allocation within cells. When cells are tasked with performing new functions, such as producing large quantities of enzymes, the materials and energy to perform this function necessarily come from the cellular pool reserved for robust growth and homeostasis. It is thought that ribosomes are the limiting quantity in a cell, and a simple model of ribosome allocation can predict reductions in growth rate due to overexpression of protein [16]. Including an additional simple model of nutrient/energy allocation can allow for predictions of the interplay between nutrient quality, heterologous protein expression level, and growth rate [26]. But cells do not always exist in steady-state, as these models assume, especially in a batch reactor, common in metabolic engineering. The pool of available ribosomes per cell expands as cells adjust to new media [27]; for instance, when starting a culture for an experiment or production reaction. As the ribosome concentration per cell increases, eventually plateauing as cells reach maximum growth rate, the expected cost per protein decreases. This suggests a benefit to waiting to induce cells until protein cost is lowest, and indeed delayed induction has been seen to result in more protein than earlier induction, up to a point where cells have exhausted the resources to create more protein [28–31].

Burden can be measured several ways, as burden can manifest several ways. The most common indicator of burden is a reduction in growth (e.g. growth rate or final OD600) [17,27,32,33], mostly stemming from the sequestration of growth-related resources mentioned above. Resource sequestration can be seen
more directly, however, by looking for evidence of retroactivity [34,35]. Retroactivity is the indirect effect of one process on another by using a shared resource; for example, two mRNA transcripts using the shared resource of ribosomes. A “capacity monitor,” such as the production rate of a constitutively expressed fluorescent protein, can be used to measure the burden of heterologous expression [17]. What’s more, changes in capacity correlate with changes in growth rate. By comparing how much heterologous protein is produced to how much capacity is reduced, inefficiencies can be found.

A more indirect and less well-studied means of measuring burden is to look at the cell’s stress response. It has been repeatedly observed that protein overexpression can lead to a heat-shock response [36–39], likely due to nascent polypeptide chains of heterologous protein sequestering chaperones, such as DnaK, GroEL, and GroES, which then frees the normally chaperone-complexed heat-shock sigma factor $\sigma^{32}$ to mount a heat-shock response [39–42]. This heat-shock response results in, amongst other things, protease expression, which can reduce protein levels [43,44]. It is therefore possible to measure burden indirectly by monitoring heat-shock activity.

Lastly, burden can be observed indirectly by the occurrence of deleterious changes to heterologous DNA, e.g., plasmids. Because burden slows growth, cells with less burden have a fitness advantage. This allows cells with broken burden-imposing components to overtake the culture. Plasmid loss [45,46], recombination [47,48], and mutations [17,48] have all been observed in cells
expressing heterologous proteins, and are thus plausible and measurable proxies for burden.

1.1.4 | Programming Gene Expression of Metabolic Pathway Enzymes in E. coli

When engineering a metabolic pathway, the most critical choices are often those of how to express the heterologous enzymes necessary for the pathway to function. The choice, conscious or not, of promoter system, ribosome binding site, mRNA stability, copy number (and plasmid- versus chromosome-based), codon choice, time of induction, etc., all have a large effect on how many enzymes are present in cells, the stoichiometry of those enzymes, how quickly cells make the enzymes, and when cells make them. These outcomes largely determine the performance of the pathway, and so much attention must be paid to these choices. A few will be discussed here.

1.1.4.1 | Promoters

Promoters can be split into two classes: inducible and uninducible. The inducibility of a promoter is typically dependent on the presence of a transcription factor that targets that promoter. Inducer molecules bind the transcription factor and then prevent it from occluding RNA polymerase binding, allowing transcription to occur. Uninducible promoters are often constitutive, that is, always active and transcribing. But some promoters, such as those that require a non-standard sigma factor, are not always active. Sigma factors are the proteins that bind DNA and
recruit RNA polymerase. In growing cells, the most abundant sigma factor is $\sigma^{70}$, encoded by the rpoD gene. In cases of stress, such as heat-shock or starvation, alternative sigma factors, such as $\sigma^{32}$ (rpoH) or $\sigma^{38}$ (rpoS) become more abundant so as to express a suite of genes to respond to the stress. As the abundance of these sigma factors is very low during normal exponential growth, their affiliated promoters are effectively silenced until a stress condition arises.

An important aspect of the inducibility of a promoter is how it affects the timing of expression. As mentioned in Section 1.1.4, induction timing is an important consideration when optimizing a metabolic pathway, and can dramatically impact titers [49]. Inducible promoters have the advantage of offering complete control over induction timing; inducer can be added manually at any concentration at any point during culture growth. Inducer efflux may preclude precise control over induction duration [50–52], however, or necessitate additional measures to remove efflux apparatus. Though some promoters suffer from low levels of “leaky,” i.e., constitutive, expression, many are tightly controlled. The primary drawback of inducible promoters is the cost of inducer molecules on an industrial scale, which has led to research into alternative, autonomous, induction systems.

Stress-responsive promoters and quorum-sensing (QS)-based promoters (which effectively generate their own inducer, HSL) have garnered a great deal of interest. They have both been used to autonomously express pathway enzymes, and QS-based systems have been used to autonomously and simultaneously express pathway enzymes and downregulate or degrade enzyme from competing
pathways. While autonomous promoters offer the benefit of not requiring expensive inducers, they have their own drawbacks. QS-based promoters are tunable but require significant screening of variants to change the rate of HSL production and sensitivity to HSL [53]. Stress-responsive promoters are not tunable and thus require specific conditions before expression occurs. Being tuned to cellular needs, many stress promoters are weak [54] or many turn on when cells are unable to produce high levels of protein [55,56]. The more-or-less fixed induction timing of some stress promoters may be fortuitous, however. Many σ38-based promoters turn on in late exponential phase [57], which may coincide with the optimal time to induce expression [30], and can reach high levels of expression.

1.1.4.2 | Ribosome Binding Sites

Promoters generate the mRNA transcripts, but ribosomes still need to translate the mRNA into the actual enzymes of the metabolic pathway. The primary means of controlling translation is the sequence of the ribosome binding site (RBS), where the 16S rRNA of the ribosome binds the mRNA. The RBS sequence determines the strength of this interaction, which governs the translation initiation rate and thus the overall translation rate. While there are well-used RBSs with characterized relative strengths, sequence context can change the strength of an RBS, and fine-grained adjustments may be required. The RBS Calculator [58], first introduced in 2009, generates bespoke RBS sequences to achieve desired translation rates in a
The predictions it makes of RBS strength can differ significantly from expectations, however, requiring at least a small library of generated sequences to find one that matches requirements. For pathways with multiple genes, finding the “right” RBS for each gene requires full combinatorial [59], or design-of-experiment-based [60], sampling of “expression space,” i.e., the possible RBS strengths of all of the genes, to find the set of RBSs that leads to the highest titers.

1.1.4.3 | mRNA Transcript Stability

Ribosomes can only translate an mRNA for as long as its lifetime, often only a few minutes [61]. RNA is degraded through multi-step pathways that can begin as soon as a transcript has been synthesized by an RNA polymerase. Degradation of mRNA typically begins with a rate-limiting, RNase E-mediated phosphodiester bond cleavage of the mRNA. This cleavage is followed by subsequent rounds of 3' → 5' degradation [62], carried out in concert by the degradosome [63–66]. RNase E [67], an endoribonuclease, is thought to bind and process transcripts primarily via 5' entry at a monophosphorylated (5'-P) end. As transcripts are synthesized natively with 5'-PPP, it was hypothesized and then shown that 5'-P RNAs are created in cells through the removal of the gamma- and beta-phosphate from 5'-PPP RNAs [68] by RppH, an RNA pyrophosphohydrolase [69].

mRNA transcript stability is determined through the collective impact of a multitude of sequence and structural features. 5' terminus identity (i.e., 5'-PPP
versus 5’-P versus 5’-OH) and the presence of stable secondary structures within
the 5' untranslated region (UTR) affect 5' end accessibility by RppH and RNase E.
Active translation creates steric hindrance and ribosome occlusion that reduces
internal accessibility by RNase E. 3' end accessibility by PNPase varies according to
3' UTR secondary structure, polyadenylation state, and the presence or absence of
sRNAs that mediate degradation. Because RNAs can be transcribed and degraded
within the space of only a few minutes, variations in transcript stability can have
dramatic effects on RNA levels. This implies that gene expression can be controlled
quickly by modulating the sequence and structural features that directly affect
transcript stability. Several mechanisms to control transcript stability have been
studied, for instance, 5' and 3' UTR hairpins [70–74], 5' UTR cleavage [7], and
antisense RNA/small RNA binding [75–79].

1.1.4.4 | Dynamic Control of Gene Expression
Generally, choices of promoters, RBS, transcript stability, and other factors, result
in an open-loop system. Open-loop systems do not contain explicit feedback loops.
The system components must be tuned just-so in order to achieve the desired
result. In contrast to this are closed-loop systems, which do contain feedback
loops, and, therefore, sensors and actuators to implement the feedback.
Closed-loop systems in a metabolic engineering context are typically referred to as
dynamic control systems. The goal of most published dynamic control systems is to
balance pathway enzyme levels and/or to control levels of a toxic metabolite to
minimize its product toxicity. This is implemented with a sensor to detect a relevant metabolite and an actuator to downregulate a relevant enzyme. Commonly, the sensor and actuator are the same thing, namely, a transcription factor that can bind a metabolite and downregulate a promoter controlling enzyme expression.

Most examples of dynamic control systems have been deployed in fatty acid biosynthesis [32,80,81], as two transcription factors, FadR and FapR, have been found to respond to the fatty acid precursors fatty acyl-CoAs and malonyl-CoA, respectively. These control systems operate by putting pathway enzymes under control of these metabolite-regulated promoters. When the metabolite-of-interest is produced at a sufficient quantity, the promoters are expected to be up- or down-regulated in order to either balance enzyme levels or reduce product toxicity. While all of these papers report increases in titer with the introduction of control systems, levels of the regulated enzymes are not reported, making it impossible to know if the dynamic control systems operate as intended. All increases in titer could potentially be attributed to changing promoter strengths of these regulated promoters; that is, all strain comparisons have different promoters, and it is unclear if the strengths of these promoters are the cause of the titer improvement. Another dynamic control system, where a pinene-sensitive transcription factor controls expression of an efflux pump to expel pinene and increase tolerance, does not contain direct evidence that the efflux pump expression is being dynamically changed due to feedback regulation. Like the
aforementioned papers, it relies on promoter engineering and its results can potentially be explained by promoter strengths alone. Thus, more rigorous quantification of the relationship between the levels of negative regulator, its target, and the molecule being sensed is needed to better understand how these systems function. Indeed, the field is in need of a demonstration that dynamic control can function on the time-scale of a typical batch culture, and moreover that it can outperform static control by some metric. Furthermore, it is not clear that tuning the expression of dynamic control components is inherently any easier than tuning the expression of the regulated enzyme. In fact, it may be better to engineer static expression to reach a set-point than to overshoot and require dynamic control to reach a set-point.

The label of “dynamic control” can also apply to work that utilizes stress-responsive promoters to regulate pathway enzyme levels [82–91]. In these cases, there is typically no feedback, and hence they are not closed-loop systems. They therefore are implemented with different goals in mind. Typically, the benefit arises from the system creating a two-stage fermentation, meaning that induction is delayed until late exponential phase, allowing cells to grow without the burden of expressing pathway enzymes. The dynamic control often stems from the use of stress-responsive promoters: these promoter systems wait until some stressor is present, such as starvation or heat-shock, to produce enzymes. As with the aforementioned closed-loop systems, it is often unclear in these studies how much promoter strength contributes to the improvement seen. The comparisons are
insufficient to rule out the possibility that the autonomous promoter is better because it makes more protein than the manual induction control, and thus the dynamics are unimportant. In one study that did track protein levels, this seems to be the case [83].

Similarly, “dynamic control” often contains work that uses quorum-sensing (QS) components to autonomously turn on pathway enzyme expression. The goal is the same as with stress-responsive promoters, but the mechanism is slightly different. In this case, cells express an enzyme to make a small molecule, e.g., HSL, that can diffuse to other cells in the culture. An HSL-responsive transcription factor, e.g., LuxR, responds to HSL by not binding its associated promoter, allowing for pathway enzyme expression. Thus, as cells accumulate HSL, by both producing it and importing HSL from other cells, pathway expression increases. By controlling levels of the HSL-producing enzyme and the HSL-responsive transcription factor, the system can be tuned to turn on at different points during culture growth. Several groups have used QS components to autonomously express enzymes [91–96]. Other groups have gone a step further to not only express enzymes, but to also downregulate competing pathways in order to redirect flux to their pathway [53,87,97].

1.1.5 | Kinetic Models of Metabolic Pathway Gene Expression

Models are used in synthetic biology and metabolic engineering for many reasons. They can suggest hypotheses and predict experimental outcomes to guide
experimental efforts. This can reduce labor costs by reducing the number of experiments that need to be performed and eliminating dead-end lines of inquiry. They can also be used to evaluate current understanding of a system, by integrating current knowledge into a model and assess its predictive capacity.

In metabolic engineering, models are often used to evaluate flux using flux-balance analysis (FBA). FBA models are linear systems of reaction stoichiometries, the fluxes of which are optimized for growth. This enables predictions of how adding or removing reactions affects flux in a metabolic pathway. However, these models are steady-state and time-invariant and thus cannot capture dynamic processes, such as genetic feedback networks.

In synthetic biology, models are often used to study dynamic processes in genetic networks. The models tend to be comprised of systems of deterministic ordinary differential equations that describe how species such as proteins and mRNAs change over time. By solving these systems, the behavior of a complex genetic network over time can be evaluated by studying the behavior of its constituent species.

A crucial choice in modeling a system is determining the complexity of the model, i.e., how realistic versus how abstract. While realistic, or mechanistic, models can potentially better capture the real-world behavior of a system, they involve large numbers of species and reaction rates, and it may be impossible to find reliable reaction rates in many cases. Abstract models can provide a parsimonious reflection of system behavior, but it can be impossible to relate
model parameters to real-world biological processes. Thus abstract models and mechanistic models can serve different purposes. Abstract models can be used to understand system dynamics, while mechanistic models can be used for more fine-grained analysis of parameters that have a real-world counterpart. Mechanistic models that accurately reflect system behavior can thus be useful for informing engineering decisions. For instance, if simulations show that a behavior only occurs with transcription factors with a certain range of binding energies, this would lead researchers to using transcription factors with that characteristic in their system. It should be noted that many aspects of genetic systems are not easily tunable, so it may be difficult or impossible to do what a model suggests would improve a system.

For systems of differential equations that cannot be solved, parameter sampling can provide an accurate picture of the system's behavior. The parameter space can be constrained to physiologically relevant ranges, such as the span of reported measurements from the literature. This allows attention and computational resources to be focused only on the subset of model input-output space that is relevant for engineering.
1.2 | Dissertation Organization

This document is organized into five chapters. The first chapter (this chapter) is an overview of topics relevant to the subsequent chapters, intended to establish a foundation of literature and ideas that supports the work that was done.

The second chapter, “Computational Analysis of Genetic Control Systems for Metabolic Pathway Engineering,” is primarily a published research article on the subject of analyzing kinetic models of genetic feedback control systems acting on a metabolic pathway. The goal of this work was to evaluate the potential for RNA-based feedback control systems to regulate metabolic pathway enzyme levels to mitigate product toxicity and thereby improve titers.

The third chapter, “Programming Enzyme Expression Dynamics to Reduce Cellular Burden and Improve Expression and In Vivo Catalysis,” is a research article that examines the burden imposed by a particular enzyme, the origins of that burden, and how burden can be mitigated by using autonomous promoters and by programming expression dynamics. There is a novel focus on expression dynamics with respect to burden, which is enabled by using a fluorescent protein fusion of the enzyme under study. The paper suggests strategies for improving production from metabolic pathways.

The fourth chapter, “Production of 4-aminocinnamic Acid from Glucose in E. coli,” is work toward building a pathway for producing 4-aminocinnamic acid (4-ACA) in E. coli, which could serve in the future as a testbed pathway for studying both product toxicity and enzyme expression burden.
The fifth and last chapter consists of overall conclusions and a discussion.
Chapter 2 | Computational Analysis of Genetic Control Systems for Metabolic Pathway Engineering

Part of the work in this chapter (Sections 2.1-2.4) is adapted and reprinted with permission from:

Abstract
Engineered metabolic pathways can be augmented with dynamic regulatory controllers to increase production titers by minimizing toxicity and helping cells maintain homeostasis. We investigated the potential for dynamic RNA-based genetic control systems to increase production through simulation analysis of an engineered p-aminostyrene (p-AS) pathway in E. coli. To map the entire design space, we formulated 729 unique mechanistic models corresponding to all of the possible control topologies and mechanistic implementations in the system under study. 2,000 sampled simulations were performed for each of the 729 system designs to relate the potential effects of dynamic control to increases in p-AS production (total of $3 \times 10^6$ simulations). Our analysis indicates that dynamic control strategies employing aptazyme-regulated expression devices (aREDs) can yield >10-fold improvements over static control. We uncovered generalizable trends in successful control architectures and found that highly performing RNA-based
control systems are experimentally tractable. Analyzing the metabolic control state space to predict optimal genetic control strategies promises to enhance the design of metabolic pathways.

2.1 | Introduction

Metabolic pathways found in natural systems have evolved complex control circuitry that responds dynamically to chemical and physical inputs to actuate genetic regulatory functions, minimize the buildup of toxic intermediates, and maintain homeostasis [98]. Many pathways that are targets for metabolic engineering contain: cytotoxic intermediates; metabolites that are inadvertently lost from the cell; or enzymes or proteins whose overexpression depletes metabolic resources, generates toxic compounds, or otherwise have deleterious effects on cellular physiology [99]. Constructing chemically-responsive genetic systems to solve control problems in engineered metabolic pathways could therefore lead to optimized production titers [25] of industrially- and medically-relevant compounds in microbial hosts.

In principle, there are a number of regulatory mechanisms that could be applied to solve pathway control problems and enable high levels of production. In seminal work, Farmer and Liao engineered a dynamic controller comprised of a promoter and its associated transcriptional activator, which induces transcription in conditions of excess flux, reducing toxic metabolic intermediate (acetate) accumulation and improving final titers of lycopene [82]. More recently, a fatty
acid/acyl-CoA-inducible transcriptional repressor (TR), FadR, was used to program enzyme expression levels in response to high levels of fatty acyl-CoA, minimizing the buildup of toxic intermediates and increasing Fatty Acid Ethyl Ester (FAEE) biodiesel production [80]. In the latter case, simulation analysis identified a broad range of genetic control parameter values under which the dynamically-regulated system was expected to generate higher levels of FAEE than a comparable system of static controllers. Elsewhere, modeling has shown that dynamic control of efflux pump expression [22] can expand the functional design space by removing toxic intermediates or products from cells more effectively than pump expression under static control [100].

An RNA-based approach for constructing sensors (e.g., ligand-responsive aptamers) and actuators (e.g., catalytic ribozymes) to control mRNA transcript stability may obviate the need to rely on pre-existing genetic components and increase the speed and efficacy with which dynamic control systems can be constructed [Stevens and Carothers 2018]. In vitro selection [101] can be used to select aptamers that bind in physiological conditions to a variety of target ligands. Using these aptamers, catalytic ribozymes with activities that are responsive to desired metabolites (aptazymes) can be generated and further assembled into dynamic regulators (sensor-actuators) of gene expression in bacteria, yeast, and mammalian cells [7,102,103]. Aptazymes and other aptamer-based RNA components have been used as pathway biosensors to implement higher-order functions [104], for in vivo enzyme evolution [103], and for screening metabolite-producing strains
Building on this work, model-driven processes were developed for engineering aptazyme-regulated expression devices (aREDs) [7], to produce quantitatively-predictable genetic outputs in E. coli. In these devices, cis-aptazyme cleavage within the 5' UTR results in a programmable increase in mRNA half-life in response to a target ligand (Fig. SI.2.1). Because aREDs can be readily built from component parts separately generated and characterized, it may be possible to engineer and then combine multiple genetic devices to give predictable systems-level functions. If so, fundamental questions about the space of genetic control system architecture designs and the corresponding functions may then become accessible to experiment, including: 1) the effect of increasing the complexity of control systems, 2) the degree to which the desired architectures can actually be implemented using RNA-based genetic controls, and 3) whether formal processes can be developed for identifying design specifications.

To begin to address these long-term questions, and better understand the space of potential designs for RNA-based genetic control systems, we are engineering an RNA-regulated p-aminostyrene (p-AS) production pathway in E. coli (Fig. 2.2.1a). p-AS is an industrially-relevant vinyl aromatic monomer with properties amenable to advanced applications in photonics, photolithography, and biomedicine [106]. Cytotoxicity of intermediates in production pathways have made substituted styrenes, like p-AS, challenging to produce in microbes[8]. Production of p-AS in E. coli therefore offers an excellent test case for demonstrating the utility of genetic control systems. The pathway enzymes and metabolic products are
well-defined, and it provides control problems to address to optimize titers (Fig. 2.2.1, in blue). These problems include an intermediate, \( p \)-AF, leaking from cells [7]; another intermediate, \( p \)-ACA, exhibiting cytotoxicity (Fig. SI.2.2a); \( L \)-amino acid oxidases (LAAOs) potentially depleting the aromatic amino acid pool and creating hydrogen peroxide as a byproduct; and efflux pump overexpression (for \( p \)-ACA removal) also generating toxicity (Fig. SI.2.2b).

Although flux balance analysis [107] and metabolic control analysis [108] can be used to study metabolic pathways and control systems, results obtained with those methods can be difficult to formulate in terms of kinetic parameters for the genetic components needed to build the system. Elsewhere, the importance of developing new formalisms to relate network topology and kinetic parameters in engineered circuits has been noted [109]. To maximize engineering tractability, the challenge is to create approaches for mapping potential design space inputs to production outputs in terms of measurable and tunable kinetic parameters. These parameters can then be associated with specific genetic components required for building metabolic control systems.

Our approach is to overlay a dynamic control layer onto an engineered metabolic pathway and sample kinetic parameters to identify functional control architectures that improve production. Specifically, we automated the generation of coarse-grained mechanistic models for 728 unique control architectures. To sample the underlying design space and account for parameter uncertainty within each of the 728 control architectures, we computed the \( p \)-AS production yields
expected with dynamic control for 2,000 unique parameter instances, for a total of $1.5 \times 10^6$ solutions. The impact of dynamic control on p-AS production was evaluated by comparison to parameter-matched systems comprised only of static control functions (i.e., those that could be obtained with promoter [110], RBS [111], or ribozyme-regulated expression device [7] (rRED) tuning). We discovered control architectures that dramatically increase the likelihood of generating high p-AS production and show that the choice of mechanistic implementation may be important for effectively coordinating the stoichiometries of intermediates, enzymes, and efflux pumps. Moreover, we find that large improvements in p-AS yield are within reach of currently-achievable controller component specifications and expect that the approach described here will be useful for driving the design of functional metabolic production systems.
Figure 2.1. Engineered p-aminostyrene (p-AS) pathway and control system design space. A) The proposed pathway includes optimized modules for converting glucose to chorismate (steps 1-9) [112], along with the PapABC operon from Streptomyces venezuelae and an E. coli deaminase to convert chorismate to p-aminophenylalanine (p-AF, steps 10-13) [10]. p-AF is converted to p-aminocinnamic acid (p-ACA) via an L-amino acid oxidase (LAAO, step 14) and p-ACA export occurs via an efflux pump (step 15), after which it can be converted to p-AS (step 16). Blue and green dashed lines indicate sensing functions, responsive to metabolites, and orange dashed lines indicate actuation functions using aptazyme-regulated expression devices [7] (aREDs) as genetic controllers to integrate the sensing information. B) There are a total of 729 ways to control the system with four actuators and two sensing molecules (p-AF and p-ACA), which comprises the control system design space. To map the effect of control on p-AS production, unique models were created for all 728 dynamic architectures (the 729th
is all static control) and the outputs from each were simulated 2,000 times with Latin Hypercube-sampled parameter instances, along with 2,000 parameter-matched static control reference instances.

2.2 | Methods

Modeling

p-AS production systems were modeled as sets of deterministic ordinary differential equations (ODEs). All 728 architectures were automatically generated as PySCeS [113] v. 0.9.0 model files using custom Python scripts to write models in the PySCeS modeling language. Models were comprised of mechanistic differential equations described in Fig. 2.2 and S. I. Table 1. Incorporating the control problems into the model was achieved by adding a first-order loss reaction for p-AF and introducing a toxicity term into the model that placed a general fitness disadvantage on cells by slowing all non-degradation and dilution reaction rates [80]. The value of this toxicity term (see S. I. Table 1) was dynamically dependent on the concentration of p-ACA, LAAO, and efflux pump. Toxicity coefficient ranges were modeled so that p-ACA slows growth by half at ~10 mM (Fig. SI.2.2), and so that the efflux pump and LAAO slow growth when present in the thousands of molecules range.

Two full sets (one static and one dynamic) of model reactions for the best Core Topology 12 architecture are in the S. I. materials, along with four parameter sets for instances with relative yield performance ranked 1<sup>st</sup>, 50<sup>th</sup>, 1000<sup>th</sup>, and 1950<sup>th</sup>.
A data chart with final relative yield values and statistics is also provided.

For time-course simulations (Fig. 2.6 and Fig. SI.2.7), data were collected from simulations of the architecture (270) of Core Topology 12 with the highest median $Y_{rel}$, using the parameter set from the best instance (highest $Y_{rel}$) of that architecture (see Fig. 2.5). All parameters except the three aRED LARs were kept constant, and the aptazyme ligand-activation ratios (LARs) were all set to 1, 5, and 100 and the models re-evaluated.

**Simulation**

Models were simulated on the University of Washington Hyak Computing Cluster using custom software for set-up and analysis. Models were simulated to 48 h, the length of a typical batch reactor run, using the LSODA ODE [114] solver packaged with PySCeS v. 0.9.0.

**Yield Calculations**

Yield ($Y$) on glucose for an architecture instance was calculated by summing internal (int) and external (ext) p-ACA molecule counts ($N_{p-ACA,int}$ and $N_{p-ACA,ext}$) and dividing that by the total number of chorismate molecules ($N_{chor}$) produced: $Y = (N_{p-ACA,int} + N_{p-ACA,ext}) / N_{chor}$. Relative yield ($Y_{rel}$), for static or dynamic cases, is $Y$ divided by the median relative yield value of all static controllers: $Y_{rel} = Y / \text{Median}(Y_{static})$. Note that, for the purposes of this study, p-AS is used
interchangeably with p-ACA, as the conversion is extracellular.

**Parameter Sampling**

Parameters were sampled using the Latin Hypercube Sampling [115] method from uniform distributions centered around nominal values gathered from the literature. Latin Hypercube Sampling aims to approximate an output distribution with fewer inputs by dividing up input space and sampling within the divided regions. This makes it less likely that large regions of input space will go unsampled. Ranges for the sampling, and sources for nominal values, are given in S. I. Table 1. The sampled parameter sets were generated using the lhs method in the lhs package for R.

**Statistical Analysis**

Mann-Whitney testing was performed in R using the wilcox.test method. The Mann-Whitney test is a non-parametric statistical test with the null hypothesis that the two populations being tested are from the same distribution. All tests had a one-sided alternative hypothesis of one population being greater than the other. Two-sided Spearman Rank Correlations were performed in R using the cor.test method. Spearman Rank Correlation is a non-parametric statistical test that assesses the monotonicity of the relationship between two variables. One-way ANOVA was performed in R using the lm and anova methods. ANOVA is a statistical test that compares the differences between group means to determine the likelihood of data originating from the same distribution. The null hypothesis is that
the data are from the same distribution.

Experimental Measurements of Toxicity
An overnight culture of E. coli MG1655 was inoculated 1:100 into fresh MOPS EZ (Teknova) media and allowed to reach an OD600 of 0.1. 200 ul of culture was then placed into wells of a black 96-well plate with a clear bottom (Corning) and amounts of inducer or p-ACA (in DMSO) were added to cultures. Final DMSO concentration in media (vol/vol) was kept at or below 2% to avoid DMSO toxicity. The plate was then placed into a Tecan M1000 plate reader with orbital shaking at 37 C for 12 hours. OD600 measurements were taken every 10 minutes.

2.3 | Results and Discussion
We first formulated a genetic control framework for the p-AS pathway (Fig. 2.2.1) consisting of three control nodes, where a node is defined as a possible target of positive or negative dynamic regulation. These three control nodes were chosen to be genetic expression of 1) the Pap operon enzymes (PapA, PapB, and PapC) that convert chorismate to p-AF, controlling pathway flux; 2) the LAAO that converts p-AF to p-ACA, producing a toxic intermediate; and 3) the efflux pump that exports p-ACA from the cell. The total number of potential ways of regulating these three nodes, that is, positive, negative, or no dynamic regulation, comprise the control state space. (Note that there are 27 combinations, or core topologies, with the 27th being no dynamic regulation at any node.) Two different ligands were examined as
sensing metabolites, and four different genetic regulatory mechanisms were examined as potential means of implementing actuation, at each node (see Fig. 2.2 for mechanistic details). Altogether, the combinations of sensors and actuators lead to 729 distinct architectures for creating a control system (Fig. 2.2.1b).

While aREDs can be fashioned into positive or negative actuators, we reasoned that coupling dynamic aREDs with transcriptional repressors (TRs) or translation-repressing sRNAs could result in composite devices with enhanced functions compared to aREDs alone. Composite devices have not been built experimentally; however, the individual components have been used with success [7,78,116], and it is expected that composite devices are experimentally tractable. We thus investigated one positive actuator and three negative actuators, two of which are composite devices. As positive actuators, aREDs were employed where ligand-binding increases genetic output by extending mRNA half-life, increasing target mRNA levels relative to the basal level (aRED<sup>ACT</sup>, controller C1; Fig. SI.2.1). Controller C2 generates negative actuation because ligand-binding represses aptazyme cleavage activity that would otherwise increase mRNA half-life (aRED<sup>REP</sup>; Fig. 2.2, in red; Fig. SI.2.1). Controller C3 is a composite device for negative actuation where ligand-binding activates aptazyme cleavage which programs an increase in the half-life of an sRNA engineered to inhibit translation from a target mRNA (aRED<sup>ACT</sup>-sRNA; Fig. 2.2, in blue). Controller C4 is a composite device for negative actuation where ligand-binding activates aptazyme cleavage, increasing the mRNA half-life of a transcriptional repressor (aRED<sup>ACT</sup>-TR; Fig. 2.2, in orange).
Figure 2.2. Mechanistic model and genetic control mechanisms. Reactions in the red module correspond to aptazyme-regulated expression device (aRED) function. aREDS program genetic expression through ligand-responsive aptazyme cleavage in the 5' UTR that increases RNA half-life by selectively removing the binding target of RppH-mediated degradation [68] (Fig. SI.2.1). Reactions in the orange module correspond to transcriptional repressors (TR) where binding operator regions in target promoter sequences downregulates transcription initiation [117]. Reactions in the blue module describe a small RNA (sRNA) mechanism that represses translation by binding a complementary sequence of a target mRNA near the ribosome binding site or start codon and inhibiting ribosomal interactions [118]. The blue or orange modules can be combined with a red module (controller C1: aRED\textsuperscript{ACT} or C2: aRED\textsuperscript{REP}) to create an aRED\textsuperscript{ACT}-sRNA controller (C3) or an aRED\textsuperscript{ACT}-TR controller (C4). Small orange circles represent ligand; A and B represent chemicals. Rate constants are drawn beside arrows. S.I. Table 1 gives full descriptions, mathematical expressions, and ranges for the sampled rate constants.

We determined architecture performance by creating 728 unique coarse-grained biochemical models (Fig. 2.2), corresponding to each possible dynamic control architecture (see Methods). Each of these 728 models was evaluated 2,000 times with Latin-hypercube-sampled [115] parameter values constrained to biochemically plausible ranges (S. I. Table 1). Each of these 2,000 parameter sets is referred to as an instance of that architecture. Models were formulated to emphasize measurable and tunable kinetic design parameters. Sampling parameter values thus allows exploration of the design space, in the case of tunable parameters, and additionally accounts for parameter uncertainty within the cell in the case of non-tunable parameters. For each architecture instance, i.e., parameter set, the yield on glucose was computed (see Methods). Yields were then normalized to the median yield for static controllers to arrive at relative yield ($Y_{rel}$).
Yield comparisons are thus relative to static control, which contains no dynamic regulation, though parameters still vary by sampling.

To broadly ascertain how dynamic control can influence production, distributions of relative yields for all of the parameter sets were analyzed. The distribution of $Y_{rel}$ for the full set of dynamic control architecture instances shows a wide range of production levels relative to static control levels, a large portion of which is worse than static control (Fig. 2.3). However, 137 out of the 728 (~19%) architectures have median $Y_{rel}$ greater than one, and expected p-AS production from the top 5% of architectures is almost three times higher than that of comparable systems with only static control. While a subset of the static control parameter sets can produce results comparable to the best dynamic controllers, as evidenced by comparing the largest $Y_{rel}$ values for dynamic and static control (Fig. SI.2.3), the distributions of highly-performing controllers are shifted toward higher $Y_{rel}$. Taken together, these results show that dynamic control can greatly increase the likelihood of obtaining high p-AS production yields across the space of system architectures that we studied.
Figure 2.3. Relative production yields across the entire design space. Log$_{10}$ relative yields ($Y_{rel}$) are plotted for all static control parameter instances (top), all dynamic control parameter instances (middle), and all instances for architectures with median $Y_{rel}$ in the top 5% (i.e., $Y_{rel} > 2.11$). The vertical red line denotes median $Y_{rel}$ within a set. Black shading indicates the $Y_{rel}$ 95% confidence interval. As depicted at right, static control systems have no sensing and no actuation, while there are 728 possible combinations of sensing and actuation functions defining the set of dynamic control system architectures under study.

We then investigated whether dynamic control system performance can be understood at the level of core topology, or the set of actuator functions combined across the three nodes. All 728 dynamic control architectures were grouped according to the 26 distinct core topologies (Fig. SI.2.4) and we then evaluated the performance of the control architectures as a set. For example, Core Topology 1 (CT 1) consists of the 12 architectures corresponding to all the possible combinations of sensor targets and controllers that can lead to positive actuation at Node 1, negative actuation at Node 2, and no actuation at Node 3. Predicted p-AS
production varies dramatically between core topologies, as indicated by median $Y_{rel}$ values that span more than two orders of magnitude (Fig. 2.4). Notably, several of the core topologies have both high $Y_{rel}$ and only small variations in the midrange, or $1^{st}$ to $3^{rd}$ quartile, of the $Y_{rel}$ values, indicating that most of the architectures in those core topologies enable high levels of production (Fig. 2.4, blue points at bottom right). Combined, these results are consistent with the idea that there are regions of the dynamic control design space defined by the core topologies that reliably outperform systems with only static controllers.

Variation in performance among the architectures within a core topology can be almost as large as the variation between core topologies, where nearly 100-fold differences in $Y_{rel}$ at the midrange are observed in some cases (Fig. 2.4, upper points). For example, Architecture 270 (Arch 270) within CT 12 (Fig. 2.4, labeled) has the second-highest median $Y_{rel}$ (3.6) out of 728 architectures examined, even though the median relative yield for CT 12 itself is only 0.7. Thus, although classifying control system architectures according to core topology can help identify high- and low- performing system designs, core topology alone does not fully-determine production yields.
Figure 2.4. Performance of core topologies. Each circle corresponds to all of the architectures within a given core topology. Placement on the horizontal axis indicates the median relative yield ($Y_{rel}$) of all instances of the core topology. Placement on the vertical axis indicates the $\log_{10}$-difference of the 25th to 75th percentile range of architecture $Y_{rel}$ (i.e., the inter-quartile range). Color denotes the 95th percentile $Y_{rel}$ for architectures within the core topology. Core Topology 12 is identified by call-out (see text).

We can begin to understand how variation in expected yields within a core topology arises by examining architecture-level details. For instance, there are 24 different architectures in CT 12, all with aRED-mediated positive actuation at Nodes 1 and 3, and negative actuation at Node 2. However, $Y_{rel}$ for the best CT 12 architecture (270) is 20-fold higher than for the worst architecture (266) (Fig. 2.5). Arch 270 and Arch 266 employ different Node 2 actuation mechanisms and have very different sets of sensing metabolite-actuator linkages. As a result, significant differences in $p$-AS production across the sampled parameter spaces are observed.
(Mann–Whitney $W = 3.7 \times 10^5$, $P < 2.2 \times 10^{-16}$). One-way ANOVA shows that the type of controller employed for negative actuation at Node 2 has a significant impact on $p$-AS production within CT 12 ($F = 435.4$, $P < 2.2 \times 10^{-16}$). For CT 12 architectures with C2 controllers ($\text{aRED}^{\text{REP}}$), more than 22% (3,611 out of 16,000) of the simulated parameter instances gave high-performing production yields better than 95% of the static control system instances. In contrast, for CT 12 architectures with controllers C3 ($\text{aRED}^{\text{ACT}}$-sRNA) and C4 ($\text{aRED}^{\text{ACT}}$-TR) at Node 2, only 4.1% and 0.42% of the parameter instances, respectively, resulted in predicted $p$-AS levels at the 95th percentile of static control system production.

Interestingly, within the CT 12 architectures, Node 2 controllers with high levels of repression appear to have deleterious effects on production (Fig. SI.2.5). For C3 controllers, there is a negative correlation (Spearman rank correlation coefficient, $R_s = -0.38$, $P < 2.2 \times 10^{-16}$) between $p$-AS production and the sRNA-target mRNA association rate, the latter being a design variable directly increasing translation repression in the model. Similarly, for C4 controllers, tighter TR-DNA binding results in less $p$-AS production ($R_s = -0.71$, $P < 2.2 \times 10^{-16}$) across a range of TR $k_d$s similar to those reported for the well-known Lac repressor (LacI-DNA $k_d \sim 10^{-7}$ to $10^{-9}$ M [119]). Counter-intuitively, these data show that actuators with strong repression characteristics are not always well-suited to the problem of implementing negative feedback in this system. Instead, the choice of controller may be important for creating fine-grained modulation that can mitigate regulatory problems while maintaining high levels of flux through the pathway.
To confirm that the observed differences in predicted p-AS production are caused by the dynamic control layer, we simulated production for the 24 CT 12 architectures with deactivated aREDs. The design variable for aptazyme folding, needed for ligand-responsive cleavage, was set to zero and $Y_{rel}$ was re-computed for all the CT 12 architectures with the same 2,000 parameter instances as before (Fig. SI.2.6). Over the entire design space, $Y_{rel}$ values obtained with deactivated aptazymes are lower than for actuators with functional controllers, indicating that dynamic, ligand-responsive control is needed for highest performance (One-Way ANOVA $F = 1914$, $P < 2.2 \times 10^{-16}$). Similarly, for the best-performing CT 12 architecture (Arch 270), predicted p-AS production with functional aptazymes is significantly better than when the controllers are disabled (Mann-Whitney $W = 2.9 \times 10^6$, $P < 2.2 \times 10^{-16}$). These results suggest that aRED-mediated dynamic control is responsible for the improvements in production obtained relative to systems comprised only of static controllers.
Figure 2.5. Relative production yields for Core Topology 12 (CT 12). CT 12 comprises the entire set of architectures that effect positive actuation on Nodes 1 (PapABC operon) and 3 (Efflux Pump), and negative actuation on Node 2 (LAAO). Shown are log10 relative yields ($Y_{rel}$) alongside a schematic for all CT 12 instances (top), all instances for the CT 12 architecture (266) with the lowest median $Y_{rel}$ (middle), and all instances for the CT 12 architecture (270) with the highest median $Y_{rel}$ (bottom). The vertical red line denotes median $Y_{rel}$ within a set. Black shading indicates the $Y_{rel}$ 95% confidence interval.

Because aRED function is central to the connection between dynamic sensing and actuation, we examined the feasibility of engineering aptazymes that can meet component specifications needed for high-performing systems. As previously described [7], aptazyme ligand-activation ratio (LAR), or cleavage rate in the presence versus absence of ligand (i.e., $k_{obs+}/k_{obs-}$), has a significant impact on aRED genetic output. In fact, differences in LAR and transcript half-life have been shown to account for as much as 85% of the variation in relative ligand-responsive
expression [7]. p-AF-responsive aptazymes with LARs of ~5-6 have been successfully engineered [7], and aptazymes responsive to other molecules with LARs >100 have been characterized [120]. New methods that combine kinetic RNA folding design, in vitro selection [121,122] and in vivo screening [123] should enable the rapid generation of p-ACA aptazymes with similar LARs.

Notably, for aRED controllers with aptazyme LAR = 5 (see Methods), predicted p-AS production from the best-performing CT 12 architecture (270) is almost 6 times higher than for static controllers and 3 times higher than for controllers without ligand-responsive actuation (i.e., LAR = 1). When LAR = 100, predicted p-AS yields are 16 times greater than for systems with static control (Fig. 2.6; Fig. SI.2.7). As expected [124], positive feedback results in faster rise times to steady-state for genes expressed from Nodes 1 (pap operon) and 3 (efflux pump operon), with the slopes of the rise increasing 4-8 times as the LAR increases from 1 to 5 to 100. Negative actuation at Node 2 results in the downregulation of LAAO expression as a function of time, with the degree of repression increasing along with aptazyme LAR. Overall, we can see that the aRED-mediated controllers program outputs from each node in response to changing sensor inputs (p-ACA concentrations), with the net effect of lowering steady-state chorismate concentrations and increasing p-AS production. The fact that even aptazymes with modest LARs can enable substantial increases in p-AS production highlights the potential for engineering aRED-based systems to dramatically-improve metabolic pathway function.
Figure 2.6. Dependence of relationship between sensor input and node output on aptazyme ligand-activation ratio. A) The predicted level of sensor input (p-ACA) and outputs (i.e., folded proteins) for each of the three nodes in the CT 12 Arch 270 system are plotted, as a function of time, relative to parameter-matched systems of static controllers. B) The relative levels of initial input (chorismate) and final output (p-AS) are plotted as a function of time. Colors indicate the modeled aptazyme ligand-activation ratio or the cleavage rate in the presence versus the absence of ligand ($k_{\text{obs+}}/k_{\text{obs-}}$) specified for the aRED controllers in the system.
2.4 | Supplementary Information

Note: some supplementary information is omitted here (Fig. SI.2.4, Table SI.1, Chart SI.1, Equation CSVs, and Parameter CSVs) due to size and format constraints. Please see the published work for these materials.

![Diagram](image)

**Figure SI.2.1.** Pictorial representation of aRED function. Non-cleaved, that is, most, mRNA is terminated with a 5’-PPP, and is subject to RppH-dependent dephosphorylation to a 5’-P, which enables RNase E-mediated degradation. Aptazyme cleavage leads to a 5’-OH, which is not recognized by RppH. Degradation therefore occurs via a slower RppH-independent mechanism and half-life of the mRNA increases, leading to increased expression\(^1,2\). This is similar in function to the \(glmS\) ribozyme\(^3\), though with the opposite outcome. Aptazymes can be designed so that ligand binding leads to either higher or lower levels of cleavage than the basal rate. If ligand binding leads to higher levels of cleavage, then \(k_{\text{obs}+}/k_{\text{obs}-} > 1\) and this is
an aRED$^{\text{ACT}}$ device. If ligand binding leads to lower levels of cleavage, then $k_{\text{obs}+}/k_{\text{obs}-} < 1$ and this is an aRED$^{\text{REP}}$ device.

**Figure SI.2.2.** Toxicity data for p-ACA and AcrAB efflux pump. A) p-ACA inhibits E. coli MG1655 growth when added to the extracellular media at increasing concentrations. Ribbon bands around the lines indicate standard error from six experimental replicates. B) AcrAB efflux pump expression inhibits E. coli MG1655 growth when expressed highly. AcrA and AcrB genes were under control of a lacUV5 promoter and expression induced with IPTG.
Figure SI.2.3. Dynamic controller performance at the highest end of the relative yield distribution. A) Relative yield values for all architecture instances for dynamic control and static control were sorted and the top 1000 values were plotted against each other (note: these are not parameter matched). This illustrates that nearly all of the space of top relative yield values reachable by dynamic control are also reachable by static control. B) The highest ten relative yield values for an architecture instance are shown, with the relative yield value on the vertical axis and the relative yield ratio with respect to parameter-matched static control on the horizontal axis. Color denotes the architecture of the instance.
Figure SI.2.5. The relationship between binding parameters and negative actuator performance. Scatterplots are shown of A) relative yield values versus sRNA duplex formation rates for a single Core Topology 12 architecture, B) relative yield values versus TR-DNA kd for a single Core Topology 12 architecture. Trend lines were generated using the default generalized additive model method implemented using the stat_smooth function from ggplot2. The trend lines indicate that weaker binding characteristics of the negative regulators are consistent with higher performance of the control system.
Figure SI.2.6. Production yield distributions for models with non-folding, de-activated aptazymes. Shown are histogram plots of log10-scaled relative yields for all Core Topology 12 parameter sets, all parameter sets of Arch 266 of Core Topology 12, and all parameter sets of Arch 270 of Core Topology 12 (chosen for comparison to Fig. 2.5). A vertical red line denotes the median value for the set. Black coloring indicates the 95% confidence interval.
Figure SI.2.7. Relationships between sensor inputs and node outputs. A) Time-course data of species levels relative to the static control case for different performance levels. These data use parameter values from the highest relative yield instance of the Core Topology 12 implementation with the highest median (see Fig. 2.5). The coloring of the lines indicates the percent rank within the 2000 instances, as measured by relative yield. The first row shows the concentration over time of the aptazyme ligand for each control node (p-ACA in each case). The second row shows the relative (v. static) concentrations of the pertinent folded proteins (PapA, PapB, PapC, shown with different line types; LAAO; Efflux Pump, respectively) for each control node. B) The third row shows the relative levels of initial input, chorismate, and final output, p-AS, over time.

2.5 | Modular Modeling with PyMDC

Much of the code written and used in this paper was developed in an ad-hoc fashion. To create a more user-friendly codebase for future modeling efforts —
organized, scalable, documented — PyMDC was developed. PyMDC is written in Python and built on top of LibSBML. The principal idea of PyMDC is that, when creating large number of similar models, much of the models are repeated. These repeated elements, such as transcription, translation, regulation, are modules in PyMDC, and these modules can be connected by their inputs and outputs to create a model. If elements are commonly reused and/or always modeled the same way, this saves a lot of time in model generation and allows for easy creation of large combinatorial sets of models. Figure 2.7 shows the base modules of PyMDC. A model is written in Python then translated automatically into a PySCeS model for simulation. A CSV file of all parameters is created so that parameter values or ranges can be specified, and users can create a template to automate this process.

**Figure 2.7.** Base module hierarchy of PyMDC. Primitive modules, such as addSpecies, addReaction, and addLinearKineticLaw, are used to develop more complex modules, which can then be used to develop even more complex modules. By combining many modules, the buildProtein module adds transcription and translation reactions along with all necessary species and reactions and rates — all with one line of code.
2.6 | Conclusions and Future Work

Our work on mechanistic modeling of metabolic pathways and genetic feedback controllers led to the development of PyMDC, a tool for easily building combinatorial sets of models out of modular model components. This strategy was used for developing a combinatorial set of RNA-based genetic feedback controller models. We then simulated each model using 2,000 parameter sets to create output distributions that allowed us to assess how the genetic feedback controllers impacted pathway titers. Our conclusions indicate that mostly the positive feedback-based controllers outperform no-feedback controls and negative feedback is only beneficial with very weak negative feedback connections. These results suggest that, in most cases, product toxicity is not enough to prohibit the strategy of making more enzyme to improve titers. Furthermore, it suggests that product toxicity is not enough to encourage the strategy of reducing enzyme levels to reduce product toxicity levels. The accuracy of these conclusions is difficult to evaluate without experimental measurements of product levels inside of cells and associated toxicity.

The predictive power of a model is predicated on the accuracy of that model. It is thus vital to fully understand and identify any assumptions made when creating a model. Ideally, these assumptions would then be tested experimentally to ensure their validity. In this case, neither was done sufficiently. In particular, the inaccuracy of assumptions made regarding toxicity and growth detracts from the
utility of the results obtained. As expounded on in Chapter 3, the effects of enzyme expression toxicity and dynamics play a large role in cell growth, expression, and enzyme catalysis. None of these effects are accounted for in our model. The work in Chapter 3 led to an understanding that inducer concentration in a cell can change rapidly as cells work to expel the inducer, which we find is possibly exacerbated by expression burden. This was not accounted for in our model, nor was its capacity to potentially cause escape subpopulations with reduced enzyme production. Furthermore, we modeled culture growth as a continuous — instead of batch — culture, which hinders the ability of the model to predict cell population dynamics in a batch culture and can lead to a poor estimation of how time-dependent processes can function. It is therefore also important to understand the time dynamics of dynamic controllers and metabolic production. More experimental data is required to model the accumulation of metabolites inside of cells to concentrations that are able to be sensed. This will enable better modeling of the applicability of particular instantiations of dynamic control and predicting what metabolite sensitivities are required for control to be successful. The requirements for cells to accumulate metabolite and mount a useful dynamic feedback response — all before the culture reaches stationary phase — are still unknown.

Many of these shortcomings could not have been anticipated at the time of modeling, though it was clear that PAL2 toxicity was poorly understood. Instead of waiting for more experimental evidence of our assumptions, our model contains
inaccurate assumptions, which render the model powerless to make accurate predictions for the real-life system under study. This underscores the importance of experimentally validating crucial assumptions of a model before embarking on an effort to make actionable predictions. Going forward, if the system is to be re-modeled, the aforementioned missing elements, along with recent work on ribosome allocation and protein cost \([17,26,27,125]\), could be integrated into the model.

In addition to improving the model, experiments to narrow the ranges of sampled parameters, in particular the enzymatic rates and toxicity, would make statistical analysis easier by reducing the potential for parameters to be sampled in ranges that turn out to be unrealistic. When working with upwards of 100 parameters, even 2,000 samples is still a sparse sampling. Focusing the parameter space as much as possible will therefore improve sampling quality. When doing Partial Rank Correlation analysis, the enzymatic rates were always the most important with respect to titer, which is sensible given that the enzymes directly produce flux. Fixing these rates or severely narrowing the sampling range would facilitate observing effects of other parameters under study.

Additionally, the RNA-based genetic controllers that were studied in our simulation analysis have not been as tractable as anticipated. Indeed, our lab has not, as yet, seen any in vivo functionality from 4-AF 5’ UTR aptazymes (aREDs) after several design iterations (see Chapter 4 for more information). Furthermore, because aREDs require some level of basal mRNA expression and have limited
expected dynamic range (2–3-fold), their utility as a means to implement genetic feedback using a transcription factor or small RNA actuator is questionable, as that feedback would have to be always present at a basal level. It is therefore recommended that any further computational studies into dynamic feedback control eschew the use of aREDs in favor of alternative ligand-based control systems, such as riboswitches or ligand-activated CRISPR guide RNAs.

Given our recent work in burden and autonomous induction (see Chapter 3), one important question to answer is: In what context is dynamic feedback control useful? It seems possible that the only important dynamic control necessary is control of expression timing. If expression occurs at the optimal time, at the optimal rate, negative feedback may be superfluous and counterproductive. It might be tempting to see negative feedback as a means for finding the optimal rate of expression, but RBS screens may be easier than tuning a negative feedback regulator, which potentially involves both screening regulators and engineering their expression via RBS tuning. The question of the utility of dynamic control may be amenable to computational analysis, if the model contains the necessary physiological phenomena as discussed above.
Chapter 3 | Programming Enzyme Expression Dynamics to Reduce Cellular Burden and Improve Expression and In Vivo Catalysis

Abstract

Engineering metabolic pathways requires expressing large quantities of enzymes. This places a heavy burden on cells and can dramatically lower pathway productivity. How the dynamics of enzyme expression impact burden and its manifestations is not well understood, however. Here, we study burden as a function of enzyme expression dynamics using an Arabadopsis thaliana phenylalanine ammonia lyase (PAL) enzyme fused to superfolder GFP (PAL2-sfGFP), expressed in Escherichia coli. We show how the timing and level of manual induction impact PAL2-sfGFP expression burden. A high level of induction of PAL2-sfGFP can lead to as much as 80% less growth compared to uninduced cells. By controlling the timing of induction, we eliminate growth reductions due to burden while simultaneously improving enzyme expression. Interestingly, we observe that burden can manifest in unexpected ways not reflected in culture growth. PAL2-sfGFP burden can lead to bimodal levels of enzyme expression and reductions in enzyme-catalyzed conversion. Controlling both induction timing and level can mitigate the effects of both manifestations. We then show that stationary phase promoters can autonomously express PAL2-sfGFP at levels similar to
manually induced levels and, despite similar enzyme levels, generate more in vivo catalysis than manually induced cases — all without noticeable burden. Our work demonstrates the importance of considering burden as a dynamic phenomenon and considering how enzyme expression dynamics influence burden. Doing so can elucidate manifestations of burden and suggest strategies for mitigation in order to improve production from engineered metabolic pathways.

3.1 | Introduction

The pursuit of high production titers from engineered metabolic pathways requires cellular hosts to produce large quantities of enzymes. It’s been well established that expression of high levels of heterologous protein is deleterious to a growing cell culture, however [16,21,33,126–129]. Resources, such as ribosomes and ATP, that could be allocated to cell growth must instead be allocated to heterologous protein production [16]. Indeed, a model of ribosome allocation can be used to predict growth rate changes in response to protein overexpression [16], suggesting that ribosome allocation is a major contributor to the deleterious impacts of high levels of heterologous protein expression. Additionally, changing cellular conditions under the strain of high levels of protein expression in many cases triggers a stress response (e.g., heat-shock response, acid response) [33,37–39,83,130,131]. This can lead to proteolysis [132] and further reductions in culture growth. Understanding and mitigating the burden of high enzyme expression levels is thus critical for optimizing production from engineered metabolic pathways.
The notion of a two-stage production has been widely adopted to reduce the impact of expression on growth [28,133]. During the first stage, cells prioritize growth and produce little or no protein. During the second stage, cells prioritize protein production, by expressing protein and sometimes simultaneously down-regulating growth-related processes or competing pathways [134]. To implement a system under this two-stage framework, the dynamics of pathway enzyme production in the second stage need to be programmed along two axes: time and rate.

The time at which the cells shift stages from growth to production is a powerful point of control. Small changes in induction time can dramatically impact protein production [135,136] and pathway titers [49]. Many studies indicate that the best time to induce is in mid-to-late exponential phase [28,30,137,138], or even later [139]. A delayed induction is thought to be beneficial at least partly due to the accumulation of ribosomes and other expression infrastructure during exponential growth in new media [27,140]. A higher concentration of ribosomes and resources creates a lower protein cost, which is defined as the negative impact of expression on growth rate [27].

At a given time, switching from growth to production can be triggered either manually or autonomously. Manual induction of promoters offers a high level of control over induction timing and transcription rate, but numerous pitfalls (e.g., high inducer cost) limit its utility on an industrial scale [ref]. Alternative, autonomous induction systems have shown promise as a means to control protein
expression and improve pathway production without inducer molecules. Several groups have used quorum-sensing components to create dynamic “metabolic valves” that turn on at high cell densities to autonomously express heterologous pathway enzymes and re-direct flux away from growth toward enzyme production by degrading or downregulating growth-related genes [53,87,97]. Other groups have used quorum sensing components to autoinduce enzyme expression without explicitly altering flux [91–96]. A similar effect can be achieved by using promoters that require alternative sigma factors that are only expressed during stress situations or stationary phase, though many of these promoters suffer from low levels of expression [54]. For example, stress- and nutrient-sensitive promoters have been used to delay induction of pathway enzyme expression until cells near stationary phase [82–87], and other stationary phase promoters have been used to control protein expression in several studies [88–91]. A library of synthetic stationary phase promoters has also been characterized [57]. Additional means of autoinducing enzyme expression include autoinducing media compositions often used in conjunction with the T7 promoter system [31,141–143] and, more recently, SILEX [144].

When cells switch to the production stage, protein production occurs at some rate over time. Many tuning knobs exist for controlling protein production rate, including the promoter system and its induction level, plasmid copy number, the ribosome binding site (RBS), mRNA stability, and others [145]. Unfortunately, maximum strength on all counts is not a fruitful strategy. To wit, the phage T7
promoter-polymerase system often leads to severe complications [31,146], thought to stem from its high transcription rate [147]. Slow induction of the trc promoter was shown to reduce protease activity and increase levels of soluble protein compared to a quick, pulsed induction [148]. And recent work suggests that strong RBSs can lead to inefficient protein expression [17]. Thus, it seems there is a balance to be struck with protein production dynamics so as to not overload cellular capacity and thereby reduce burden. Much work remains to be done to enable predictably finding this balance.

Many studies track some measure of burden from protein expression. Often this measure is related to growth, e.g., final OD600 [32], growth rate [17,33], or time-course OD600 [17,27,33]. Other measures include expression capacity, i.e., the ability for cells to constitutively produce a protein [17]; cell proliferation in a competitive culture environment [149,150]; and mRNA levels of the rrsG ribosomal RNA [84], which may be responsive to levels of the stress alarmone ppGpp [151]. However, most studies neglect either the time or rate axis when measuring burden; that is, burden is not assessed as a function of both expression level and expression timing. And how autonomous induction can potentially mitigate burden is not well studied. Thus, the relationship between enzyme expression dynamics and enzyme expression burden remains poorly understood, as does how to use this relationship to engineer better autonomous induction.

Phenylalanine Ammonia-Lyase (PAL, EC 4.3.1.24) enzymes natively convert L-phenylalanine to trans-cinnamic acid [152]. One isozyme, PAL4 from A. thaliana,
has been shown to perform the same reaction using 4-amino-L-phenylalanine (4-AF) as substrate, producing 4-amino-cinnamic acid (4-ACA) as product [13], albeit at a much lower rate [14]. Elsewhere, PAL2 from A. thaliana was shown to have higher activity than other isozymes on L-phenylalanine [9,153,154]. We thus sought to study the use of PAL2 to convert 4-AF to 4-ACA, with the goal of ultimately producing an engineered metabolic pathway to produce 4-aminostyrene (4-AS) from glucose in E. coli [155]. Preliminary experiments with PAL2 showed that PAL2 expression imposed severe burden on E. coli cells (data not shown), suggesting it could be a model system for studying burden in E. coli.

We created a C-terminal fusion [156] of PAL2 to superfolder GFP (sfGFP) [157], PAL2-sfGFP, to enable protein expression tracking. PAL2-sfGFP was placed under the control of a Tn10 pA tet promoter (pTet) and the strong BBa_B0030 RBS from the pBbE2A BglBrick plasmid [158] to create plasmid pPAL2.1 (Figure 1a). As a point of comparison, we created pSFGFP.1 (Figure 1b), which consists of pTet controlling sfGFP expression, with a strong, designed RBS [58,159] in place of BBa_B0030.

In this work, we study enzyme expression dynamics and burden with A. thaliana PAL2, expressed using both manual and autonomous induction systems. We show how manual induction timing and increasing enzyme expression burden impact cell growth, enzyme accumulation, and in vivo enzyme activity. We elucidate a relationship between induction timing and bimodal enzyme expression levels in the population, wherein, presumed escape pressure due to enzyme
expression burden leads to a population of low-expressing cells. A delayed induction reduces the proportion of cells in this low-expressing population. We find that enzyme abundance does not always correlate with enzymatic activity, suggesting that the dynamics of enzyme expression plays an important role in producing enzyme that is active in vivo. We lastly show that stationary phase promoters can recapitulate and even exceed the best characteristics of manual induction, which we suggest can be attributed to improved expression dynamics compared to manual induction, and which supports their potential utility as a simple yet effective means to operate a “metabolic valve.” We hope this work serves to reinforce the importance of enzyme expression dynamics in metabolic engineering and furthers our understanding of how to achieve optimal dynamics.

3.2 | Materials and Methods

**Construct assembly**

All DNA constructs were assembled using Circular Polymerase Extension Cloning (CPEC) [160]. Primers were designed manually to have binding Tms and overlap Tms of at least 70 C and minimal homodimer ΔG. Linear fragments were amplified with a 50 uL PCR reaction using Q5 high-fidelity DNA polymerase (New England Biolabs). Fragments were then gel-extracted using Qiagen kit reagents. All fragments were assembled with a 20 uL CPEC reaction. The resulting CPEC mixture was used directly to transform competent cells.
**E. coli transformation**

Chemically competent cells are mixed 1:1 with 50 μL 1X KCM buffer and incubated on ice with at least 1 ng of miniprepped DNA or 5 μL of CPEC reaction mixture in a 14 mL polypropylene culture tube for 30 minutes. Cells are then heat-shocked for 20 seconds at 42°C. 200-400 μL of SOC is then added and cells are shaken at 200 rpm at 37°C for at least one hour. All of the mixture is then transferred to an LB-Miller plate with appropriate antibiotics, inverted, and stored at 37°C for at least 12-18 hours, or until colonies are clearly visible. Plates are then transferred to 4°C for storage. Plates are typically used within one week.

**E. coli strains used**

All cloning and assembly procedures were performed using DH10b (Invitrogen) cells. All PAL2-sfGFP-related experiments were performed using MG1655 (ΔrecA ΔendA) (Addgene) cells.

**Time-course measurement of PAL2-sfGFP expression**

An overnight culture is grown by selecting around five colonies with a pipette tip and adding to 1 ml of LB-Miller media with appropriate antibiotics, then left to grow overnight (typically 12-18 hours) at 37°C with shaking at 200 rpm. The following morning, overnight cultures are diluted 1:100 into 1 ml LB-Miller media with antibiotics in a 96-deep-well plate with round-bottom 2 ml wells (Axygen). A Breathe-Easier membrane (Diversified Biotech) is placed on top of the plate for
sterility and oxygen permeability. Cultures are grown at 37 C with shaking at 900 rpm. Cells are induced at indicated optical densities or times using indicated concentrations of anhydrotetracycline (Fisherbrand). For fluorescence and optical density measurements, 25 uL of culture is added to 100 uL 1x PBS, and briefly mixed by pipetting up and down, in a 96-well, black-side, clear-bottom plate (Corning). Bubbles are removed before measurements are taken. Measurements are performed with a Biotek Synergy HTX plate reader. Absorbance is measured at 600 nm. Fluorescence is measured with a 485/528 nm filter using a bottom read with a gain setting of 40. All operations involving the primary growing culture are performed near a lit bunsen burner to maintain sterility.

**Measurement of 4-AF to 4-ACA conversion**

Cultures are grown as described in the “Time-course measurement PAL2-sfGFP expression” section. At induction, 50 uL of 100 mM 4-amino-L-phenylalanine (4-AF) (Alfa Aesar, CAS 943-80-6) in Milli-Q H2O is added to cultures along with inducer. At 24-hours post-inoculation, 500 uL of culture is spun-down at 14k rpm for 2 minutes in 1.5 ml microcentrifuge tubes (USA Scientific). 300 uL of culture supernatant is then transferred to an Amicon 10 kDa centrifuge filter and centrifuged at 14k rpm for 20 min, or until enough volume has been filtered. 98 uL of culture with 2 uL 10% TFA (i.e., 0.2% TFA) is then used for HPLC analysis.
HPLC quantification of 4-ACA

The column used for all HPLC analysis is an Agilent ZORBAX Eclipse Plus Phenyl-Hexyl, 4.6 x 250 mm, with 5 uM particle size. HPLC is performed on an Agilent 1200 series. The program for 4-ACA detection is as follows. With Solution A of Milli-Q H2O + 0.2% TFA and Solution B of Methanol + 0.2% TFA: 100% A for 4 min, gradient to 70% B for 4 minutes, gradient to 95% B for 4 minutes, 95% B for 4 minutes, 100% A for 4 min.

Flow cytometry of PAL2-sfGFP expression

At indicated times, 1 uL of culture (already diluted 1:5 into 1x PBS for plate reader measurements) is added to 200 uL 1x PBS in a 96-well, black-side, clear-bottom plate (Corning). Flow cytometry is performed with a Milltenyi MACSQuant flow cytometer. 10,000 events are read with a low flow rate and medium mixing. FSC is set to log2, 550 nm. SSC is set to log2, 350 nm. B1 is set to log5, 350 nm. Triggering is set to SSC, with a threshold of 15 nm.

Data analysis

All data analysis was performed with R in RStudio, with ggplot2 for visualizations and ggcyto for cytometry data analysis. T-tests were performed with the t.test() method in R.
3.3 | Results

PAL2-sfGFP expression burden can be attenuated by a delayed induction

To assess the burden that PAL2-sfGFP and sfGFP alone impose on cells, pPAL2.1 and pSFGFP.1 were expressed in *E. coli* MG1655 (ΔrecA ΔendA) cells. Overnight cultures were diluted 1:100 and the resulting cultures were induced at nine distinct induction times (each culture being induced once) along the growth curve and grown in LB-Miller media at 37 C in 96-deep-well plates. To monitor both protein expression and growth, fluorescence and OD600 were measured at regular intervals using a plate reader. Flow cytometry was performed at 8 hours and 24 hours with all samples to measure per-cell fluorescence (Fig 3.1c).
We find that PAL2-sfGFP expression at all tested aTc induction levels can lead to a reduction, often severe, in growth, when observed at three hours following induction (Table SI.3.1). At three hours post-induction, the growth rate
(over those three hours) can be as low as 22% of that of an uninduced culture (Fig 3.2a). However, for medium and high induction levels, the reduction in growth is attenuated with an extra hour of induction delay, up to an induction time of 2.5 hours (Table SI.3.2). In contrast, sfGFP expression at the highest level appears to only cause a slight reduction in growth (Fig 3.2a), despite potentially more fluorescence (Fig 3.2b), indicating that the culture may be able to accumulate higher levels of sfGFP than PAL2-sfGFP before growth attenuation occurs. This is likely explained by the four-fold size disparity between the proteins (239 amino acids for sfGFP v 973 amino acids for PAL2-sfGFP), differences in rare codon abundance, or other reasons. The enzymatic activity of PAL2 does not noticeably contribute to burden: an enzymatically dead PAL2 mutant (S204A) [161] led to similar levels of expression and growth attenuation as the wild-type PAL2 (Table SI.3.3).

**Figure 3.2.** Measurements of burden and expression of PAL2-sfGFP and sfGFP. Induction levels: low (10 nM), medium (25 nM), high (100 nM). All fluorescence and rate values were normalized to a maximum of 100 for each panel. A) % uninduced (UI) growth rate at 3 hours post-induction, where wild-type is uninduced cells. A
lower growth rate indicates slower growth due to PAL2-sfGFP expression burden. A darker red color symbolizes a lower growth rate compared to wild-type. B) Per-cell fluorescence measurements using flow cytometry. A darker green color indicates higher fluorescence. Values are the average of 2-3 independent biological replicates.

**PAL2-sfGFP expression level can increase with a delayed induction**

Along with the decrease in growth attenuation with a delayed induction, per-cell mean enzyme expression levels, as measured by flow cytometry, can increase with an extra hour of induction time, up to an induction time of 3 hours, for medium- and high-level inductions (Fig 3.2b, Table SI.3.4). Lower expression levels at the latest induction times (Table SI.3.4) are expected to be due to a reduced nutrient pool for making protein as cells near stationary phase.

**PAL2-sfGFP expression can result in bimodal fluorescence levels**

Per-cell expression levels at 24 hours (Fig 3.3a) form a bimodal distribution in the cases of high and medium induction levels of PAL2-sfGFP. Distributions are unimodal for all sfGFP induction levels and the low induction level of PAL2-sfGFP. Comparing the 8-hour and 24-hour measurements (Fig 3.3b) shows that reductions in mean expression, for the high induction levels of PAL2-sfGFP, occur over the sixteen hours between measurements (Table SI.3.5). For the medium induction level, the proportion of cells in the low-fluorescence population does not change significantly from 8 to 24 hours for almost all induction times (Table SI.3.5),
suggesting that the low-fluorescence population arises more quickly for the medium induction level case.

A. Flow cytometry at 24 hours post-inoculation

B. Fluor$_{24h}$ $\div$ Fluor$_{6h}$

C. Fraction of cells in high-fluor. population
**Figure 3.3.** Flow cytometry measurements. A) 10,000 fluorescence intensity events grouped into 200 bins. The darker blue the bin color, the higher the number of events in that bin, and thus the proportion of cells with that level of fluorescence. Two distinct bands of dark blue indicate a bimodal population. B) $\frac{F_{\text{fluor}_{24h}}}{F_{\text{fluor}_{8h}}}$, mean fluorescence at 24 hours divided by mean fluorescence at 8 hours. This demonstrates the change in mean fluorescence from the 8 hour timepoint to the 24 hour timepoint. Redder color symbolizes more low-fluorescence cells at 24 hours than at 8 hours. C) Fraction of cells expressing pPAL2.1 (i.e., pTet : PAL2-sfGFP) in the high-fluorescence population at the indicated timepoint. A darker red color symbolizes more low-fluorescence cells. Values are the average of 2-3 independent biological replicates.

This reduction is especially dramatic (up to 69%) for the high induction level of PAL2-sfGFP, and the reduction diminishes with an extra hour delay in induction, for induction times between one and three hours (Table S1.3.6). The reduction in the aforementioned mean expression levels corresponds with a reduction in the fraction of cells present in the high-expression subpopulation (Fig 3.3c). Thus, we suggest that changes in population dynamics may be largely responsible for the increase in mean culture expression seen with delayed induction timing.

**Bimodal fluorescence levels are due to a non-genetic process**

The time-dependent nature of the bimodality observed in expression suggests two hypotheses. The first is that enzyme expression burden exerts mutational pressure on the population; thus, slow-growing enzyme-expressers are out-competed by cells with expression-reducing mutations that confer a fitness advantage. The second is that enzyme expression burden exerts a pressure to increase efflux and/or decrease influx of the aTc inducer molecule, potentially by overexpressing
AcrAB, which happens to some degree regardless of expression burden [52]. In both cases, there is an advantage to inducing later, which gives “cheater” cells fewer possible cells divisions, before nutrients are exhausted, with which to overtake the population.

To test the mutation hypothesis, we plated cells harboring pPAL2.1 that had been induced at 2 and 3.5 hours with a high-level aTc induction then grown for 24 hours. Flow cytometry data indicate that roughly 60% and 20% of cells should be in the low-fluorescence population for the 2- and 3.5-hour inductions, respectively (Fig 3b). Under the mutation hypothesis, we would expect 60% and 20% of cells to have genetic changes that confer a fitness advantage leading to faster growth in the presence of inducer. However, when ten colonies from each plate (i.e., the 2- and 3.5-hour inductions) were grown in the presence of high-level aTc, all cultures appeared to grow at similar rates that were not different from a typical 0-hour induction (data not shown). Furthermore, the 5’ UTR, a likely target for loss-of-function mutations [17], of all colonies were sequenced and no mutations were found. Thus, we reject the mutation hypothesis. The population dynamics observed are not due to a genetic change but instead a transient change.
**Figure 3.4.** Induction times relative to ininduced culture growth rate. The gray line is the mean of uninduced cultures with pPAL2.1. Points and the numbers above them indicate the time at which cultures were induced. The gray line thus shows the culture growth rate at the induction times used in this study. Values are the average of 2-3 independent biological replicates; error bars are standard deviations.

PAL2-sfGFP expression burden is minimized with an induction near stationary phase

By looking at the culture growth rates that correspond to induction times, we can start to understand how culture conditions influence burden. Cultures containing uninduced pPAL2.1 have a maximal growth rate in the 1.5-2.5 hour region (Fig 3.4), which is consistent with expectations for growth in LB [162]. As growth rate nears its maximum, i.e., during exponential phase, cellular ribosome concentration increases, then plateaus following maximal growth rate [27,140]. Accordingly, protein production cost is expected to decrease as growth rate increases, then plateau following maximal growth rate [27]. This would imply that maximal
production levels and rates should occur with induction times of 2 and 2.5 hours for the system under study. However, most maximums for sfGFP are at 3.5 hour inductions, and for PAL2-sfGFP are at 4 or 4.5 hour inductions (Fig 3.2). These times coincide with very late exponential or early stationary phase (Fig 3.4).

**PAL2-sfGFP in vivo catalysis can be improved with a delayed induction**

With the goal of improving production from metabolic pathways containing the PAL2 enzyme, we turned to measuring in vivo catalysis from cells expressing PAL2-sfGFP. Cells were grown and induced as described above, plus 5 mM 4-AF was added at the time of induction. Conversion of 4-AF to 4-ACA after 24 hours post-inoculation was measured using HPLC.

![Conversion Chart](chart.png)

**Figure 3.5.** In vivo conversion of 4-AF to 4-ACA at 24 hours post-inoculation. Comparison of conversion from medium and high induction levels of PAL2-sfGFP. Bluer color symbolizes higher conversion. A darker blue color symbolizes higher conversion. Color is on the same scale as the high pulsed induction level. Values are the average of 2-3 independent biological replicates.

Contrary to expectations, maximal conversion is likely seen not with a high induction level but with a medium induction level (Fig 3.5, Table SI.3.7). This is
despite significantly lower per-cell enzyme levels (Table SI.3.8) and potentially larger escape population fractions (Fig 3.3c) compared to the high-induction case. As with expression level and growth rate, there appears to be an improvement in conversion with induction time (Fig 3.5).

**Figure 3.6.** Comparison of tet and $\sigma$38 promoters. Induction levels: high (100 nM). All fluorescence and rate values were normalized to a maximum of 100 per panel. A) Mean per-cell fluorescence at 24 hours. A darker green color indicates high fluorescence. B) Mean fraction of cells in high-fluorescence population. A darker red color indicates a large fraction of low-producing cells. C) Mean conversion of 5 mM 4-AF to 4-ACA at 24 hours post-inoculation. A darker blue color indicates higher conversion. Values are the average of 2-3 independent biological replicates.

**Expression dynamics of stationary phase promoters can improve in vivo catalysis and reduce expression burden**

We lastly sought to develop an autonomous induction system that would perform as well as the best cases of manual induction with pTet. We hypothesized that this would require an induction time of around 4 hours and levels of expression in between the medium- and high-level manual inductions. An RpoS ($\sigma$38) promoter
consensus sequence [54,163] (SP1) and two of the highest-expressing synthetic promoters from the Miksch, et al. σ 38 promoter library [57] (SP2, SP3), paired with a strong, designed RBS, were used to drive expression of PAL2-sfGFP.

We find that per-cell expression levels are between the medium and high induction levels (Fig 3.6a, Tables SI.3.9 and SI.3.10), and nearly all cells are in the high-expression population (Fig 3.6b). Cells with the s38 promoters achieved higher levels (up to 50-80%) of in vivo conversion of 4-AF to 4-ACA compared to a manual, pulsed induction of the tet promoter (Fig 3.6c, Tables SI.3.11-13). We expect that much of the improvement in conversion is due to the expression level being in-between the medium and high induction levels. This result indicates that stationary phase promoters could potentially be used to drive high levels of enzyme expression in an autonomous fashion.

3.5 | Discussion

In this work, we examine the impact of expression dynamics on burden from overexpressing PAL2-sfGFP. We see, consistent with previous research, that delaying induction can lead to increased protein expression and faster growth. Prior research [27] suggests that much of this benefit arises from decreasing protein cost as cells reach maximal growth rate and ribosome concentration. Our work suggests that high levels of enzyme expression burden can introduce another factor, namely, a subpopulation that escapes the negative effects of burden by producing less enzyme. We see two populations: one with high levels of
PAL2-sfGFP expression and one with low levels of expression. The low-expression population appears to be enabled not by a genetic change but by a transient one.

We suspect this transient change is an upregulation of aTc efflux by the AcrAB pump [52], which leads to population splitting as follows. In the initial population of cells that are induced, there will be a distribution of AcrAB expression. Cells with more AcrAB will pump out more aTc and produce less PAL2-sfGFP. Less PAL2-sfGFP leads to faster growth and more expression capacity to express more AcrAB in response to aTc accumulation. With this positive feedback loop producing more AcrAB and less PAL2-sfGFP, this subpopulation will be able to grow much faster than the cells that had, initially, lower levels of AcrAB expression, and overtake the culture, leading to a split population. Our observation that a larger proportion of escape cells occurs with a medium-level induction compared to a high-level induction can be explained, counterintuitively, by the lower burden from a medium-level induction. The lower level of burden enables a larger fraction of cells to start the AcrAB expression positive feedback loop, and these cells will have less aTc to pump out compared to a high-level induction. With higher induction, there is more aTc to pump out and more burden; thus, the fraction that can escape decreases and escape takes longer.

The benefit of later induction thus appears to be two-fold, in this case. Firstly, protein cost is lower, due to higher ribosome concentration; secondly, there are fewer division cycles left for “cheater” cells to overtake the population. Given the explanation above, it seems possible that most protein accumulation occurs in
the cells present at the time of induction, as any new cells likely arise from cells that are growing quickly and not producing much protein. Therefore, maximizing the number of cells present at the time of induction would maximize culture protein accumulation, provided induction is done while there are sufficient nutritional resources for protein production.

While these observations are specific to aTc efflux and the pTet promoter system, inducer efflux has been seen with IPTG [50,51] and arabinose [51], suggesting that burden could impact these systems in a similar fashion. If indeed all manual induction systems suffer from the possibility of large inducer-effluxing subpopulations, strains with pump knockouts or autonomous induction systems would appear to be a solution to this problem, as they allow no inducer escape and obviate the need for an inducer, respectively. Our data shows that population enzyme expression bimodality does not occur with our autonomous induction systems, and that high levels of protein expression are achievable, on par with a manual induction with a tet promoter. Higher levels of protein expression are potentially achievable, but will depend on the limiting resource. In the cases of limiting nutrient availability or limiting $\sigma^{38}$ in the sigma factor pool, expressing additional $\sigma^{38}$ with a stationary phase promoter could lead to both earlier expression and a faster expression rate, as there would be more $\sigma^{38}$ to compete against $\sigma^{70}$ for RNA polymerases.

Our results show the highest levels of 4-AF to 4-ACA conversion from $\sigma^{38}$ promoters expressing PAL2-sfGFP. We see $\sigma^{38}$ promoters generating mean
per-cell PAL2-sfGFP levels at around 60-70% the level of the high-induction case, and roughly twice as much as the medium-induction case. With this level of PAL2-sfGFP expression, we see roughly 50% and 80% increases in conversion compared to the best medium- and high-induction cases, respectively. There are several interpretations of this data. Firstly, it is possible that the increase in PAL2-sfGFP compared to the medium-induction case is mostly properly folded protein than can convert 4-AF to 4-ACA, and thus explains the improvement in conversion. This would imply that being at 60-70% the level of PAL2-sfGFP from a high-level induction avoids the negative effects of overloading the cell, which leads to decreases in conversion and active enzyme in the high-induction case. It is also possible that the rate of enzyme expression in cells with a σ38 promoter construct is lower than in cells with a tet promoter, and that this lower rate allows cells to avoid overloading the cell’s capacity to create active enzyme. Lastly, it is possible that substrate is limiting, and that higher levels of burden lead to lower levels of substrate influx, and thus less substrate for PAL2 to convert. Further experiments are necessary to deconvolute the potential contributions from these mechanisms.

Our work shows that induced cells exhibit growth similar to an uninduced control when induction is sufficiently delayed. Protein cost and inducer efflux can explain why an earlier induction leads to reduced growth, but these effects are not eliminated with a later induction, though growth is indistinguishable from an uninduced control. This would lead one to believe that burden has been eliminated. It is more likely that, instead, using growth as a proxy for burden breaks down in
this case. While reduction in growth is important to mitigate, there are other manifestations of burden that may be important to account for.

Expression capacity is one alternative measure of burden, though it is unclear how much expression capacity is necessary in the context of a late-exponential-phase induction of metabolic pathway enzymes, where the end goal is to produce as much functional enzyme as possible. It may therefore be more profitable to measure not the cellular capacity for expressing enzyme, but rather the cellular capacity for creating enzymes that are functional. Our results seem to indicate that a high level of expression can lead to less functional enzyme than a lower level of expression. A real-time measure of misfolding or aggregation would allow for monitoring when cells are being pushed past their ability to create functional enzyme. This measure would allow for understanding the utility of modulating enzyme expression rate in order to stay within the cellular capacity to create functional enzyme.

For a culture of cells expressing metabolic pathway enzymes, there is an optimal function of enzyme expression rate over time. To understand what this function looks like, we need to study how modulating the timing of enzyme expression and the rate of enzyme expression impacts cell fitness and pathway performance. We hope our results will serve to inform these efforts and point toward fruitful further studies.

3.6 | Supplementary Information

Statistics Tables
<table>
<thead>
<tr>
<th>induction_time</th>
<th>induction_level</th>
<th>rate_induced</th>
<th>rate_uninduced</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>High</td>
<td>c(0.018, 0.007, 0.007)</td>
<td>c(0.059, 0.039, 0.054)</td>
<td>0.00410</td>
<td>-5.700</td>
</tr>
<tr>
<td>1</td>
<td>High</td>
<td>c(0.026, 0.022, 0.021)</td>
<td>c(0.111, 0.102, 0.116)</td>
<td>0.0037</td>
<td>-20.000</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>c(0.055, 0.057, 0.057)</td>
<td>c(0.153, 0.159, 0.135)</td>
<td>0.00280</td>
<td>-13.000</td>
</tr>
<tr>
<td>2.5</td>
<td>High</td>
<td>c(0.083, 0.087, 0.1)</td>
<td>c(0.156, 0.172, 0.149)</td>
<td>0.00086</td>
<td>-8.100</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>c(0.123, 0.129, 0.12)</td>
<td>c(0.143, 0.185, 0.131)</td>
<td>0.11000</td>
<td>-1.700</td>
</tr>
<tr>
<td>3.5</td>
<td>High</td>
<td>c(0.138, 0.161, 0.135)</td>
<td>c(0.131, 0.172, 0.117)</td>
<td>0.59000</td>
<td>0.250</td>
</tr>
<tr>
<td>4</td>
<td>High</td>
<td>c(0.152, 0.177, 0.136)</td>
<td>c(0.119, 0.165, 0.098)</td>
<td>0.84000</td>
<td>1.200</td>
</tr>
<tr>
<td>4.5</td>
<td>High</td>
<td>c(0.154, 0.199)</td>
<td>c(0.136, 0.189)</td>
<td>0.64000</td>
<td>0.400</td>
</tr>
<tr>
<td>5</td>
<td>High</td>
<td>c(0.143, 0.171, 0.099)</td>
<td>c(0.122, 0.158, 0.091)</td>
<td>0.68000</td>
<td>0.490</td>
</tr>
<tr>
<td>0</td>
<td>Medium</td>
<td>c(0.017, 0.016, 0.017)</td>
<td>c(0.048, 0.034, 0.052)</td>
<td>0.01800</td>
<td>-5.100</td>
</tr>
<tr>
<td>1</td>
<td>Medium</td>
<td>c(0.048, 0.052, 0.054)</td>
<td>c(0.103, 0.096, 0.117)</td>
<td>0.00430</td>
<td>-8.400</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>c(0.082, 0.068, 0.104)</td>
<td>c(0.143, 0.153, 0.136)</td>
<td>0.00820</td>
<td>-5.100</td>
</tr>
<tr>
<td>2.5</td>
<td>Medium</td>
<td>c(0.115, 0.113, 0.128)</td>
<td>c(0.144, 0.16, 0.146)</td>
<td>0.00530</td>
<td>-4.500</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>c(0.134, 0.148, 0.132)</td>
<td>c(0.133, 0.164, 0.135)</td>
<td>0.32000</td>
<td>-0.540</td>
</tr>
<tr>
<td>3.5</td>
<td>Medium</td>
<td>c(0.123, 0.162, 0.124)</td>
<td>c(0.128, 0.162, 0.126)</td>
<td>0.45000</td>
<td>-0.130</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>c(0.147, 0.158, 0.113)</td>
<td>c(0.126, 0.153, 0.097)</td>
<td>0.73000</td>
<td>0.660</td>
</tr>
<tr>
<td>4.5</td>
<td>Medium</td>
<td>c(0.14, 0.159)</td>
<td>c(0.136, 0.162)</td>
<td>0.51000</td>
<td>0.031</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>c(0.12, 0.139, 0.111)</td>
<td>c(0.11, 0.123, 0.094)</td>
<td>0.85000</td>
<td>1.200</td>
</tr>
<tr>
<td>0</td>
<td>Low</td>
<td>c(0.046, 0.033, 0.03)</td>
<td>c(0.054, 0.046, 0.051)</td>
<td>0.04300</td>
<td>-2.600</td>
</tr>
<tr>
<td>1</td>
<td>Low</td>
<td>c(0.099, 0.087)</td>
<td>c(0.135, 0.101)</td>
<td>0.18000</td>
<td>-1.400</td>
</tr>
<tr>
<td>2</td>
<td>Low</td>
<td>c(0.149, 0.129, 0.129)</td>
<td>c(0.154, 0.131, 0.135)</td>
<td>0.34000</td>
<td>-0.450</td>
</tr>
<tr>
<td>2.5</td>
<td>Low</td>
<td>c(0.153, 0.145, 0.135)</td>
<td>c(0.157, 0.141, 0.141)</td>
<td>0.40000</td>
<td>-0.270</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>c(0.154, 0.136, 0.136)</td>
<td>c(0.161, 0.134, 0.129)</td>
<td>0.52000</td>
<td>0.057</td>
</tr>
<tr>
<td>3.5</td>
<td>Low</td>
<td>c(0.133, 0.137, 0.124)</td>
<td>c(0.141, 0.133, 0.128)</td>
<td>0.32000</td>
<td>-0.490</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>c(0.12, 0.11, 0.116)</td>
<td>c(0.111, 0.114, 0.11)</td>
<td>0.83000</td>
<td>1.200</td>
</tr>
<tr>
<td>4.5</td>
<td>Low</td>
<td>c(0.097, 0.106)</td>
<td>c(0.108, 0.107)</td>
<td>0.20000</td>
<td>-1.300</td>
</tr>
<tr>
<td>5</td>
<td>Low</td>
<td>c(0.095, 0.101, 0.093)</td>
<td>c(0.092, 0.099, 0.087)</td>
<td>0.78000</td>
<td>0.870</td>
</tr>
</tbody>
</table>

**Table SI.3.1.** Comparisons of induced v uninduced growth rates at 3 hours post-induction. Shown are sets of induced and uninduced growth rates at each induction time for each induction level of PAL2-sfGFP from pTet. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that induced rate is lower than the uninduced rate.
<table>
<thead>
<tr>
<th>induction_level</th>
<th>comparison</th>
<th>first</th>
<th>second</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0 to 1</td>
<td>c(0.305, 0.179, 0.13)</td>
<td>c(0.234, 0.216, 0.181)</td>
<td>0.4600</td>
<td>-0.10</td>
</tr>
<tr>
<td>High</td>
<td>1 to 2</td>
<td>c(0.234, 0.216, 0.181)</td>
<td>c(0.359, 0.358, 0.422)</td>
<td>0.0020</td>
<td>-6.40</td>
</tr>
<tr>
<td>High</td>
<td>2 to 3</td>
<td>c(0.359, 0.358, 0.422)</td>
<td>c(0.86, 0.697, 0.916)</td>
<td>0.0071</td>
<td>-6.40</td>
</tr>
<tr>
<td>High</td>
<td>2.5 to 3.5</td>
<td>c(0.532, 0.506, 0.671)</td>
<td>c(1.05, 0.936, 1.15)</td>
<td>0.0023</td>
<td>-5.90</td>
</tr>
<tr>
<td>High</td>
<td>3 to 4</td>
<td>c(0.86, 0.697, 0.916)</td>
<td>c(1.28, 1.07, 1.39)</td>
<td>0.0130</td>
<td>-3.70</td>
</tr>
<tr>
<td>High</td>
<td>3.5 to 4.5</td>
<td>c(1.05, 0.936, 1.15)</td>
<td>c(1.13, 1.05)</td>
<td>0.2900</td>
<td>-0.61</td>
</tr>
<tr>
<td>High</td>
<td>4 to 5</td>
<td>c(1.28, 1.07, 1.39)</td>
<td>c(1.17, 1.08, 1.09)</td>
<td>0.8600</td>
<td>1.40</td>
</tr>
<tr>
<td>Medium</td>
<td>0 to 1</td>
<td>c(0.354, 0.471, 0.327)</td>
<td>c(0.466, 0.542, 0.462)</td>
<td>0.0620</td>
<td>-2.10</td>
</tr>
<tr>
<td>Medium</td>
<td>1 to 2</td>
<td>c(0.466, 0.542, 0.462)</td>
<td>c(0.573, 0.444, 0.765)</td>
<td>0.1900</td>
<td>-1.10</td>
</tr>
<tr>
<td>Medium</td>
<td>2 to 3</td>
<td>c(0.573, 0.444, 0.765)</td>
<td>c(1.01, 0.902, 0.978)</td>
<td>0.0230</td>
<td>-3.70</td>
</tr>
<tr>
<td>Medium</td>
<td>2.5 to 3.5</td>
<td>c(0.799, 0.706, 0.877)</td>
<td>c(0.961, 1, 0.984)</td>
<td>0.0280</td>
<td>-3.70</td>
</tr>
<tr>
<td>Medium</td>
<td>3 to 4</td>
<td>c(1.01, 0.902, 0.978)</td>
<td>c(1.17, 1.03, 1.16)</td>
<td>0.0270</td>
<td>-2.80</td>
</tr>
<tr>
<td>Medium</td>
<td>3.5 to 4.5</td>
<td>c(0.961, 1, 0.984)</td>
<td>c(1.03, 0.981)</td>
<td>0.2500</td>
<td>-0.88</td>
</tr>
<tr>
<td>Medium</td>
<td>4 to 5</td>
<td>c(1.17, 1.03, 1.16)</td>
<td>c(1.09, 1.13, 1.18)</td>
<td>0.4100</td>
<td>-0.26</td>
</tr>
<tr>
<td>Low</td>
<td>0 to 1</td>
<td>c(0.852, 0.717, 0.588)</td>
<td>c(0.733, 0.861)</td>
<td>0.2500</td>
<td>-0.78</td>
</tr>
<tr>
<td>Low</td>
<td>1 to 2</td>
<td>c(0.733, 0.861)</td>
<td>c(0.968, 0.985, 0.956)</td>
<td>0.1100</td>
<td>-2.70</td>
</tr>
<tr>
<td>Low</td>
<td>2 to 3</td>
<td>c(0.968, 0.985, 0.956)</td>
<td>c(0.957, 1.01, 1.05)</td>
<td>0.1600</td>
<td>-1.30</td>
</tr>
<tr>
<td>Low</td>
<td>2.5 to 3.5</td>
<td>c(0.975, 1.03, 0.957)</td>
<td>c(0.943, 1.03, 0.969)</td>
<td>0.5700</td>
<td>0.20</td>
</tr>
<tr>
<td>Low</td>
<td>3 to 4</td>
<td>c(0.957, 1.01, 1.05)</td>
<td>c(1.08, 0.965, 1.05)</td>
<td>0.2900</td>
<td>-0.59</td>
</tr>
<tr>
<td>Low</td>
<td>3.5 to 4.5</td>
<td>c(0.943, 1.03, 0.969)</td>
<td>c(0.898, 0.991)</td>
<td>0.7100</td>
<td>0.68</td>
</tr>
<tr>
<td>Low</td>
<td>4 to 5</td>
<td>c(1.08, 0.965, 1.05)</td>
<td>c(1.03, 1.02, 1.07)</td>
<td>0.4200</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

**Table SI.3.2.** Comparisons of differences in growth rate at induction times one hour apart. Shown are sets of growth rate differences (induced rate / uninduced rate) for two induction times, the first and second, of PAL2-sfGFP from pTet. For a comparison, the first (e.g., 0-hour) induction time set is compared to the second (e.g., 1-hour), to see how growth rate difference changes with induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the first growth rate difference is lower than the second.
**Table SI.3.3.** Comparisons of PAL2 v dead PAL2 differences in growth rate. Shown are sets of growth rate differences (induced rate / uninduced rate) for regular PAL2-sfGFP and a catalytically dead PAL2-sfGFP, PAL2mut-sfGFP, for each induction time and induction level. Comparisons between the two sets will determine if there is a significant difference between the growth rate differences for the two populations. P-values and t-values are obtained by performing a two-sided t-test.

<table>
<thead>
<tr>
<th>induction_time</th>
<th>induction_level</th>
<th>pTet : PAL2mut-sfGFP</th>
<th>pTet : PAL2-sfGFP</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>High</td>
<td>c(0.173, 0.17)</td>
<td>c(0.305, 0.179, 0.13)</td>
<td>0.780</td>
<td>-0.30</td>
</tr>
<tr>
<td>1</td>
<td>High</td>
<td>c(0.174, 0.214)</td>
<td>c(0.234, 0.216, 0.181)</td>
<td>0.850</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>c(0.424, 0.432)</td>
<td>c(0.359, 0.358, 0.422)</td>
<td>0.390</td>
<td>-0.94</td>
</tr>
<tr>
<td>2.5</td>
<td>High</td>
<td>c(0.611, 0.542)</td>
<td>c(0.532, 0.506, 0.671)</td>
<td>0.500</td>
<td>-0.73</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>c(0.962, 0.851)</td>
<td>c(0.86, 0.697, 0.916)</td>
<td>0.410</td>
<td>-0.87</td>
</tr>
<tr>
<td>3.5</td>
<td>High</td>
<td>c(1.13, 1.12)</td>
<td>c(1.05, 0.936, 1.15)</td>
<td>0.130</td>
<td>-1.70</td>
</tr>
<tr>
<td>4</td>
<td>High</td>
<td>c(1.29, 1.46)</td>
<td>c(1.28, 1.07, 1.39)</td>
<td>0.500</td>
<td>-0.71</td>
</tr>
<tr>
<td>4.5</td>
<td>High</td>
<td>c(1.31, 1.21)</td>
<td>c(1.13, 1.05)</td>
<td>0.700</td>
<td>-0.41</td>
</tr>
<tr>
<td>5</td>
<td>High</td>
<td>c(0.843, 1.07)</td>
<td>c(1.17, 1.08, 1.09)</td>
<td>0.054</td>
<td>2.80</td>
</tr>
<tr>
<td>0</td>
<td>Medium</td>
<td>c(0.558, 0.404)</td>
<td>c(0.354, 0.471, 0.327)</td>
<td>0.780</td>
<td>-0.30</td>
</tr>
<tr>
<td>1</td>
<td>Medium</td>
<td>c(0.436, 0.495)</td>
<td>c(0.466, 0.542, 0.462)</td>
<td>0.850</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>c(0.784, 0.746)</td>
<td>c(0.573, 0.444, 0.765)</td>
<td>0.390</td>
<td>-0.94</td>
</tr>
<tr>
<td>2.5</td>
<td>Medium</td>
<td>c(0.967, 0.993)</td>
<td>c(0.799, 0.706, 0.877)</td>
<td>0.500</td>
<td>-0.73</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>c(0.981, 0.971)</td>
<td>c(1.01, 0.902, 0.978)</td>
<td>0.410</td>
<td>-0.87</td>
</tr>
<tr>
<td>3.5</td>
<td>Medium</td>
<td>c(1.06, 1.02)</td>
<td>c(0.961, 1, 0.984)</td>
<td>0.130</td>
<td>-1.70</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>c(1.09, 1.17)</td>
<td>c(1.17, 1.03, 1.16)</td>
<td>0.500</td>
<td>-0.71</td>
</tr>
<tr>
<td>4.5</td>
<td>Medium</td>
<td>c(0.905, 0.938)</td>
<td>c(1.03, 0.981)</td>
<td>0.700</td>
<td>-0.41</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>c(1.04, 0.915)</td>
<td>c(1.09, 1.13, 1.18)</td>
<td>0.054</td>
<td>2.80</td>
</tr>
</tbody>
</table>
Table SI.3.4. Comparisons of mean RFU from high and medium induction levels at induction times one hour apart. Shown are sets of mean RFU from flow cytometry runs (comprised of 10,000 events) for two induction times, the first and second, of pTet : PAL2-sfGFP for each induction level and timepoint. For a comparison, the first (e.g., 0-hour) induction time set is compared to the second (e.g., 1-hour), to see how mean RFU changes with induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the first mean RFU is lower than the second.
<table>
<thead>
<tr>
<th>induction_level</th>
<th>induction_time</th>
<th>8 hours</th>
<th>24 hours</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0</td>
<td>c(0.90288, 0.87762, 0.9246)</td>
<td>c(0.27102, 0.2836, 0.24585)</td>
<td>0.0000025</td>
<td>36.00</td>
</tr>
<tr>
<td>High</td>
<td>1</td>
<td>c(0.91992, 0.8863, 0.92812)</td>
<td>c(0.30815, 0.26166, 0.30175)</td>
<td>0.000033</td>
<td>32.00</td>
</tr>
<tr>
<td>High</td>
<td>2</td>
<td>c(0.94318, 0.89177, 0.94968)</td>
<td>c(0.41751, 0.35223, 0.33234)</td>
<td>0.0000600</td>
<td>18.00</td>
</tr>
<tr>
<td>High</td>
<td>2.5</td>
<td>c(0.94723, 0.81717, 0.95448)</td>
<td>c(0.42506, 0.37613, 0.43705)</td>
<td>0.0016000</td>
<td>10.00</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>c(0.92891, 0.80819, 0.92902)</td>
<td>c(0.56972, 0.53549, 0.64104)</td>
<td>0.0023000</td>
<td>6.00</td>
</tr>
<tr>
<td>High</td>
<td>3.5</td>
<td>c(0.90407, 0.85968, 0.90655)</td>
<td>c(0.79912, 0.79541, 0.84825)</td>
<td>0.0100000</td>
<td>3.30</td>
</tr>
<tr>
<td>High</td>
<td>4</td>
<td>c(0.87201, 0.85981, 0.88201)</td>
<td>c(0.77893, 0.81595, 0.87146)</td>
<td>0.1000000</td>
<td>1.80</td>
</tr>
<tr>
<td>High</td>
<td>4.5</td>
<td>c(0.89863, 0.86312, 0.88373)</td>
<td>c(0.81705, 0.82502, 0.82215)</td>
<td>0.0120000</td>
<td>5.70</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>c(0.91454, 0.85382, 0.87991)</td>
<td>c(0.80132, 0.84531, 0.75089)</td>
<td>0.0300000</td>
<td>2.60</td>
</tr>
<tr>
<td>Medium</td>
<td>0</td>
<td>c(0.23722, 0.099721, 0.20259)</td>
<td>c(0.12818, 0.10569, 0.12489)</td>
<td>0.1400000</td>
<td>1.40</td>
</tr>
<tr>
<td>Medium</td>
<td>1</td>
<td>c(0.2079, 0.10717, 0.22221)</td>
<td>c(0.12916, 0.10435, 0.13805)</td>
<td>0.1300000</td>
<td>1.50</td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>c(0.19759, 0.13705, 0.24855)</td>
<td>c(0.13743, 0.13448, 0.16136)</td>
<td>0.1300000</td>
<td>1.50</td>
</tr>
<tr>
<td>Medium</td>
<td>2.5</td>
<td>c(0.31091, 0.18485, 0.30901)</td>
<td>c(0.20583, 0.19439, 0.19561)</td>
<td>0.1200000</td>
<td>1.70</td>
</tr>
<tr>
<td>Medium</td>
<td>3</td>
<td>c(0.35311, 0.24944, 0.42126)</td>
<td>c(0.27252, 0.25382, 0.26581)</td>
<td>0.1300000</td>
<td>1.50</td>
</tr>
<tr>
<td>Medium</td>
<td>3.5</td>
<td>c(0.45187, 0.45971, 0.5692)</td>
<td>c(0.3527, 0.43821, 0.37881)</td>
<td>0.0470000</td>
<td>2.30</td>
</tr>
<tr>
<td>Medium</td>
<td>4</td>
<td>c(0.70482, 0.67463, 0.83688)</td>
<td>c(0.5425, 0.60899, 0.63643)</td>
<td>0.0420000</td>
<td>2.50</td>
</tr>
<tr>
<td>Medium</td>
<td>4.5</td>
<td>c(0.84521, 0.47482, 0.78761)</td>
<td>c(0.64766, 0.38093, 0.59777)</td>
<td>0.1600000</td>
<td>1.10</td>
</tr>
<tr>
<td>Medium</td>
<td>5</td>
<td>c(0.77917, 0.19721, 0.73015)</td>
<td>c(0.61922, 0.1671, 0.57436)</td>
<td>0.3300000</td>
<td>0.49</td>
</tr>
</tbody>
</table>

**Table SI.3.5.** Comparisons of fraction of cells in the high-RFU population at 8 v 24 hours. Shown are sets of proportions of cells with an RFU (as determined by flow cytometry) from pTet : PAL2-sfGFP in the high-RFU population for each induction time and level at two timepoints, 8 hours and 24 hours. For a comparison, the 8-hour set is compared to the 24-hour set, to see how high-RFU proportion changes with time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the 8-hour proportion is lower than the 24-hour proportion.
<table>
<thead>
<tr>
<th>induction_level</th>
<th>timepoint</th>
<th>comparison</th>
<th>first</th>
<th>second</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>8 hours</td>
<td>0 to 1</td>
<td>c(0.9, 0.88, 0.92)</td>
<td>c(0.92, 0.89, 0.93)</td>
<td>0.230000</td>
<td>-0.80</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>1 to 2</td>
<td>c(0.92, 0.89, 0.93)</td>
<td>c(0.94, 0.89, 0.95)</td>
<td>0.290000</td>
<td>-0.60</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>2 to 3</td>
<td>c(0.94, 0.89, 0.95)</td>
<td>c(0.93, 0.81, 0.93)</td>
<td>0.760000</td>
<td>0.83</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>2.5 to 3.5</td>
<td>c(0.95, 0.82, 0.95)</td>
<td>c(0.9, 0.86, 0.91)</td>
<td>0.630000</td>
<td>0.36</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>3 to 4</td>
<td>c(0.93, 0.81, 0.93)</td>
<td>c(0.87, 0.86, 0.88)</td>
<td>0.670000</td>
<td>0.49</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>3.5 to 4.5</td>
<td>c(0.9, 0.86, 0.91)</td>
<td>c(0.9, 0.86, 0.88)</td>
<td>0.680000</td>
<td>0.52</td>
</tr>
<tr>
<td>High</td>
<td>8 hours</td>
<td>4 to 5</td>
<td>c(0.87, 0.86, 0.88)</td>
<td>c(0.91, 0.85, 0.88)</td>
<td>0.310000</td>
<td>-0.55</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>0 to 1</td>
<td>c(0.24, 0.1, 0.2)</td>
<td>c(0.21, 0.11, 0.22)</td>
<td>0.500000</td>
<td>0.00</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>1 to 2</td>
<td>c(0.21, 0.11, 0.22)</td>
<td>c(0.2, 0.14, 0.25)</td>
<td>0.370000</td>
<td>-0.35</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>2 to 3</td>
<td>c(0.2, 0.14, 0.25)</td>
<td>c(0.35, 0.25, 0.42)</td>
<td>0.041000</td>
<td>-2.40</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>2.5 to 3.5</td>
<td>c(0.31, 0.18, 0.31)</td>
<td>c(0.45, 0.46, 0.57)</td>
<td>0.008000</td>
<td>-3.90</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>3 to 4</td>
<td>c(0.35, 0.25, 0.42)</td>
<td>c(0.7, 0.67, 0.84)</td>
<td>0.002700</td>
<td>-5.50</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>3.5 to 4.5</td>
<td>c(0.45, 0.46, 0.57)</td>
<td>c(0.85, 0.47, 0.79)</td>
<td>0.110000</td>
<td>-1.70</td>
</tr>
<tr>
<td>Medium</td>
<td>8 hours</td>
<td>4 to 5</td>
<td>c(0.7, 0.67, 0.84)</td>
<td>c(0.78, 0.2, 0.73)</td>
<td>0.770000</td>
<td>0.86</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>0 to 1</td>
<td>c(0.27, 0.28, 0.25)</td>
<td>c(0.31, 0.26, 0.3)</td>
<td>0.140000</td>
<td>-1.30</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>1 to 2</td>
<td>c(0.31, 0.26, 0.3)</td>
<td>c(0.42, 0.35, 0.33)</td>
<td>0.044000</td>
<td>-2.50</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>2 to 3</td>
<td>c(0.42, 0.35, 0.33)</td>
<td>c(0.57, 0.54, 0.64)</td>
<td>0.002900</td>
<td>-5.40</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>2.5 to 3.5</td>
<td>c(0.43, 0.38, 0.44)</td>
<td>c(0.79, 0.8, 0.85)</td>
<td>0.000056</td>
<td>-15.00</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>3 to 4</td>
<td>c(0.57, 0.54, 0.64)</td>
<td>c(0.78, 0.82, 0.87)</td>
<td>0.001900</td>
<td>-6.10</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>3.5 to 4.5</td>
<td>c(0.79, 0.8, 0.85)</td>
<td>c(0.82, 0.83, 0.82)</td>
<td>0.320000</td>
<td>-0.53</td>
</tr>
<tr>
<td>High</td>
<td>24 hours</td>
<td>4 to 5</td>
<td>c(0.78, 0.82, 0.87)</td>
<td>c(0.8, 0.85, 0.75)</td>
<td>0.710000</td>
<td>0.60</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>0 to 1</td>
<td>c(0.13, 0.11, 0.12)</td>
<td>c(0.13, 0.1, 0.14)</td>
<td>0.410000</td>
<td>-0.25</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>1 to 2</td>
<td>c(0.13, 0.1, 0.14)</td>
<td>c(0.14, 0.13, 0.16)</td>
<td>0.130000</td>
<td>-1.30</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>2 to 3</td>
<td>c(0.14, 0.13, 0.16)</td>
<td>c(0.27, 0.25, 0.27)</td>
<td>0.000300</td>
<td>-11.00</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>2.5 to 3.5</td>
<td>c(0.21, 0.19, 0.2)</td>
<td>c(0.35, 0.44, 0.38)</td>
<td>0.007700</td>
<td>-7.00</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>3 to 4</td>
<td>c(0.27, 0.25, 0.27)</td>
<td>c(0.54, 0.61, 0.64)</td>
<td>0.002900</td>
<td>-11.00</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>3.5 to 4.5</td>
<td>c(0.35, 0.44, 0.38)</td>
<td>c(0.65, 0.38, 0.6)</td>
<td>0.099000</td>
<td>-1.80</td>
</tr>
<tr>
<td>Medium</td>
<td>24 hours</td>
<td>4 to 5</td>
<td>c(0.54, 0.61, 0.64)</td>
<td>c(0.62, 0.17, 0.57)</td>
<td>0.790000</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table SI.3.6. Comparisons of fraction of cells in the high-RFU population at induction times one hour apart. Shown are sets of proportions of cells with an RFU (as determined by flow cytometry) from pTet : PAL2-sfGFP in the high-RFU population for each induction level and timepoint at two induction times, first and second. For a comparison, the first (e.g., 0-hour) induction time set is compared to the second (e.g., 1-hour), to see how high-RFU proportion changes with induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the first set proportion is lower than the second set.
Table SI.3.7. Comparison of conversion from medium and high induction levels. Shown are fractional levels of conversion (where 0.01 = 1%) of 5 mM 4-AF to 4-ACA from medium and high induction levels of PAL2-sfGFP from pTet at each induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the high-level induction conversion is lower than the medium-level conversion.

<table>
<thead>
<tr>
<th>induction_time</th>
<th>High Induction</th>
<th>Medium Induction</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>c(0.000966, 0.0012)</td>
<td>c(0.00355, 0.0029)</td>
<td>0.03400</td>
<td>-6.20</td>
</tr>
<tr>
<td>1</td>
<td>c(0.00128, 0.00133)</td>
<td>c(0.00333, 0.00332)</td>
<td>0.00120</td>
<td>-74.00</td>
</tr>
<tr>
<td>2</td>
<td>c(0.00154, 0.0016)</td>
<td>c(0.0035, 0.00353)</td>
<td>0.00095</td>
<td>-58.00</td>
</tr>
<tr>
<td>2.5</td>
<td>c(0.00374, 0.00378)</td>
<td>c(0.00416, 0.00412)</td>
<td>0.00270</td>
<td>-13.00</td>
</tr>
<tr>
<td>3</td>
<td>c(0.00537, 0.00518)</td>
<td>c(0.00543, 0.00487)</td>
<td>0.63000</td>
<td>0.42</td>
</tr>
<tr>
<td>3.5</td>
<td>c(0.0047, 0.00473)</td>
<td>c(0.00784, 0.00675)</td>
<td>0.06600</td>
<td>-4.70</td>
</tr>
<tr>
<td>4</td>
<td>c(0.00499, 0.00447, 0.0038)</td>
<td>c(0.00513, 0.00515)</td>
<td>0.08600</td>
<td>-2.10</td>
</tr>
<tr>
<td>4.5</td>
<td>c(0.00568, 0.00638, 0.00652)</td>
<td>c(0.0041, 0.0045)</td>
<td>0.99000</td>
<td>5.80</td>
</tr>
<tr>
<td>5</td>
<td>c(0.00546, 0.00601, 0.00547)</td>
<td>c(0.0034, 0.00367)</td>
<td>1.00000</td>
<td>9.30</td>
</tr>
<tr>
<td>induction_time</td>
<td>timepoint</td>
<td>High</td>
<td>Medium</td>
<td>p_value</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>0</td>
<td>8 hours</td>
<td>c(14056, 12842, 14333)</td>
<td>c(2448, 12316, 1771.9)</td>
<td>0.0089000</td>
</tr>
<tr>
<td>1</td>
<td>8 hours</td>
<td>c(16385, 14103, 15157)</td>
<td>c(2343.4, 1245.2, 2121.9)</td>
<td>0.0070000</td>
</tr>
<tr>
<td>2</td>
<td>8 hours</td>
<td>c(19898, 17731, 19381)</td>
<td>c(2715.7, 1732.7, 3211.4)</td>
<td>0.0053000</td>
</tr>
<tr>
<td>2.5</td>
<td>8 hours</td>
<td>c(20459, 13019, 18748)</td>
<td>c(3729.4, 1663.3, 3051.9)</td>
<td>0.0072000</td>
</tr>
<tr>
<td>3</td>
<td>8 hours</td>
<td>c(20558, 14644, 21432)</td>
<td>c(3099.7, 1862.4, 3879.7)</td>
<td>0.0018000</td>
</tr>
<tr>
<td>3.5</td>
<td>8 hours</td>
<td>c(19879, 24112, 20425)</td>
<td>c(3415.1, 2770.4, 4649.1)</td>
<td>0.0000170</td>
</tr>
<tr>
<td>4</td>
<td>8 hours</td>
<td>c(18521, 16805, 18082)</td>
<td>c(5133, 3625.9, 7548.9)</td>
<td>0.0000006</td>
</tr>
<tr>
<td>4.5</td>
<td>8 hours</td>
<td>c(13325, 6460.4, 10648)</td>
<td>c(5900.6, 2147.3, 4421.8)</td>
<td>0.0016000</td>
</tr>
<tr>
<td>5</td>
<td>8 hours</td>
<td>c(8246, 4445.3, 6305.8)</td>
<td>c(3610, 1563.4, 3036.4)</td>
<td>0.0012000</td>
</tr>
<tr>
<td>0</td>
<td>24 hours</td>
<td>c(4426.4, 4213.7, 3962.3)</td>
<td>c(1516.1, 1105.7, 1231)</td>
<td>0.0089000</td>
</tr>
<tr>
<td>1</td>
<td>24 hours</td>
<td>c(5909.9, 4262.4, 4980.8)</td>
<td>c(1590.5, 1111.4, 1437.5)</td>
<td>0.0070000</td>
</tr>
<tr>
<td>2</td>
<td>24 hours</td>
<td>c(8409.2, 6219.2, 6234.6)</td>
<td>c(1965.7, 1492.7, 2092.2)</td>
<td>0.0053000</td>
</tr>
<tr>
<td>2.5</td>
<td>24 hours</td>
<td>c(7288.1, 5342.3, 7181.4)</td>
<td>c(2400.8, 1589.9, 1919)</td>
<td>0.0072000</td>
</tr>
<tr>
<td>3</td>
<td>24 hours</td>
<td>c(9958.9, 7927, 10888)</td>
<td>c(2362.2, 1706.6, 2360.9)</td>
<td>0.0018000</td>
</tr>
<tr>
<td>3.5</td>
<td>24 hours</td>
<td>c(14588, 18288, 17236)</td>
<td>c(2561.5, 2427.2, 2822)</td>
<td>0.0000170</td>
</tr>
<tr>
<td>4</td>
<td>24 hours</td>
<td>c(15870, 13410, 19679)</td>
<td>c(3473.5, 3103.5, 4514.2)</td>
<td>0.0000006</td>
</tr>
<tr>
<td>4.5</td>
<td>24 hours</td>
<td>c(10104, 5462.5, 9159.7)</td>
<td>c(3763.9, 1969.9, 3001.1)</td>
<td>0.0016000</td>
</tr>
<tr>
<td>5</td>
<td>24 hours</td>
<td>c(6503, 4170.9, 4732)</td>
<td>c(2827.1, 1464.5, 2435.1)</td>
<td>0.0012000</td>
</tr>
</tbody>
</table>

**Table SI.3.8.** Comparisons of mean RFU from medium and high induction levels. Shown are mean RFU levels, as measured by flow cytometry from 10,000 events, for medium and high induction levels of PAL2-sfGFP from pTet at each induction time and timepoint. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the medium-level induction RFU is lower than the high-level conversion.
Table SI.3.9. Comparisons of RFU from stationary phase promoters against high induction level. Shown are sets of mean RFU from flow cytometry runs (comprised of 10,000 events) for a high induction level of PAL2-sfGFP from pTet at the best induction time, and each stationary phase promoter. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the stationary phase mean RFU is lower than the high-level induction.

<table>
<thead>
<tr>
<th>induction_time</th>
<th>timepoint</th>
<th>pTet</th>
<th>pS38</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>8 hours</td>
<td>c(18521, 16805, 18082)</td>
<td>c(8238, 8762)</td>
<td>0.00038</td>
<td>16.0</td>
</tr>
<tr>
<td>SP2</td>
<td>8 hours</td>
<td>c(18521, 16805, 18082)</td>
<td>c(8984, 9424.8)</td>
<td>0.00061</td>
<td>15.0</td>
</tr>
<tr>
<td>SP3</td>
<td>8 hours</td>
<td>c(18521, 16805, 18082)</td>
<td>c(9441.5, 10394)</td>
<td>0.00100</td>
<td>11.0</td>
</tr>
<tr>
<td>SP1</td>
<td>24 hours</td>
<td>c(15870, 13410, 19679)</td>
<td>c(9584.6, 10756)</td>
<td>0.03400</td>
<td>3.2</td>
</tr>
<tr>
<td>SP2</td>
<td>24 hours</td>
<td>c(15870, 13410, 19679)</td>
<td>c(10755, 11139)</td>
<td>0.04800</td>
<td>2.9</td>
</tr>
<tr>
<td>SP3</td>
<td>24 hours</td>
<td>c(15870, 13410, 19679)</td>
<td>c(11399, 11786)</td>
<td>0.06000</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Table SI.3.10. Comparisons of RFU from stationary phase promoters against medium induction level. Shown are sets of mean RFU from flow cytometry runs (comprised of 10,000 events) for a medium induction level of PAL2-sfGFP from pTet at the best induction time at each timepoint, and each stationary phase promoter. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the medium-level mean RFU is lower than the stationary phase induction.

<table>
<thead>
<tr>
<th>induction_time</th>
<th>timepoint</th>
<th>pTet</th>
<th>pS38</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>8 hours</td>
<td>(5133, 3625.9, 7548.9)</td>
<td>(8238, 8762)</td>
<td>0.05500</td>
<td>-2.6</td>
</tr>
<tr>
<td>SP2</td>
<td>8 hours</td>
<td>(5133, 3625.9, 7548.9)</td>
<td>(8984, 9424.8)</td>
<td>0.03800</td>
<td>-3.2</td>
</tr>
<tr>
<td>SP3</td>
<td>8 hours</td>
<td>(5133, 3625.9, 7548.9)</td>
<td>(9441.5, 10394)</td>
<td>0.02300</td>
<td>-3.6</td>
</tr>
<tr>
<td>SP1</td>
<td>24 hours</td>
<td>(3473.5, 3103.5, 4514.2)</td>
<td>(9584.6, 10756)</td>
<td>0.00580</td>
<td>-9.0</td>
</tr>
<tr>
<td>SP2</td>
<td>24 hours</td>
<td>(3473.5, 3103.5, 4514.2)</td>
<td>(10755, 11139)</td>
<td>0.00052</td>
<td>-16.0</td>
</tr>
<tr>
<td>SP3</td>
<td>24 hours</td>
<td>(3473.5, 3103.5, 4514.2)</td>
<td>(11399, 11786)</td>
<td>0.00041</td>
<td>-17.0</td>
</tr>
<tr>
<td>induction_time</td>
<td>Medium Induction</td>
<td>SP1</td>
<td>p_value</td>
<td>t_value</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>c(0.00355, 0.0029)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.014</td>
<td>-5.3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>c(0.0033, 0.00332)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.017</td>
<td>-5.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>c(0.0035, 0.00353)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.017</td>
<td>-5.2</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>c(0.00416, 0.00412)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.020</td>
<td>-4.8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>c(0.00543, 0.00487)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.025</td>
<td>-4.0</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>c(0.00784, 0.00675)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.052</td>
<td>-2.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>c(0.00513, 0.00515)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.027</td>
<td>-4.1</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>c(0.0041, 0.0045)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.020</td>
<td>-4.6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>c(0.0034, 0.00367)</td>
<td>c(0.00956, 0.0101, 0.0143)</td>
<td>0.017</td>
<td>-5.2</td>
<td></td>
</tr>
</tbody>
</table>

Table SI.3.11. Comparison of conversion from medium induction level and stationary phase promoter SP1. Shown are fractional levels of conversion (where 0.01 = 1%) of 5 mM 4-AF to 4-ACA from a medium induction level of PAL2-sfGFP from pTet and the SP1 stationary phase promoter at each induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the medium induction level conversion is lower than the SP1 conversion.
<table>
<thead>
<tr>
<th>induction_time</th>
<th>Medium Induction</th>
<th>SP2</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>c(0.00355, 0.0029)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0015</td>
<td>-9.3</td>
</tr>
<tr>
<td>1</td>
<td>c(0.0033, 0.00332)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0044</td>
<td>-11.0</td>
</tr>
<tr>
<td>2</td>
<td>c(0.0035, 0.00353)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0047</td>
<td>-10.0</td>
</tr>
<tr>
<td>2.5</td>
<td>c(0.00416, 0.00412)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0058</td>
<td>-9.1</td>
</tr>
<tr>
<td>3</td>
<td>c(0.00543, 0.00487)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0045</td>
<td>-6.6</td>
</tr>
<tr>
<td>3.5</td>
<td>c(0.00784, 0.00675)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0470</td>
<td>-2.5</td>
</tr>
<tr>
<td>4</td>
<td>c(0.00513, 0.00515)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0089</td>
<td>-7.4</td>
</tr>
<tr>
<td>4.5</td>
<td>c(0.0041, 0.0045)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0036</td>
<td>-8.4</td>
</tr>
<tr>
<td>5</td>
<td>c(0.0034, 0.00367)</td>
<td>c(0.00861, 0.00884, 0.0104)</td>
<td>0.0035</td>
<td>-9.9</td>
</tr>
</tbody>
</table>

Table SI.3.12. Comparison of conversion from medium induction level and stationary phase promoter SP2. Shown are fractional levels of conversion (where 0.01 = 1%) of 5 mM 4-AF to 4-ACA from a medium induction level of PAL2-sfGFP from pTet and the SP2 stationary phase promoter at each induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the medium induction level conversion is lower than the SP2 conversion.
<table>
<thead>
<tr>
<th>induction_time</th>
<th>Medium Induction</th>
<th>SP3</th>
<th>p_value</th>
<th>t_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>c(0.00355, 0.0029)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0170</td>
<td>-16.0</td>
</tr>
<tr>
<td>1</td>
<td>c(0.0033, 0.00332)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0025</td>
<td>-95.0</td>
</tr>
<tr>
<td>2</td>
<td>c(0.0035, 0.00353)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0019</td>
<td>-89.0</td>
</tr>
<tr>
<td>2.5</td>
<td>c(0.00416, 0.00412)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0014</td>
<td>-76.0</td>
</tr>
<tr>
<td>3</td>
<td>c(0.00543, 0.00487)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0220</td>
<td>-12.0</td>
</tr>
<tr>
<td>3.5</td>
<td>c(0.00784, 0.00675)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.1200</td>
<td>-2.4</td>
</tr>
<tr>
<td>4</td>
<td>c(0.00513, 0.00515)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0040</td>
<td>-62.0</td>
</tr>
<tr>
<td>4.5</td>
<td>c(0.0041, 0.0045)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0100</td>
<td>-21.0</td>
</tr>
<tr>
<td>5</td>
<td>c(0.0034, 0.00367)</td>
<td>c(0.00867, 0.00856)</td>
<td>0.0032</td>
<td>-35.0</td>
</tr>
</tbody>
</table>

Table SI.3.13. Comparison of conversion from medium induction level and stationary phase promoter SP3. Shown are fractional levels of conversion (where 0.01 = 1%) of 5 mM 4-AF to 4-ACA from a medium induction level of PAL2-sfGFP from pTet and the SP3 stationary phase promoter at each induction time. P-values and t-values are obtained by performing a one-sided t-test with the alternative hypothesis that the medium induction level conversion is lower than the SP3 conversion.
Chapter 4 | Production of 4-aminocinnamic Acid from Glucose with E. coli

Abstract

This chapter describes work toward developing a biosynthetic pathway for producing 4-aminocinnamic acid (4-ACA) in E. coli. To do so, we developed a biosynthetic pathway for producing 4-aminophenylalanine (4-AF), the immediate upstream precursor to 4-ACA, and HPLC methods for 4-AF detection. Titers in the range of 100 mg/L were achieved in test-tube cultures, and suggestions are provided to increase titers further and extend the pathway to produce 4-ACA. Also discussed are attempts to use 4-AF-responsive aptazymes in the 5’ mRNA untranslated region (UTR) to dynamically control gene expression, and the use of in vitro RNA-based biosensors to detect 4-AF concentrations in E. coli culture supernatants. The chapter ends with a discussion of lessons from this work and future directions.

4.1 | Introduction

4-aminocinnamic acid (4-ACA) is a high-value aromatic compound amenable for use in advanced polymers [13]. 4-ACA has not been produced from glucose in microbes, though it has been produced by addition of 4-amino-phenylalanine (4-AF) to resting E. coli cells expressing AtPAL4, a phenylalanine ammonia lyase
enzyme from *A. thaliana* [13]. 4-ACA is a presumed to be a challenging compound to produce at high levels in microbes due to its toxicity (see Fig SI.2.2a). The toxicity makes this pathway an attractive testbed for studying toxicity mitigation as a means to improve pathway production. Furthermore, our lab has developed ligand-responsive self-cleaving RNA ribozymes, or aptazymes, that respond to 4-AF, the precursor to 4-ACA, that could be used to regulate pathway gene expression to mitigate toxicity. Hence, this pathway is of great interest to our lab.
Figure 4.1. Biosynthesis pathway for chorismate from D-erythrose-4-phosphate (E4P). Enzymes catalyzing the shown reactions are indicated next to reaction arrows. Created using the BioCyc Pathway Collage tool.

The pathway to produce 4-ACA from glucose is comprised of 19 or 20 enzymes, most of which are part of the native E. coli pathway to produce aromatic amino acids. Two pathways, comprised of eight or seven genes, convert glucose to phosphoenolpyruvate (PEP), as part of glycolysis, and glucose to erythrose-4-phosphate (E4P), as part of the pentose-phosphate pathway. Either
PEP or E4P can be used to create shikimate with four enzymes, and chorismate with an additional three (Fig. 4.1). A further three enzymes, the PapABC operon (PapA, PapB, and PapC) can convert chorismate to 4-aminophenyl pyruvate (4-APPA) (Figure 4.2). Functional PapABC operons have been taken from the genomes of *Streptomyces venezuelae* [10,11,13] and *Pseudomonas fluorescens* [12]. A transaminase enzyme from *E. coli*, likely TyrB [10,164], then converts 4-APPA to 4-aminophenylalanine (4-AF). A final enzyme, PAL, can convert 4-AF to 4-ACA (Fig. 4.2).

Attempts to overproduce 4-AF have focused on increasing flux through the aromatic amino acid pathway and reducing flux to the amino acids themselves. Much of this can be accomplished through the use of the strain NST37 [5], which was developed to overproduce L-phenylalanine. It does this by nonfunctionalizing TyrR, which eliminates negative feedback on AroG and AroL; deleting TrpE and TyrA, which prevents tryptophan and tyrosine biosynthesis; and overexpressing aroF, aroG, and pheA, which increases flux to chorismate and then L-phenylalanine. By deleting pheA [12], much of the metabolic pull on chorismate is eliminated and can be redirected to 4-AF production.
Figure 4.2. Biosynthesis of 4-ACA from chorismate. Five-gene biosynthesis pathway for 4-ACA (at right) from chorismate (at left), by way of 4-AF (middle). PapA, PapB, and PapC comprise the PapABC operon. PAL2 is the specific phenylalanine ammonia lyase under study in Chapter 3.

4.2 | Materials and Methods

Construct assembly

All DNA constructs were assembled using Circular Polymerase Extension Cloning (CPEC) [160]. Primers were designed manually to have binding Tms and overlap Tms of at least 70 C and minimal homodimer ΔG. Linear fragments were amplified with a 50 uL PCR reaction using Q5 high-fidelity DNA polymerase (New England Biolabs). Fragments were then gel-extracted using Qiagen kit reagents. All fragments were assembled with a 20 uL CPEC reaction. The resulting CPEC mixture was used directly to transform competent cells.

E. coli transformation

Chemically competent cells are mixed 1:1 with 50 uL 1X KCM buffer and incubated on ice with at least 1 ng of miniprepped DNA or 5 uL of CPEC reaction mixture in a 14 mL polypropylene culture tube for 30 minutes. Cells are then heat-shocked for
20 seconds at 42 C. 200-400 uL of SOC is then added and cells are shaken at 200 rpm at 37 C for at least one hour. All of the mixture is then transferred to an LB-Miller plate with appropriate antibiotics, inverted, and stored at 37 C for at least 12-18 hours, or until colonies are clearly visible. Plates are then transferred to 4 C for storage. Plates are typically used within one week.

E. coli strains used

All cloning and assembly procedures were performed using either DH10b (Invitrogen) or 10 β (New England Biolabs) cells. 4-AF production experiments were performed using either MG1655 (ΔrecA ΔendA) (Addgene), NST37 (ATCC), or DH10b (Invitrogen) cells. Aptazyme experiments were performed using T7 Express (New England Biolabs) cells.

Shikimate production in E. coli

An overnight culture is grown by selecting around five colonies with a pipette tip and adding to 1 ml of LB-Miller media with appropriate antibiotics, then left to grow overnight (typically 12-18 hours) at 37 C with shaking at 200 rpm. The following morning, overnight cultures are diluted 1:100 into 2 ml MOPS EZ-Rich defined media (Teknova). At an OD600 of around 0.4 (as measured using 150 uL culture in 300 uL well in plate reader), cultures are induced with indicated concentrations of IPTG. Cultures are grown at 37 C with shaking at 200 rpm. At 24-hours post-inoculation, 500 uL of culture is spun-down at 14k rpm for 2 minutes in 1.5 ml
microcentrifuge tubes (USA Scientific). 300 uL of culture supernatant is then transferred to an Amicon 10 kDa centrifuge filter and centrifuged at 14k rpm for 20 min, or until enough volume has been filtered. 99 uL of culture with 1 uL 10% TFA (i.e., 0.1% TFA) is then used for HPLC analysis.

4-AF production in E. coli

An overnight culture is grown by selecting around five colonies with a pipette tip and adding to 1 ml of LB-Miller media with appropriate antibiotics, then left to grow overnight (typically 12-18 hours) at 37 C with shaking at 200 rpm. The following morning, overnight cultures are diluted 1:100 into 2 ml MOPS EZ-Rich defined media (Teknova) with 10% LB-Miller media with antibiotics in a 14 ml polypropylene culture tube and grown at 37 C with shaking at 200 rpm. Optical densities are occasionally sampled using a Biotek Synergy HTX plate reader: 150 uL of culture is placed into a well on a black-side, clear-bottom 96-well plate and absorbance at 600 nm is measured. Cells are induced at indicated optical densities using indicated concentrations of anhydrotetracycline (aTc, Fisherbrand). After induction, cells are transferred to a 30 C shaker or left shaking at 37 C, as indicated. At 24-hours post-inoculation, 500 uL of culture is spun-down at 14k rpm for 2 minutes in 1.5 ml microcentrifuge tubes (USA Scientific). 300 uL of culture supernatant is then transferred to an Amicon 10 kDa centrifuge filter and centrifuged at 14k rpm for 20 min, or until enough volume has been filtered. 98 uL of culture with 2 uL 10% TFA
(i.e., 0.2% TFA) is then used for HPLC analysis. Culture without TFA is used for RNA aptamer ribosensor assays.

Aptazyme experiments in E. coli

An overnight culture is grown by selecting around five colonies with a pipette tip and adding to 1 ml of LB-Miller media with appropriate antibiotics, then left to grow overnight (typically 12-18 hours) at 37 C with shaking at 200 rpm. The following morning, overnight cultures are diluted 1:100 into 2 ml LB-Miller media or MOPS EZ-Rich defined media (Teknova), as indicated, with antibiotics in a 14 ml polypropylene culture tube and grown at 37 C with shaking at 200 rpm. Cells are induced at indicated optical densities using indicated concentrations of IPTG. Optical densities and fluorescence measurements are sampled as indicated using a Biotek Synergy HTX plate reader: 150 uL of culture is placed into a well on a black-side, clear-bottom 96-well plate. Absorbance is measured at 600 nm. Fluorescence is measured with a 485/528 nm filter using a bottom read with a gain setting of 40.

HPLC quantification of shikimate and 4-AF

The column used for all HPLC analysis is an Agilent ZORBAX Eclipse Plus Phenyl-Hexyl, 4.6 x 250 mm, with 5 uM particle size. HPLC is performed on an Agilent 1200 series. The program for 4-AF detection is as follows. With Solution A of Milli-Q H2O + 0.2% TFA and Solution B of Methanol + 0.2% TFA: 100% A for 4 min,
gradient to 100% B for 14 minutes, 100% A for 4 minutes. The program for shikimate detection is as follows. With Solution A of Milli-Q H2O + 0.1% TFA and Solution B of Methanol + 0.1% TFA: 100% A for 4 min, gradient to 70% B for 4 minutes, gradient to 95% B for 4 minutes, 95% B for 4 minutes, 100% A for 4 min.

4.3 | Shikimate Production

As shikimate is an important intermediate in the 4-ACA pathway, and its production the subject of much study [6], we attempted to produce high levels of shikimate using a construct, pS4, from Juminaga, et al. [112]. pS4 is a large plasmid with six enzymes from the shikimate biosynthesis pathway. Two, tktA and ppsA, are upstream of E4P and PEP, respectively, and thus are expected to increase accumulation of these precursors. The other four convert E4P and PEP to shikimate. One, aroG, a 3-Deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthase, is a feedback-resistant version (termd aroG\textsuperscript{fr}) of the initial step that converts E4P and PEP to DAHP. Without feedback on aroG, a pathway bottleneck is eliminated. The paper reports low protein expression levels and low promoter inducibility, but indicates that this construct can produce shikimate titers in excess of 700 mg/L in shake flasks after 24 hours.

Attempts to produce high levels of shikimate with pS4 and pS3, a construct similar to pS4, failed to recapitulate the published results, however. HPLC analysis of \textit{E. coli} cultures growing with pS3 or pS4 in DH10b cells in similar conditions to Juminaga, et al. (See Methods) for 24 hours showed titers of 55-57 mg/L of
shikimate regardless of induction (Table 4.1). While the cause of the discrepancy in titers is unclear, it seemed that expressing the full seven genes to convert PEP and E4P to chorismate would be unnecessary to reach high levels of 4-AF production [12]. A feedback-resistant DAHP synthase, aroG\textsuperscript{fbr}, was reported to contribute the most to high 4-AF levels.

<table>
<thead>
<tr>
<th>Inducer level</th>
<th>Plasmid pS3</th>
<th>Plasmid pS4</th>
<th>pS4 + 10 mM shikimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 uM IPTG</td>
<td>26 / 0.32 mM / 56 mg/L (3.55 min)</td>
<td>no data</td>
<td>no data</td>
</tr>
<tr>
<td>100 uM IPTG</td>
<td>26.5 / 0.33 mM / 57 mg/L (3.52 min)</td>
<td>no data</td>
<td>no data</td>
</tr>
<tr>
<td>1000 uM IPTG</td>
<td>no data</td>
<td>25.4 / 0.31 mM / 55 mg/L (3.56 min)</td>
<td>837 / 10.31 mM / 1790 mg/L (3.55 min)</td>
</tr>
</tbody>
</table>

**Table 4.1.** Shikimate production data. Format: Shikimate HPLC peak area / concentration in mM / concentration in mg/L. Peaks were observed at 3.6 min. Retention times are in parentheses. Measurements were taken at 260 nm. n=1.

### 4.4 | HPLC Detection of 4-AF

Concurrent with the work described in Sections 4.5 and 4.6, an HPLC program to detect 4-AF was developed. The high polarity of 4-AF leads to low retention when using common, polar HPLC mobile phases such as acetonitrile and methanol, and a common, polar stationary phase, such as water. Moreover, the molecule can adopt several protonation states, leading to a strong pH-dependence of retention. When the supernatant of a culture grown in unbuffered LB is used, the 4-AF peak elutes
when running only water through a phenyl-hexyl column, and moreover is unresolvable amongst other peaks. This low retention is partly due to low pH caused by the addition of 0.1% TFA, a common additive that improves peak resolution. Using a buffered media, such as MOPS EZ-Rich, and 0.2% TFA, the 4-AF peak is retained longer and is resolvable amongst nearby peaks. This concentration of TFA, while it reduces the pH, does not reduce it past the buffered range, and leads to a pH of around 7. A similar strategy is used by Masuo, et al., who use potassium phosphate at pH 7 as their stationary phase. Thus, a pH in the 7 range is expected to result in a lower-polarity state of 4-AF, allowing it to retain longer than the low-pH-protonation-state 4-AF and enable it to be detected more easily.

To detect 4-AF, we initially looked at the 280 nm wavelength, as 4-AF has an absorbance peak in the 280 nm range, similar to other aromatics. However, the limit of detection is lower with a 210 nm wavelength, as used by Masuo, et al. In fact, in all cases of 4-AF production detected with HPLC (see next section), including a control with 1 mM 4-AF, 4-AF was seen only at 210 nm and 220 nm and not at 280 nm.

### 4.5 | 4-AF Production Using S. venezuelae Operon

Initial attempts to produce 4-AF from glucose in E. coli were performed using a plasmid containing a PapABC operon from S. venezuelae, which was derived from pJC9A from Carothers, et al. [7], which was itself derived from pLASC-lacpw from Mehl, et al. [10]. This operon was codon-optimized and placed in front of a pBAD
arabinose-inducible promoter to create plasmid pJTS16. This construct was never seen to produce any 4-AF, using methods similar to those described in Methods and grown at 37 C. Subsequent analysis of the original papA sequence in the plasmid map PapABC operon showed many sequence deviations from the genomic sequence (91% amino acid identity) from which it was obtained, S. venezuelae (ATCC 10712) [10], including a single-base frameshift deletion. This deletion was fixed, the papA gene re-codon-optimized, and the promoter replaced with a Tn10 pA tet promoter, creating plasmid pJTS76D. This construct was also never seen to produce any 4-AF. SDS-PAGE analysis showed undetectable bands for all PapABC enzymes amongst whole-cell lysate. Designed RBSs [159] were introduced to improve expression of the enzymes, which SDS-PAGE confirmed. Additionally, the original papB sequence (in the obtained plasmid map) was analyzed against the genomic sequence and found to have a frameshift deletion, so the papB gene was replaced with the genomic version. (Note: papC is equivalent to the genomic version.) Plasmids with the genomic papB sequence include pJTS76W, pJTS76X, and pJTS76Y, with optimized RBSs for papA and papB, papA and papB and papC, and papA, respectively. Despite these many changes, no 4-AF production was observed. While it was not surprising that all constructs and conditions tested prior to finding and fixing the two frameshift deletions produced no 4-AF, it was surprising to see no 4-AF subsequently. It is expected that the differences remaining in the papA gene compared to the genomic version may create a nonfunctional PapA protein. A
fully genomic PapABC has not been tested.

### 4.6 | 4-AF Production Using *P. fluorescens* Operon

As an alternative to the *S. venezuelae* operon described in Section 4.5, the PapABC operon from *P. fluorescens* SBW25, described in Masuo, et al., was used. The papABC genes were placed downstream of a Tn10 pA tet promoter, and all RBS sequences were computationally designed for maximal expression using the RBS Calculator (v2) [58], to create plasmid pJTS123. Following the methods described in Methods, 4-AF was produced and detected via HPLC. Maximum titers observed using pJTS123 in strain NST37 were 57 mg/L. Potentially higher titers were observed using a two-plasmid system: pJTS123 and pJTS111. pJTS111 contains aroG^fbr^ and aroL, the two genes downregulated by tyrR, under control of the Tn10 pA tet promoter. With both pJTS123 and pJTS111 in DH10b (notably, a common cloning strain), titers of 104 mg/L were reached. A plasmid combining papABC and aroG^fbr^ and aroL into one five-gene operon controlled by pTet, plasmid pJTS128, yielded a maximum of 45 mg/L in the NST37 strain. We thus expect that there is a disadvantage to having a long, five-gene mRNA, and that smaller operons lead to improved expression, perhaps due to increased mRNA stability. Table 4.2 shows a construct list with associated maximum titers and conditions.
<table>
<thead>
<tr>
<th>Plasmid(s)</th>
<th>123</th>
<th>128</th>
<th>123 + 111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Strain</td>
<td>NST37</td>
<td>NST37</td>
<td>DH10b</td>
</tr>
<tr>
<td>[aTc] in nM</td>
<td>40</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Induction OD600</td>
<td>0.26</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Growth Temp.</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Peak Area</td>
<td>2584</td>
<td>2007</td>
<td>4686</td>
</tr>
<tr>
<td>[4-AF] in uM</td>
<td>229</td>
<td>178</td>
<td>415</td>
</tr>
<tr>
<td>[4-AF] in mg/L</td>
<td>57</td>
<td>45</td>
<td>104</td>
</tr>
</tbody>
</table>

**Table 4.2.** Highest observed 4-AF titers for the constructs shown. n=1, though similar conditions have been studied (different induction levels and times), which gives us confidence in the general trend of these results.

The impact of temperature on pathway titers was studied by growing cells with both pJTS123 and pJTS111, as described in Methods, identically up until induction, then growing one tube at 30 C and one at 37 C. As Table 4.3 shows, a 37 C post-induction growth appears to lead to a reduction of pathway titer, with approximately 11% of the 4-AF made compared to the 30 C tube. This indicates that the 4-AF enzymes are likely more active at 30 C than 37 C. While 30 C appears to be better for 4-AF production, the A. thaliana PAL2 enzyme that converts 4-AF to 4-ACA has been shown to have optimal turnover at 48 C [152]. Thus, this may indicate that a pathway comprised of papABC and PAL2 may require a temperature such as 37 C to find a balance between the opposing optimal temperatures for these enzymes.

We also assessed the impact of induction time on pathway performance. The data in Chapter 3 suggests that induction time may have a large impact on pathway
titer, and that delaying induction could improve titer. We saw that a delay from an OD600 of 0.1 to OD600 of 0.4 led to titers of 57 and 104 mg/L, respectively (Table 4.4). This suggests that an induction delay may improve titer.

<table>
<thead>
<tr>
<th>Construct</th>
<th>123 + 111</th>
<th>123 + 111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Strain</td>
<td>DH10b</td>
<td>DH10b</td>
</tr>
<tr>
<td>[aTc] in nM</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Induction OD</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Growth Temp.</td>
<td>30</td>
<td>37</td>
</tr>
<tr>
<td>Peak Area</td>
<td>3958</td>
<td>429</td>
</tr>
<tr>
<td>[pAF] in uM</td>
<td>350</td>
<td>38</td>
</tr>
<tr>
<td>[pAF] in mg/L</td>
<td>88</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 4.3.** Impact of growth temperature on 4-AF titer. Growing at 30 C post-induction results in titers nearly 9 times higher than post-induction growth at 37 C. n=1.

There are numerous potential explanations for the discrepancy between our 4-AF titers and those observed in Masuo, et al. Perhaps the most important is differences in media composition. They saw best performance in MOPS-M9 medium with what is essentially 40% LB added: 2 g/L yeast extract + 5 g/L tryptone + 4 g/L ammonium chloride. We used MOPS EZ-Rich medium, which contains around 0.5 g/L ammonium chloride, with 10% LB: 0.5 g/L yeast extract + 1 g/L tryptone. Of at least equal importance are presumed differences in expression and/or enzyme stoichiometry. They split the papABC operon in two, putting papA and papBC on separate plasmids, which improved titers around 25% from a one-plasmid papBAC operon. The gene order of papBAC v. our papABC gene order
could potentially lead to higher titers and better enzyme stoichiometry [112]. Lastly, their NST37 strain had an additional knockout of pheLA (pheL being the pheA leader peptide), which resulted in a 50% improvement in titer for them.

<table>
<thead>
<tr>
<th>Construct</th>
<th>123 + 111</th>
<th>123 + 111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Strain</td>
<td>DH10b</td>
<td>DH10b</td>
</tr>
<tr>
<td>[aTc] in nM</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Induction OD600</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Growth Temp.</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Peak Area</td>
<td>2543</td>
<td>4686</td>
</tr>
<tr>
<td>[4-AF] in uM</td>
<td>225</td>
<td>415</td>
</tr>
<tr>
<td>[4-AF] in mg/L</td>
<td>57</td>
<td>104</td>
</tr>
</tbody>
</table>

**Table 4.4.** Impact of induction time on pathway titer. Delaying induction from an OD600 of 0.1 to 0.4 resulted in an almost two-fold increase of 4-AF production. n=1.

### 4.7 4-AF-responsive Dynamic Control of Gene Expression

A new frontier of dynamic control in metabolic pathways is using RNA-based sensors instead of protein-based transcription factors. RNA aptazymes have been proposed as a means for achieving this [7]. Aptazymes contain an aptamer domain that can bind a metabolite of interest. When this aptamer domain is bound, it leads to a structural rearrangement that causes the aptazyme's ribozyme domain to self-cleave. When placed in the 5' UTR of an mRNA, this self-cleavage event leaves a 5'-OH. In *E. coli*, this 5'-OH instead of the typical 5'-PPP is thought to prevent RNase E binding at the 5' end, thus extending the lifespan of the mRNA. Different amounts of mRNA longevity improvement from a 5'-OH have been reported
[68,165], and one group has shown very context-specific effects of 5' UTR ribozyme cleavage on gene expression [166], including decreases in gene expression. In any case, it seems possible to develop metabolite-responsive changes in gene expression with 5' UTR aptazymes.

![Figure 4.3](image)

**Figure 4.3.** 4-AF-HCl effect on expression from 5’ UTR aRED constructs. T7 Express (lacIq/LysY) cells with pDSY2F were grown for 2.5 hours then induced with 200 uM IPTG and the indicated amount of Tris, Tris + pAF, Tris + pAF-HCl, and nothing (null) was added. Only the case with pAF-HCl shows an increase in fluorescence above the control, indicating that HCl is needed for the increase in fluorescence, and that 4-AF alone does not increase fluorescence. n=1.

Our lab's work to-date, however, has not led to the development of a 4-AF-responsive aptazyme that is capable of modulating gene expression in *E. coli*, though cleavage has been demonstrated *in vitro*. We have shown, however, a possible relationship between pH and presumed aptazyme cleavage (i.e., by seeing increased fluorescence) that suggests a strong need to perform aptazyme
experiments in buffered conditions. The 4-AF used in initial experiments was 4-AF-HCl, the HCl being for solubility, which is a common commercial formulation. With an aptazyme in the 5’ UTR of superfolder GFP (sfGFP), in plasmid pDSY2F, the addition of 10 mM 4-AF-HCl solubilized in 10 mM Tris potentially leads to greater in vivo fluorescence levels when cells were grown in unbuffered LB-Miller media. However, 10 mM 4-AF without HCl leads to potentially no improvement in expression (Fig. 4.3). A similar improvement in fluorescence to that seen with 4-AF-HCl addition can possibly be achieved with the addition of 10 mM HCl alone (Fig. 4.4), further suggesting that HCl and the concomitant decrease in pH is the agent causing the increase in expression.

**Figure 4.4.** HCl effect on expression from 5’ UTR aRED constructs. T7 Express (lacIq/LysY) cells with pDSY2F were grown for 2.5 hours then induced with 200 uM IPTG and the indicated amount of HCl. Fluorescence normalized to OD600 for the HCl+ case very quickly overtakes the HCl- case, ending around 3 times higher. Note that OD600 of the culture converges at 6 hours post-induction (not shown). n=1.
4.8 | Detecting 4-AF Using RNA Aptamer Ribosensors In Vitro

One of the bottlenecks in developing a high-titer metabolic pathway is screening production of many pathway variants. Often this is done with high-throughput chromatography, such as ultra performance liquid chromatography or gas chromatography, which limits screening throughput to around 1000 samples per instrument per day. The combinatorial explosion that can result from changes to pathway enzymes and culture conditions can result in millions of variants to screen, however. While design-of-experiment-based sampling can reduce the number of necessary variants, screening remains a costly and time-intensive endeavour. To address this challenge, our lab developed a methodology for creating RNA aptamer-based ribosensors that can be used for inexpensive, multiplexed, in vitro quantitative detection of metabolites for strain screening [167]. These ribosensors work by a strand-displacement reaction that occurs when the metabolite-of-interest binds the RNA aptamer and causes a structural rearrangement of the ribosensor. The rearranged ribosensor can then displace a strand of the output RNA duplex, freeing a fluorophore, which will then fluoresce for detection.

A proof-of-concept ribosensor sensitive to 4-AF was developed using a previously characterized 4-AF aptamer [168]. After a set of ambiguous HPLC results with S. venezuelae papABC pathway constructs (it was unclear if 4-AF was present), we tested these culture samples with the 4-AF ribosensors. As noted in Burke, et al.,
the ribosensors showed outputs consistent with titers in the 100 mg/L range. Subsequent analysis of the HPLC results, in addition to subsequent production experiments and HPLC analysis, confirmed that these strains do not produce 4-AF in observable quantities. It must then be concluded that the ribosensors did not detect 4-AF but instead detected another compound, likely similar in structure, that was generated from pathway induction. This false positive illustrates a potential limitation of RNA aptamer-based sensors for use with cell cultures containing a milieu of metabolites. It is difficult to ascertain which metabolite the ribosensors bound and thus which counterselection could have been done in order to prevent this.

4.9 | Conclusions and Future Directions

In this work, we produced shikimate and 4-AF in *E. coli* using previously developed components. We also developed a protocol for detection of 4-AF using HPLC. This work lays a foundation for developing a 4-AF and 4-ACA production platform, which could be used as a testbed to study how enzyme expression burden and product toxicity can be mitigated by controlling genetic expression.

There are several easy improvements to the 4-AF production pathway, outlined at the end of Section 4.6, that are expected to improve titers substantially, and there are many others that are worth noting. As an alternative to creating a NST37 (ΔpheLA) strain would be to use CRISPRi or small RNA (sRNA) to knock down expression of pheLA in the NST37 strain, or to knock down expression of tyrA, trpE,
pheA, and tyrR simultaneously in an MG1655 strain. The latter would presumably lead to better cell growth and not require supplementary amino acids to overcome auxotrophy. The timing of the CRISPRi or sRNA knockdown could be optimized for strain productivity. In addition to the knockdown program, cells would express papABC and aroGfr and possibly aroL and/or tktA. Masuo, et al. showed an improvement with tktA expression but have not examined expression of aroL. It is unclear if a one-plasmid system is inherently worse than a two-plasmid system in terms of reaching a high level of gene expression. Therefore, it will be necessary to construct and test sets of the 4-AF pathway genes in one-plasmid and two-plasmid configurations using various operon sizes and thus 2-4 different promoters and terminators. To further maximize titers, all sets of genes will need to be screened using RBSs of differing strengths in order to balance enzyme expression.

The primary barrier to high 4-ACA titers is the activity of the PAL enzyme. We have seen a maximum of around 1% conversion of extracellular 4-AF after 24 hours using the A. thaliana PAL2 isozyme. Others have reported similarly poor conversion with other isozymes [14]. Therefore, even with 1 or 2 g/L 4-AF, we would expect around 10-20 mg/L of 4-ACA. A priority for improving production should therefore be finding isozymes with greater specificity for 4-AF, potentially PAL/TAL or TAL (tyrosine ammonia lyase) enzymes that would be more likely to tolerate the para-position amine. Beyond that, either rational modifications or directed evolution could be used to create an enzyme more tolerant of said amine.
Together, the aforementioned strategies could lead to 4-ACA titers high enough to lead to product toxicity, though this hinges on finding a better PAL isozyme or engineering/evolving one. Alternative pathways, such as one for producing styrene [9] or 4-hydroxystyrene [8], have properties similar to the 4-ACA production pathway. That is, they require expression of a PAL or TAL enzyme, which could be used to study enzyme expression burden, and lead to production of a toxic compound, which could be used to study product toxicity. While it is convenient that RNA aptamer ribosensors exist to detect 4-AF, a rough cost-benefit analysis would indicate that the amount of effort required to develop a 4-ACA pathway into a testbed for studying both enzyme expression toxicity and product toxicity outweighs the benefit of having such a testbed. As mentioned, other pathways exist with these qualities, and it would likely be better to develop RNA-based sensors and gene expression controllers tailored to a pathway with desired properties that already exists.
Chapter 5 | Conclusions

The work in this dissertation is focused on the issue of toxicity and burden in engineered metabolic pathways, with the aims of better understanding why toxicity and burden occur, and how to mitigate negative impacts on pathway titers. While many outstanding questions remain, this work has substantively advanced the study of pathway dynamic control modeling and enzyme burden mitigation, and suggests fruitful further work.

In Chapter 2, we developed a computational framework for generating and analyzing large combinatorial sets of genetic feedback controllers that control enzyme expression in an engineered metabolic pathway with product toxicity. This framework could be used to generate and analyze large sets of kinetic genetic network models for virtually any area of synthetic biology.

In Chapter 3, we performed an in-depth experimental analysis of enzyme expression burden from one enzyme, PAL2. This work recapitulates previous results demonstrating the importance of expression timing and expression dynamics. Cell cultures with delayed expression have increased enzyme expression levels and substrate conversion levels, and also show better growth. Our work shows that burden from enzyme expression can induce large changes in population dynamics and encourage escape phenotypes, which may be a result of upregulated inducer efflux. We find that stationary phase promoters are capable of high levels of expression similar to a tet promoter, but also create less burden and facilitate
better enzyme conversion activity, which may result from slower enzyme expression rates. This work suggests new considerations and areas of inquiry with respect to burden in metabolic pathways, namely, how burden exacerbates inducer efflux; the relationship between enzyme expression rate, burden, and protein folding; and the utility of strong stationary phase promoters in metabolic engineering.

In Chapter 4, we demonstrated production and HPLC detection of roughly 100 mg/L of 4-AF in *E. coli*, recapitulating previously reported levels of production. Studies with PAL2 suggest roughly a 1% conversion *in vivo* of 4-AF to 4-ACA is possible. It thus seems likely that a pathway in *E. coli* to produce levels of 4-ACA relevant for the study of product toxicity is out of reach without finding or engineering a PAL isozyme with a higher conversion rate. The 4-AF pathway could still be useful as a testbed for development of RNA aptamer-based biosensors, however.

Going forward, this work suggests that perhaps the most fruitful area of further study surrounds the questions raised in Chapter 3. Namely, what is the relationship between enzyme expression rate, burden, and protein folding? Can we use strong stationary phase promoters to express large metabolic pathways? What does the optimal curve of enzyme expression rate over time look like, and how do we engineer that curve? These questions should provide the basis for much impactful research in the future.
References


12. Masuo S, Zhou S, Kaneko T, Takaya N. Bacterial fermentation platform for


49. Andrew Jones J, Vernacchio VR, Lachance DM, Lebovich M, Fu L, Shirke AN, et


Smolke CD, Carrier TA, Keasling JD. Coordinated, differential expression of two


91. He X, Chen Y, Liang Q, Qi Q. Autoinduced AND Gate Controls Metabolic Pathway Dynamically in Response to Microbial Communities and Cell Physiological State. ACS Synth Biol. 2016; doi:10.1021/acssynbio.6b00177


98. Martínez-Antonio A, Janga SC, Salgado H, Collado-Vides J. Internal-sensing


136. Huber R, Roth S, Rahmen N, Büchs J. Utilizing high-throughput experimentation to enhance specific productivity of an E.coli T7 expression


160. Quan J, Tian J. Circular polymerase extension cloning of complex gene


Appendix A | Genetic Sequences

pPAL2.1

LOCUS pPAL2.1      5731 bp  ds-DNA  circular  05-FEB-2018
DEFINITION .
FEATURES Location/Qualifiers
  protein_bind 698..714
    /label="Q2"
  misc_feature 3674..3802
    /label="dbl term"
  CDS complement(20..646)
    /label="tetR"
  promoter 647..719
    /label="tet promoter region"
  misc_feature complement(4759..5614)
    /label="Amp"
  CDS 755..2905
    /label="A. thaliana PAL2"
  rep_origin complement(3936..4618)
    /label="colE1 origin"
  misc_feature 3712..3782
    /label="recombined out"
  misc_feature 2906..2956
    /label="A(EAAAK)3A linker"
  CDS 2957..3673
    /label="sfGFP"
  terminator 4624..4729
    /label="T0"
  RBS 735..754
    /label="B0030 RBS"
  protein_bind 668..686
    /label="Q1"
ORIGIN
  1 AAAATGGCCA CCTGACGCTT TAAGACCCAC TTTCACATTT AAGTTGTTTT TCTAATCCGC
     61 ATATGATCAA TTCAAGGCCG AATAAGAAGG CTGGCTCTGC ACCTTGGTGA TCAAATAATT
     121 CGATAGCTTG TCGTAATAAT GGCGGCATAC TATCAGTAGT AGGTGTTTCC CTTTCTTCTT
     181 TAGCGACTTG ATGCTCTTGA TCTTCCAATA CGCAACCTAA AGTTAAAATG CCCACACGGC
     241 TTAGTGCAAT TAATGCATTC TCTTGAGAAA AACCTTGGTG GCATAAAAG GCTAATTGAG
     301 TTTCGGAGGT TTTCTACGT TTTTCTGTAG GCGGTGTACCC TTAATTGACT TTTGCTCCAT
     361 CGGATGACT TAGTAAAGCA CATCTAAAAC TTTTAGCGTT ATTACGTAAA AAATCTTGCC
     421 AGCTTTCCCC TTCTAAAGGG CAAAAGTGAG TATGGTGCCT ATCTAACATC TCAATGGCTA
     481 AGGCGTCGAG CAAAGCCCGC TTATTTTTTA CATGCCAATA CAATGTAGGC TGCTCTACAC
     541 CTAGCTTCTG GCGGATGTTA GGGGTGGTTA AAACCTTGAT TCGGACTTCA TTAAGGACGT
     601 CTATATGCCG CTATTATCCT TTTTCTTTSAT CTAAATCTAGA CATCATTAAT TCCTAATTTT
     661 TGTTGCACT CATGCGATTG TTAGACATT TTTGCACCT CCATAGTGA TTAAGAACAG
     721 AATTCAAAAG ATCTTTCATAG AAGAGATAT CACATATGAG CATGCTGAGC AAGTGTGTG
     781 CGGCAGGAGA GAGAAGACAA AAGTTGCGCT TAATACGAG ACTTTGTCGAG ATCTTCAAGA
     841 TTGGCCGTTTA GCCGCGGTAC AAATGAAGAG ATAGCAATT TATGGAAGT GAAAGATAGT
     901 CGAAGAGATAT CGTACAGCTT TCGTGAATCT TGGCGGAGAA ACATCTGAGA TGCGACAAGT
     961 TGTCGCGCATC TCCACCGTAG GAGGACGCG TGAGGTTAG TTAAGGAGGA ATTCAGAAGG
    1021 CGGTGTGAAGA GCTATACGAGT ATGAGGTTAT GAGAGATATG ATAAATGCTG CATGACGAT
    1081 CGGAGTGCGCC ATGGGCGTTTG GTCCTACCTTG TCCAGGGAAT ACAAATGGG ACAGCCGATT
    1141 AAAAACAGA GCGGATCATG TTGTGACAC GCAGGAAGGA AAGAGAGAGT
    1201 TCAGTCGGCA GCAGGCGCTC CGGAACAGAC CCTGATGGC CACATCTGGC CATCAGGATG
    1261 CGAAGGATAC TCGGGCACTC ACAGGCGGCA GACATCATAGC ATCAGCTGCC
    1321 CGAAGGATTAC TCGGGCACTC ACAGGCGGCA GACATCATAGC ATCAGCTGCC
    1381 TCTCTACTAC ATCGGCGCCG TTTCACCTGA CACATCTGGC
    1441 CGGTGAACTCG TATCGGCGGCA AAGACCGCT TGGACAGCTG TACAGGCTTT
    1501 CGATATTTAC GTAAGAAGGA GCTTACGCG CAAGGCAGTT
    1561 GCCGCTGGATG TGTACGATAC AAGCAGGACG CATCGGCACTG
    1621 GATCATATTCG GAGGAGAGT GAGAGAGAGT

136
5641 GCATTATCA GGGTTATGT CTCATGACCG GATACATATT TGAATGTATT TAGAAAATA
5701 AACAAATAGG GGTTCCGCGC ACATTTCCCC G
//
**pSFGFP.1**

**LOCUS** pSFGFP.1   3543 bp ds-DNA  circular  09-OCT-2017

**DEFINITION**

**FEATURES** Location/Qualifiers

<table>
<thead>
<tr>
<th>CDS</th>
<th>complement(136..762)</th>
<th>label=&quot;tetR&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>misc_feature</td>
<td>complement(2687..3542)</td>
<td>label=&quot;Amp&quot;</td>
</tr>
<tr>
<td>promoter</td>
<td>763..838</td>
<td>label=&quot;Tn10 TetA Promoter&quot;</td>
</tr>
<tr>
<td>rep_origin</td>
<td>complement(1864..2546)</td>
<td>label=&quot;colE1 origin&quot;</td>
</tr>
<tr>
<td>misc_feature</td>
<td>803..808</td>
<td>label=&quot;-10&quot;</td>
</tr>
<tr>
<td>misc_feature</td>
<td>1602..1730</td>
<td>label=&quot;dbl term&quot;</td>
</tr>
<tr>
<td>misc_feature</td>
<td>779..784</td>
<td>label=&quot;-35&quot;</td>
</tr>
<tr>
<td>misc_feature</td>
<td>851..872</td>
<td>label=&quot;sfGFP_RBS1&quot;</td>
</tr>
<tr>
<td>CDS</td>
<td>873..1589</td>
<td>label=&quot;sfGFP&quot;</td>
</tr>
<tr>
<td>protein_bind</td>
<td>814..832</td>
<td>label=&quot;TetO2&quot;</td>
</tr>
<tr>
<td>protein_bind</td>
<td>784..802</td>
<td>label=&quot;TetO1&quot;</td>
</tr>
<tr>
<td>terminator</td>
<td>2552..2657</td>
<td>label=&quot;T0&quot;</td>
</tr>
</tbody>
</table>

**ORIGIN**

```
1 CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC
61 ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA
121 GTGCCACCTG ACGTCTTAAG ACCCACTTTC ACATTTAAGT TGTTTTTCTA ATCCGCATAT
181 GATCAATTCA AGGCCGAATA AGAAGGCTGG CTCTGCACCT TGGTGATCAA ATAATTCGAT
241 AGCTTGTCGT AATAATGGCG GCATACTATC AGTAGTAGGT GTTTCCCTTT CTTCTTTAGC
301 ATGACCTAAT AAACACCACACT TAAAATTTTG AGCGTTATTTA CGTAAAAAT CATGCGAGCT
361 GACCTCTTCC AAAGGGGCAA AGTGGATATG GTGCCTATCT AACATCTCAA TGGCTAAGGC
421 GTGACGGGAAA GTGGGTTTCA CTGCTGTGTCT TGGCAGGCTG AGCAGACGTTG
481 ATGGGATATG GAAGCTTCTGA CAGTTTTTAC TCAGTCTCAT TTTTCACTTT TTTTCAACTT
541 GTGACGCTTTA ATCCATTTCTG CTGCTGTTTTA AGGATTTGAT TGGAGGCCC CGAACGCTCT
601 GTGACCGAGAA TGGGAGGTTA GGTGAGCTAG GTGAGCTAGT GAGGAGATTT TTTTCAAGTG
661 CTGGGACGTTA TTTTCTTACG TACATCATTT AATTTTTCTG TGGGAGAAGT TGTGATCAAA
721 CGCTACGATG CTGCTGCTGG GGTGAGCTAG GGGCGAGATG TGGGAGAAGT TGTGATCAAA
781 GAGACTATGT AGGAGACTAT GGGCGAGATG TGGGAGAAGT TGTGATCAAA
841 CAAAAGATCT GGAATAATTA AGGAGGTTAA ATAGTACGAA AGAAGGAAA TCTTTCCTCT
901 GAGTGTTTGG ATTATTCTGA AAAAAATAGT GGTTAAATTA TTTTCTTCAAA
961 ATGGGTTAGAAAAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1021 CTGGAAATATG TGGGTTTAAAG GGGCGAGATG TGGGAGAAGT TGTGATCAAA
1081 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1141 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1201 CCCTGCAAGG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG
1261 CTGCTGTTTTA TTTTCTTACG TACATCATTT AATTTTTCTG TGGGAGAAGT TGTGATCAAA
1321 GAGACTATGT AGGAGACTAT GGGCGAGATG TGGGAGAAGT TGTGATCAAA
1381 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1441 ATGGGTTAGAAAAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1501 ATGGGTTAGAAAAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1561 CACATGCACTA GGGCGAGATG TGGGAGAAGT TGTGATCAAA
1621 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1681 GTGCTGTTTTA TTTTCTTACG TACATCATTT AATTTTTCTG TGGGAGAAGT TGTGATCAAA
1741 CGCTGCTGCTG CTGCTGCTGCT GGGCGAGATG TGGGAGAAGT TGTGATCAAA
1801 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1861 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
1921 TGGGTTTAAAG GGGCGAGATG TGGGAGAAGT TGTGATCAAA
1981 AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA AAAAAAAATAA
2041 CACATGCACTA GGGCGAGATG TGGGAGAAGT TGTGATCAAA
```

139
pJTS123

LOCUS  pJTS123     6149 bp ds-DNA  circular  09-OCT-2017
DEFINITION .
FEATURES    Location/Qualifiers
misc_feature complement(138..993)
    /label="Amp"
misc_feature 3965..4306
    /label="pfPapB_Opt"
promoter 1757..1829
    /label="tet promoter region"
misc_feature 4307..4328
    /label="pfPapC_RBS1"
rep_origin 5464..6146
    /label="ColE1"
misc_feature 1845..1869
    /label="pfPapA_RBS1"
misc_feature 1870..3939
    /label="pfPapA_Opt"
CDS    complement(1130..1756)
    /label="tetR"
protein_bind 1778..1796
    /label="Q1"
misc_feature 3940..3964
    /label="pfPapB_RBS1"
terminator 3..108
    /label="T0"
misc_feature 4329..5201
    /label="pfPapC_Opt"
terminator 5202..5330
    /label="dbl_terminator"
protein_bind 1808..1824
    /label="Q2"

ORIGIN
1 GTGCTTGAG TCTCACCAC AAAAAACGCC CGGCGGCAAC CGAGCGTTCT GAACAAATCC
61 AGATGGAGTT CTGAGGTCAT TACTGGATCT ATCAACAGGA GTCCAAGCGA GCTCGTAAAC
121 TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT
181 TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT
241 ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT
301 ATCAGCAATA AACCAAGCAG CCGAAAGGCC CGAGCGCAGA AGTGGTCCCT CAACTTTATC
361 CGCCTCCATC CAGTCTATTA ATTTGTCGCC GGAAGCTAGA GTAAGTATTT CGCCAGTTAAA
421 TAGTGGGTTC AACGTTGGTG CCATTGCTAC AGGCATCGTG GTGTGAAGCT GTGCGTTGGG
481 TAGTGGTCCA TCCAGTCGCG GTTCCCAAGC ATCAAGCGCA GCTTATGTAT CCCCCATGTT
541 GTGCAAAAAA GCGGTGGTGC GCCATGGTTC CCGATGGTGG GTGCATTGGA AATGTGATGG
601 AGTGGTTATCA GTCAATGTTA TGGCCAGACT GCATAATTCG CTTACTGCATA TGCCATTCGG
661 AAGATGCTTC TCTGACACTG GTGATAGACT TACCAAGATCA TTCTGAAAA ATGTGATGCG
721 GCCGGCAGAT TGGCTCTGCC CGGGCCGCAAT ACCGGAATAT ACCGGGCCAC ATAGCAGAAC
781 GGTTAAAAAT GTCAATGTTA GAAAAAGTGC TCTGGGCAGA AAACCTCAA GAACCTTTCC
841 GCCTGTTGCG GCCAGTGGCA TTAAACCCC TCGTGACACC AACCTGATCT CACACTCATT
901 TCTCTTACA AACGGTTGCTG GTGTGAGGCC AAAAGCGGCG CAAAAAGCGG CAAAAAAGG
961 AAATAAGGCC ACACCGGAGA GTTGAAACTC ATACAATTC TCACTTCAAA ATATATGAAG
1021 CATTATCATG GTTATATGGC TCTAGAGGCC ATATAGATTT GAAAGTAAAAAATTAG
1081 AGAAATAAGG GTTCCGCCGCA CATTTCCTCCG CAAAGTCCCA CCTGAGCTCT TAAGACCCAC
1141 TCTCACCATT AAGTGTTTTT TCTAATTCGCC ATATGATCAA TTCAAGCGCG AATAAGGAGG
pJTS111

LOCUS JTS111 4514 bp ds-DNA circular 06-DEC-2017
DEFINITION
FEATURES Location/Qualifiers
  CDS complement(1000..1626)
    /label="tetR"
  CDS 1742..2794
    /label="aroG_fbr"
  protein_bind 1648..1666
    /label="Q1"
  misc_feature complement(21..815)
    /label="Kan/neoR"
  misc_feature 1715..1741
    /label="PapA_RBS2"
  terminator 3340..3468
    /label="dbl_termimator"
  rep_origin complement(3602..4313)
    /label="p15a"
  terminator 4404..4509
    /label="TO"
  promoter 1627..1699
    /label="tet promoter region"
  CDS 2815..3339
    /label="aroL"
  protein_bind 1678..1694
    /label="Q2"
ORIGIN
1  CTGCAAGCCCC AGAGTCCCGC TCAGAAGAAC TCGTCAAGAA GGCGATAGAA GGCGATGCGC
61  TGGCAATCGG GAGCGGCAGT ACCCTATTAC ACCCTATTAC ACCTATTACG CCGCTTACGC
121  AGCCGGGCAA AGATGAGTAG CCGCTTACGC TCTGCGTCGG GCATGCGCGC CTTGAGCCTG
181  GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA
241  GAGCCCGCTTT CCAATATCAG GGTAGCCAAC GCTATGTCCT GATAGCGGTC CGCCACACCC
301  AGCCGGGCAA AGATGAGTAG CCGCTTACGC TCTGCGTCGG GCATGCGCGC CTTGAGCCTG
361  GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA
421  GAGCCGGCTT CCAATATCAG GGTAGCCAAC GCTATGTCCT GATAGCGGTC CGCCACACCC
481  TTCTGCAGGC GACGAAAGCT AGATGACAGG AGATCGGAGC CCGCATCTTC GCCCAATAGG
541  GTGCGGCAGCC ATGATGGCGC GCGCTTACGC TCTGCGTCGG GCATGCGCGC CTTGAGCCTG
601  GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA
661  TTCTGCAGGC GACGAAAGCT AGATGACAGG AGATCGGAGC CCGCATCTTC GCCCAATAGG
721  TGGCGCAGCC ATGATGGCGC GCGCTTACGC TCTGCGTCGG GCATGCGCGC CTTGAGCCTG
781  GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA
841  TTCTGCAGGC GACGAAAGCT AGATGACAGG AGATCGGAGC CCGCATCTTC GCCCAATAGG
901  CTGCGCTTAC CGGCTGCAAG ATGATGACAGG AGATCGGAGC CCGCATCTTC GCCCAATAGG
961  GCTAAATGAG TCGACCATAC TACATTAATT GCCTGGCGCT TACGATCCAC TTCCACATTT
1021 AAGTTGTTTT TCTAATGCCA ATAGATCCAA TTTAACCAGG ATTTAAGAGG CGTGCCTTGG
1081 ACCTGAAGGA CTATATACAGT ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1141 AGGTGTTTTT TCTAATGCCA ATAGATCCAA TTTAACCAGG ATTTAAGAGG CGTGCCTTGG
1201 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1261 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1321 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1381 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1441 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1501 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1561 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG
1621 ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG ATGACAGGAG

144
pDSY2F

LOCUS pDSY2f 3014 bp ds-DNA circular 14-SEP-2016
DEFINITION .
FEATURES Location/Qualifiers
misc_feature 2806..2911 /label="T0"
             /ApEinfo_revcolor=#ffef86
             /ApEinfo_fwdcolor=#ffef86
CDS 1139..1855 /label="sfGFP"
             /ApEinfo_revcolor=#ff9ccd
             /ApEinfo_fwdcolor=#ff9ccd
misc_feature 969..1112 /label="pAF_sRSV1_155 Aptazyme"
             /ApEinfo_revcolor=#f58a5e
             /ApEinfo_fwdcolor=#f58a5e
misc_feature 961..968 /label="5' Insulator"
             /ApEinfo_revcolor=#85dae9
             /ApEinfo_fwdcolor=#85dae9
misc_feature complement(2118..2800) /label="colE1 origin"
             /ApEinfo_revcolor=#9eafd2
             /ApEinfo_fwdcolor=#9eafd2
misc_feature 1113..1119 /label="3' Insulator"
             /ApEinfo_revcolor=#faac61
             /ApEinfo_fwdcolor=#faac61
misc_feature 1856..1984 /label="dbl term"
             /ApEinfo_revcolor=#75c6a9
             /ApEinfo_fwdcolor=#75c6a9
misc_feature 919..935 /label="T7 promoter"
             /ApEinfo_revcolor=#faac61
             /ApEinfo_fwdcolor=#faac61
misc_feature 1120..1138 /label="Bglbrick RBS"
             /ApEinfo_revcolor=#d6b295
             /ApEinfo_fwdcolor=#d6b295
misc_feature complement(2941..3014) /label="Amp"
             /ApEinfo_revcolor=#b1ff67
             /ApEinfo_fwdcolor=#b1ff67
misc_feature complement(1..782) /label="Amp"
             /ApEinfo_revcolor=#b1ff67
             /ApEinfo_fwdcolor=#b1ff67
misc_feature 936..960 /label="T7 UTR"
             /ApEinfo_revcolor=#ffef86
             /ApEinfo_fwdcolor=#ffef86
ORIGIN
1 GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG
61 CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC
pJTS16

LOCUS pJTS16               6724 bp ds-DNA    circular   20-JUN-2016
DEFINITION Location/Qualifiers
FEATURES Location/Qualifiers
  misc_feature complement(5739..6594)
    /label="Amp"
    /ApEinfo_revcolor=#ff9ccd
    /ApEinfo_fwdcolor=#ff9ccd
  misc_feature 1124..1162
    /label="Operator_I1_I2"
    /ApEinfo_revcolor=#85dae9
    /ApEinfo_fwdcolor=#85dae9
  RBS 3625..3642
    /label="PapC_Mehl_RBS"
    /ApEinfo_revcolor=#ff9ccd
    /ApEinfo_fwdcolor=#ff9ccd
  terminator 5604..5709
    /label="T0"
    /ApEinfo_revcolor=#b1ff67
    /ApEinfo_fwdcolor=#b1ff67
  promoter complement(1036..1064)
    /label="araC_promoter"
    /ApEinfo_revcolor=#c6c9d1
    /ApEinfo_fwdcolor=#c6c9d1
  promoter 1161..1188
    /label="pBAD"
    /ApEinfo_revcolor=#b4abac
    /ApEinfo_fwdcolor=#b4abac
  RBS 1226..1241
    /label="PapA_Mehl_RBS"
    /ApEinfo_revcolor=#f58a5e
    /ApEinfo_fwdcolor=#f58a5e
  rep_origin 4916..5598
    /label="ColE1"
    /ApEinfo_revcolor=#9eafd2
    /ApEinfo_fwdcolor=#9eafd2
  misc_feature 1072..1093
    /label="Operator_01"
    /ApEinfo_revcolor=#c7b0e3
    /ApEinfo_fwdcolor=#c7b0e3
  CDS complement(7..885)
    /label="araC"
    /ApEinfo_revcolor=#84b0dc
    /ApEinfo_fwdcolor=#84b0dc
  CDS 3295..3624
    /label="PapB_Opt"
    /ApEinfo_revcolor=#f58a5e
    /ApEinfo_fwdcolor=#f58a5e
  CDS 3643..4629
    /label="PapC_Opt"
    /ApEinfo_revcolor=#f58a5e
    /ApEinfo_fwdcolor=#f58a5e
  RBS 3274..3294
    /label="PapB_Mehl_RBS"
    /ApEinfo_revcolor=#d6b295

148
ORIGIN

1 GACGTCTTAT GACAACCTGGA CCGCTACATC ATTCACCTTT TCTTCACACCG CGGCACGGAA
61 CTCGCTCGGG CTGGCCCCGG TGCATTTTTT AAATACCGGC GAGAAATAGA GTTGATCGTC
121 AAAACCAACA TTGCGACCGA CGGTGGCGAT AGGCATCCGG GTGGTGCTCA AAAGCAGCTT
181 CGCCTGGCTG ATACGTTGGT CCTCGCGCCA GCTTAAGACG CTAATCCCTA ACTGCTGGCG
241 GAAAAGATGT GACAGACGCG ACGGCGACAA GCAAACATGC TGTGCGACGC TGGCGATATC
301 AAAATGTGCTG CTGGCTACGG GTACGTCTATG AACTGACGA AACTGACCTT
361 CATCGGCGGA TGGAGCGACT CGTTAATCGC TTCCCATGCGC CGCAGTAACA ATTGCTCAAG
421 CAGATTTATGC CCGCCCGAGCG CAGATCGCGT GCTCGGCTTA CCCGATATC
481 CCCGCAAGCG CGCTTCAAGTT CCGCTGAGCT CCGCTGGAGG CCGCTGGCGC
541 GCCAAATATT GACGGCCAGTG TGGACCGTAG CGCTCGAGCA AGAACTGACAC
601 CCCATGCGGCA TACCATGGCC CGGGCCGGCC TGCCTGGAGG GACGGCAAGC
661 GTTTTTTGTG GACTATCGTT CCGAATTCGG CACAAATTAAT GTGTTAAGTTT
721 CCCGTAGCCCG CAGATTGCTGC GATGCGGATG ATGTTGCGCC ATCAACCGCT
781 AAAACAAATCG CGCTACGTGG GCTCGGCTTA TGCCAGAATT CGCACTGCGT
841 AAACGAGGTA TCCGGCGTCG CCGATGCGCC ATCGAGCGCT
901 GCTACACCGA GGCAACGCAG CACGTACGTT CCGCTGAGCT CCGCTGAGCT
961 TTACTTGCGTC TCGTCCGCTA CAAACCGGAGT AAACCCGCTT ATTTAAAAGCA CTTTCTTAAC
1021 AACGCGGACG GCCAAGACGCT CAAGCTGCGC TGGAGCAGGG TTCTTAGCTT
1081 AGTTCGCACAA GATCCTGGTGG TACGGCGGCT TGCCTGAGGC CAAGCTGGCG
1141 AAGATTTGAG GATCTGCACCC GCTGGTCTGG AACTACCGCC TGGACGGTTT
1201 GTATTTCTGT GATCATACCAA ACTGCGGTTT CAGTACGCTT
1261 ATACTTTGAGT CCGGCGGCTT CTCGCGGCTT CTTTGGGCTT
1321 CTGCGGCCTT TCGGCGGCTT TCTGCGGCTT CTTTGGGCTT
1381 CAATTTTCTG TACGCGGCTT TTCACTGGTG AGCCTCGCGT
1441 GTTGCGGAGT CAGTCGAGGG TCTGCGGCTT CTTGCGGCTT TCTTGGGCTT
1501 CTGGCGGCTT CAGTCGAGGG TCTGCGGCTT CTTTGGGCTT
1561 GCTGCGGCTT AATCCTGCGG CCGGCGGCTT TCTTGGGCTT
1621 TATGCGGCTT CAGTCGAGGG TCTGCGGCTT
1681 CAGATTTGAG GATCTGCACCC GCTGGTCTGG AACTACCGCC TGGACGGTTT
1741 GCACGGCGTG CAGTGCCGAC CCGGCGGCTT CTTTGGGCTT
1801 GTGCTGCGGCT CACGTAGCAG TCCGCGGCTT CTTTGGGCTT
1861 CACGCGGCGC GACACTGCGG TTGCTCGGCTT CTTTGGGCTT
1921 GTGCTGCGGCT CACGTAGCAG TCCGCGGCTT CTTTGGGCTT
1981 CACGCGGCGC GACACTGCGG TTGCTCGGCTT CTTTGGGCTT
2041 GCCATGCGG TCTGGCGGCTT CTTGCGGCTT CTTTGGGCTT
2101 CTGCTGCGGCT CACGTAGCAG TCCGCGGCTT CTTTGGGCTT
2161 AGCAGCGGCT CAGTGGCGCAG CACGTAGCAG TCCGCGGCTT
2221 GTGCTGCGGCT TCTGCGGCTT CTTTGGGCTT CTTTGGGCTT
2281 GCCACCGCGGC ACGTGGCGTCT AAGTGGCGGCT CTTTGGGCTT

149
pJTS76D

DEFINITION

FEATURES

LOCUS pJTS76D 6202 bp ds-DNA circular 08-APR-2017

CDS 4244..5230 /label="PapC_Opt" /ApEinfo_revcolor=#84b0dc

misc_feature complement(138..993) /label="Amp" /ApEinfo_revcolor=#ff9ccd

protein_bind 1778..1796 /label="Q1" /ApEinfo_revcolor=#84b0dc

protein_bind 1808..1824 /label="Q2" /ApEinfo_revcolor=#84b0dc

terminator 5255..5383 /label="dbl_termimator" /ApEinfo_revcolor=#85dae9

RBS 1845..1864 /label="BglBrick RBS" /ApEinfo_revcolor=#b7e6d7

RBS 1757..1829 /label="tet promoter region" /ApEinfo_revcolor=#f58a5e

RBS 1130..1756 /label="tetR" /ApEinfo_revcolor=#f58a5e

RBS 5517..6199 /label="ColE1" /ApEinfo_revcolor=#9eafd2

RBS 4226..4243 /label="PapC_Mehl_RBS" /ApEinfo_revcolor=#f58a5e

RBS 3875..3895 /label="PapB_Mehl_RBS" /ApEinfo_revcolor=#f58a5e

terminator 3..108 /label="T0" /ApEinfo_revcolor=#b1ff67

CDS 3896..4225 /label="PapB_Opt" /ApEinfo_revcolor=#f58a5e
misc_feature 1865..3874

ORIGIN

1  GTGCTTGGAT TCTCACCAAT AAAAAACGCC CGGCGGCAAC CGAGCGTTCT GAACAAATCC
61  AGATGGAGTT CTGAGGTCAT TACTGGATCT ATCAACAGGA GTCCAAGCGA GCTCGTAAAC
121  TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT
181  TCGTTCATCC ATATGGGCTG GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT
241  ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT
301  ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC
361  CGCCTCCATC CAGTCTATTA ATTTGGCCGG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
421  GCTGTCGTTG AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG
481  TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT
541  GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC
601  AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAAATTG CTCTACTGCTA TGCAATGCCT
661  AAGATGCCTT TTGCTGACCT GTGAGACTCT AACAAGTCGA TCTGGAAGAT AGTGTAGGCG
721  GCACGCAGGT TGGCTCTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC
781  TTATAAAGTG CTCATCATTT AATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
841  GCTGTCGTTG AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG
901  TACCTTTACC AGGGTTCTTG GTGAGAAGAA AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
961  AAAATTAGGG CCAGCGTGTG GCTATAGGAG AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
1021 CATTATTACG GGGTATTGTC TCTAGAGGCT GCTACTATTT ATATTGTTATT GAGAAAAAT
1081 ACAAAATAGAG GTTCGGCGCC GATCCCGACT AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
1141 TTTACCATTT AATTTTGTGG TCTCATCCGC ATATAGCATA CTCCAGCAGC AAATAAGGAG
1201 CTGTGCTTGG ACCTTCTGGG TGAATAAGTT AAGTCTGGCTA CGCTCGTTCTG CAACTTTATC
1261 TATCATGTG TGGGTTTGTTC TCTTCTCCTT TACAGCTTCTG ATGCTGTGTTG CCCCCATGTT
1321 CGCAACCTAA AGTAAAATGC CCCACAGCGC TGAGTGCATA TAATGCATTC TCTAGGAAA
1381 AACCTTGCCG GCATAAAAGT CCATCGAGCA TCTAGGAAA
1441 GCCGTGTACC TAAATGTACT TTTGCTCCAT CGCGATGACT TAGTAAAGCA CATCTAAAC
1501 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
1561 TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT
1621 CATGTCGTTG AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG
1681 AACCTCTGGT CTCCGACACT TAAATGCATG TCTACGACTG CAAAAAGGAA
1741 CTGTGACCTG ATTTTGAATT CTGCTCGTTG GATCTCGGTT GTAGTTTGTGG CCCCCATGTT
1801 TATCATGTG TGGGTTTGTTC TCTTCTCCTT TACAGCTTCTG ATGCTGTGTTG CCCCCATGTT
1861 ACATATGCG TGGCTC GCCGATTTTG GTGGTTATGG GCATGTGTTG CCCCCATGTT
1921 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
1981 GCCGTGTACC TAAATGTACT TTTGCTCCAT CGCGATGACT TAGTAAAGCA CATCTAAAC
2041 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2101 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2161 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2221 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2281 ATGATAGGTT GCTTCGCTAC TCTGAGCTGT GTGAGATGATG GTTCCCAACG ATCAAGGCGA
2341 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2401 GCAGCGTGGC CATCATGCTG TGGTGAACCT TGGAGAGAAA AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
2461 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2521 AGTGTTATCA TGGGTTTGTTC TCTTCTCCTT TACAGCTTCTG ATGCTGTGTTG CCCCCATGTT
2581 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2641 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2701 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2761 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2821 GCAGCGTGGC CATCATGCTG TGGTGAACCT TGGAGAGAAA AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
2881 GAAGATGATG GTTCCCAACG ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA
2941 GCAGCGTGGC CATCATGCTG TGGTGAACCT TGGAGAGAAA AAAAAACGCC CGGCGGCAAC CGAGCGTTCT
3001 TCTGTCGTTG ACGCTGGGCT ATTTTTGAAT TGAGATGATG GTTCCCAACG ATCAAGGCGA

153
pJTS76W

DEFINITION

LOCUS       pJTS76W                 6214 bp ds-DNA     circular     10-OCT-2017

FEATURES             Location/Qualifiers

CDS             complement(1130..1756)
/label="tetR"
/ApEinfo_revcolor=#d59687
/ApEinfo_fwdcolor=#d59687

terminator      3..108
/label="T0"
/ApEinfo_revcolor=#b1ff67
/ApEinfo_fwdcolor=#b1ff67

misc_feature    complement(138..993)
/label="Amp"
/ApEinfo_revcolor=#ff9ccd
/ApEinfo_fwdcolor=#ff9ccd

protein_bind    1808..1824
/label="Q2"
/ApEinfo_revcolor=#d59687
/ApEinfo_fwdcolor=#d59687

protein_bind    1778..1796
/label="Q1"
/ApEinfo_revcolor=#84b0dc
/ApEinfo_fwdcolor=#84b0dc

promoter        1757..1829
/label="tet promoter region"
/ApEinfo_revcolor=#f58a5e
/ApEinfo_fwdcolor=#f58a5e

misc_feature    3882..3925
/label="PapB_RBS1"
/ApEinfo_revcolor=#75c6a9
/ApEinfo_fwdcolor=#75c6a9

misc_feature    3926..4237
/label="PapB_Opt (genomic)"
/ApEinfo_revcolor=#d59687
/ApEinfo_fwdcolor=#d59687

CDS             complement(4256..5242)
/label="PapC_Opt"
/ApEinfo_revcolor=#84b0dc
/ApEinfo_fwdcolor=#84b0dc

terminator      5267..5395
/label="dbl_terminator"
/ApEinfo_revcolor=#85dae9
/ApEinfo_fwdcolor=#85dae9

rep_origin      5529..6211
/label="ColE1"
/ApEinfo_revcolor=#9eafd2
/ApEinfo_fwdcolor=#9eafd2

misc_feature    1845..1871
/label="PapA_RBS2"
/ApEinfo_revcolor=#f58a5e
/ApEinfo_fwdcolor=#f58a5e

RBS             complement(4238..4255)
/label="PapC_Mehl_RBS"
/ApEinfo_revcolor=#ff9ccd
ORIGIN

1 GTGCTTGGAT TCTCACAAT AAAAAGGCC CGGCACGAC CGAGGCTTCT GAACAAATCC
61 AGATGGAGTT CTGGAGCTAC TATGGAATCT ATCAACAGGA GTCCAAGCGA GCTCGTAAAC
121 TTGTCCTGAC AGTTACCAC GTTTAATCAG TGGCTGTACC ATCTCAGCGA TCTGTCTATT
181 TCGTTCATCC ATAGTGCCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT
241 ACCATCTGCC CCAAGTCGTT CAATGATACC CGCAGACCCA CGCTCACCGG CTCCAGATTT
301 ATACGAAATGA AACCAGCAGG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC
361 CGCCTCCATC CAGTCTATTA ATTGTCAGCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
421 TACCTTCCCAG TACCTTCCCG GTTCACCGAG ATCAAGCGTA GCTGCTACGT CTGCGGTGGG
481 TAGTGGTCTGA TCTGTTCAGGT GAAACGGCAA CAGCAAGCGA GAAGAAGGGG CAAAACAGGG
541 ATGCCAAATTA AATCCCGGGC AGGAGGCGAC GTGCTTGGAT GCTCTGTCAT TCTTGTCCCT
601 AGTGTATATC CTCATGTTCA TGGCAGCAGT CATATAATTCT TTACTATGC TAAGGACAGT
661 AAGAGACCGT TCTGCTGACT GTTGTGACCT ATCAATCAGA GTCCAAGCGA GCTCGTAAAC
721 CGGCACGAGT TGCTCTTGCC GCAGGCTCCT ATAGTGCCCT GACTCCCCGT CGTGTAGATA
781 TTATTATTAG TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC TAAGGACAGT
841 GTGTGTGGAA TCCACCTGGA TGTCAACCGA CAGTTACGTT AGTCCCAGTT AGTTTATTCT
901 TACCTTCCCAG TACCTTCCCG GTTCACCGAG ATCAAGCGTA GCTGCTACGT CTGCGGTGGG
961 ATGCCAAATTA AATCCCGGGC AGGAGGCGAC GTGCTTGGAT GCTCTGTCAT TCTTGTCCCT
1021 CATTATATCA CTATTATATCA TGGCAGCAGT CATATAATTCT TTACTATGC TAAGGACAGT
1081 ACAATTTAGGG GTTCGGCCAG CATTCCCCCG AAAAGTGCCCA CTCAGACCTG TAAAGACCCAC
1141 TTACACATTT AAAGTTGTCT TCTATCCCGC ATATGACCTA TCCAGACCAC AATAAGAACG
1201 CTGCTCCTGA ACCTTGGTCA TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
1261 TACATCAAGT AGTTGCAATGA CTATTATCT TTACTATGC TAAGGACAGT CATATAATTCT
1321 CGACACCTAA AGTAAAGATC CCAACAGCGT TGATGCTGCA TAAATAGTAC TTACTATGC
1381 AACCTTGTGTG CATAAAGAG GCTATATGTAT TTTTGTGTTA TACATTATC TACATTATC
1441 GCCGATCTGC TAAATGTACT TTTGCTCCAT CGCGATGACT TAGTAAAGCAG CACCTAATAC
1501 TTTGTTTGCG CCACTGAAAT GAAAGACCGT AAATTATGCAT TTGTTTATTG AATGGTGATC
1561 TAGTGCTCCAG ATCCTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT
1621 CATGCTACAG CATTCCCCCG AAAAGTGCCCA CTCAGACCTG TAAAGACCCAC
1681 AACCTTCTGAT CTCAGACCTG AAATTATGCAT TTGTTTATTG AATGGTGATC CATATAATTCT
1741 CTAATCTTGA CATCATTATC TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
1801 TTAACACATTT AAAGTTGTCT TCTATCCCGC ATATGACCTA TCCAGACCAC AATAAGAACG
1861 AGGAGGCGAC GTCTTGGTCA TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
1921 TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC TAAGGACAGT CATATAATTCT
1981 GCCGATCTGC TAAATGTACT TTTGCTCCAT CGCGATGACT TAGTAAAGCAG CACCTAATAC
2041 ATCCTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2101 TGGGTTGTCC GATAAAAAAG CATTAAAAAT AACCTCAGAT GTCTTGGTCA TCAAAATATT
2161 GTTGTTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2221 CGGAGGCGAC GTCTTGGTCA TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
2281 CATGCTACAG CATTCCCCCG AAAAGTGCCCA CTCAGACCTG TAAAGACCCAC
2341 TTCCTTTGTC ATGCTTGGAT GCTCTGTCAT TCTTGTCCCT ATAGTGCCCT GACTCTTATT
2401 GTGCTTGGAT CAGCAGCTGAGT GCAGAGCAGT AGTGGTCCTG CAACTTTATC TTACTATGC
2461 CGGAGGCGAC GTCTTGGTCA TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
2521 CGGAGGCGAC GTCTTGGTCA TCAAAATATT CAGATGGATC TCCTTAAAT ATGATGGCAAC
2581 TTAATCTTGA TTGTTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT
2641 TAGTTCGTTTGA AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2701 TTAGTTCGTTTGA AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2761 TTGTTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2821 GTGCTTGGAT CAGCAGCTGAGT GCAGAGCAGT AGTGGTCCTG CAACTTTATC TTACTATGC
2881 TTGTTTGGAT AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
2941 TAGTTCGTTTGA AGTTTATTCT TTACTATGC TAAGGACAGT CATATAATTCT TTACTATGC
3001 GTGTTTGGAT CAGCAGCTGAGT GCAGAGCAGT AGTGGTCCTG CAACTTTATC TTACTATGC