GENETIC TOOLS AND APPROACHES FOR ENGINEERING METABOLISM AND METABOLIC PATHWAYS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Cameron James Glasscock

May 2019

GENETIC TOOLS AND APPROACHES FOR ENGINEERING METABOLISM AND METABOLIC PATHWAYS

Cameron James Glasscock, Ph. D.

Cornell University 2019

The goal of biomanufacturing is to create cost-effective, sustainable and renewable routes for production of biologics, pharmaceuticals, biofuels, commodity chemicals, and materials. Central to these efforts is the endowment of cells with new enzymatic machinery and altering of cellular metabolism to support carbon flux to the desired products. These tasks create a large array of unsolved challenges, from the determination and heterologous expression of enzymes that can perform requisite chemical transformations to the toxicity caused to the host by enzymes, pathway intermediate metabolites, and products. Furthermore, the task of re-directing carbon flux from natural pathways to the desired product is often non-intuitive since cells can drastically shift their metabolism in response to engineering manipulations, necessitating technologies that can rapidly produce and evaluate many manipulations of cellular metabolism in high throughput. Altogether, these issues create a need for new tools and approaches that resolve long-standing challenges in the engineering of cells for biomanufacturing. In this work, we explore the challenges associated with this goal as well as develop and introduce new synthetic biology technologies, particularly involving new advances in RNA gene expression regulation, that will speed the pace of cellular engineering and enhance the production capabilities of engineered cells. We begin by applying classical metabolic engineering tools to a pathway for eukaryotic protein glycosylation in *Escherichia coli* that lays the groundwork for efficient production of designer therapeutic glycoproteins. Building off this work, we explored the application of synthetic RNA regulators for high-throughput screening of endogenous gene knockdown in the context of the protein glycosylation pathway and propose a general approach for high-throughput strain engineering with RNA regulators. Finally, we developed and applied a new technology

for implementing dynamic feedback control of gene expression in response to toxicity and burden imposed on cells by heterologous pathways and toxic intermediates. In addition to other examples, we applied this technology to a pathway for Taxol precursor production to improve its titer and productivity. Altogether, this work elucidates contemporary obstacles in engineering cells for biomanufacturing and provides new tools and approaches to address these obstacles.

BIOGRAPHICAL SKETCH

Cameron was raised by parents Lori and Shannon Glasscock in Sacramento, California, Brandon, Mississippi, and Monmouth, Oregon. He earned a B.S. in Biological Engineering from Oregon State University in 2013 where he performed research in cryopreservation of cells and tissues under the mentorship of Adam Higgins. During this time, he also pursued research developing MRI contrast agents for tracking transplanted cells *in vivo* in the lab of Thomas Meade at Northwestern University. In August 2013, Cameron entered the PhD program at Cornell's School of Chemical and Biomolecular Engineering joining the labs of Julius Lucks and Matthew DeLisa, where he worked on engineering protein glycosylation pathways in bacteria. During his PhD, Cameron also became a visiting predoctoral scholar at Northwestern University's Center for Synthetic Biology, where he developed novel RNA-based technologies for controlling enzyme expression.

This work is dedicated to my parents and siblings.

This work was supported by the National Science Foundation Graduate Research Fellowship Program (Grant Number 1144153).

ACKNOWLEDGEMENTS

To my parents, for encouraging and cultivating my interests in science from a young age and your incredible support throughout my life. Thank you for your endless drive to ensuring I had the right experiences to explore and pursue my dreams growing up. I always look forward to phone calls and restful visits to home in the Oregon woods. Dad, thank you for teaching me the importance of grueling exercise! Marathon training is the perfect emotional distraction from the frustrations of failed grad school experiments (I haven't decided yet if the distraction leads to greater happiness, but I have at least determined that it is a totally different type of painful experience!). Running the Crater Lake Rim Run with you and the crew every summer is a favorite tradition of my year.

To my siblings Allison, Bethany, Daniel, Elaina, Faerynn, and Gavin for all of your friendship and fun times. Allison, your dedication to pursuing a PhD in philosophy was an inspiration for me in heading to grad school and it was great to have a family member on the same side of the country during my time in Ithaca. Juan, Cassio, Jessica, Jeremy, and Cody, you are all a blast and I've greatly enjoyed spending time with each of you.

To Julius and Matt for your excellent vision, guidance, and mentorship. Because of your enthusiasm, I am even more fascinated by the potential of synthetic biology for solving societal problems than when I started. I certainly wouldn't have gotten far in the serpentine path of engineering biology without your training and insight.

To all of the DeLisa and Lucks lab members and the broader SynBio community at Cornell and Northwestern, for your camaraderie, help, and great advice. I've thoroughly enjoyed

working alongside and getting to know each of you. Paul, thank you for your endless thoughtful advice, encouragement, and patience – you are truly an incredible colleague and scientist. Additionally, thank you to Jack, Bradley, Tyler, and Min for the great advice and enthusiastic help in the rSFP project.

Angela, thank you for being such an amazing partner. You've been a great support for me and I'm looking forward to tackling our next steps together.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

aTc - anhydrotetracycline

C. jejuni - *Campylobacter jejuni*

CPR - cytochrome P450 reductase

CYP - cytochrome P450

DMAP - dimethylallyl diphosphate

DNA - deoxyribonucleic acid

E. coli - Escherichia coli

FACE - fluorophore assisted carbohydrate electrophoresis

FACS - fluorescence-assisted cell sorting

FSC - forward scatter

G3P - glyceraldehyde-3-phosphate

GDP - guanidine diphosphate

GGPP - geranylgeranyl diphosphate

glycoSNAP - glycosylation of secreted *N*-linked acceptor proteins

GlycTag - glycosylation tag

GOI - gene of interst

GTase - glycosyltransferase

HSL - homoserine lactone

iFFL - incoherent feedforward loop

IPP - isopentenyl diphosphate

LASO - looped antisense oligonucleotide

LLO - lipid-linked oligosaccharide

LPS - lipopolysaccharide

MAGE - multiplexed automated genome engineering

MALDI-MS - matrix assisted laser desorption/ionization-mass spectrometry

$Man_3GlcNAc_2$ - $Mannose_3$-*N*-$acetylglucosamine_2$

MBP - maltose binding protein

MEFL - molecules of equivalent fluorescein

MEP - methylerythritol phosphate

mRNA - messenger RNA

N-linked - asparagine-linked

OTase - oligosacharyltransferase

PBS - phosphate-buffered saline

QS - quorum-sensing

RBS - ribosome binding site

RNA - ribonucleic acid

rSFP - riboregulated switchable feedback promoter

rSSP - riboregulated switchable stabilized promoter

S. cerevisiae - *Saccharomyces cerevisiae*

sfGFP - super folder GFP

SFP - switchable feedback promoter

sRNA - small RNA

SSC - side scatter

STAR - small transcription activating RNA

TALE - transcription-activator-like-effector

TG - taxadiene synthase/geryanlgeranyl diphosphate

TRACE - tracking combinatorial engineered libraries

TRMR - trackable recursive multiplex recombineering

UDP - uridine diphosphate

Und-PP - undecaprenyl diphosphate

UTR - untranslated region

CHAPTER 1- Introduction

1.1 Engineering cells for microbial biomanufacturing

Biomanufacturing holds great promise in the creation of cost-effective, sustainable and renewable routes for production of biologics, pharmaceuticals, biofuels, commodity chemicals, and materials (Keasling, 2010). In order to develop scalable biomanufacturing platforms, microbial cell hosts must be engineered with new enzymes and finely tuned to optimize their properties, such as titer, rate, yield, and purity – altogether creating a need for new tools and approaches that resolve long-standing challenges in the engineering of microbial hosts for biomanufacturing. In chapter 1, a few of the many promising goals of biomanufacturing are explored, from production of therapeutic glycoproteins in bacteria to the sustainable synthesis of natural products useful for human health and agriculture. Next is an overview of the extensive efforts that have thus far been made in metabolic engineering and synthetic biology to develop tools and approaches for engineering of endogenous metabolism and metabolic pathways, with a particular focus on tools relevant to the microbial workhorse *Escherichia coli*, including the development of versatile RNA regulators that control gene expression. Finally, this chapter presents an overview of the work presented in this dissertation, including the optimization of a pathway for eukaryotic protein glycosylation in *E. coli* and the development of several novel technologies that may be helpful in solving a few of the long-standing challenges encountered in metabolic engineering, such as screening gene knockdown targets in strain engineering and managing burden and stress.

1.2 Biomanufacturing of N-glycoproteins with living (and non-living) bacteria

Protein-based therapeutics are a highly successful class of drugs that command a market of more than $100 billion in the US and the European Union (Dimitrov, 2012). These drugs are typically produced in large reaction vessels by cells that have been genetically engineered to express a desired protein. An important consideration in these processes is the choice of host cell, which is partially driven by the need for protein modifications that are necessary for therapeutic function. Among the most common of these is asparagine-linked (*N*-linked) glycosylation (Helenius & Aebi, 2001), a process which decorates proteins with complex sugars (oligosaccharides) called glycans. This glycoconjugation can help proteins fold into their proper form (Wormald & Dwek, 1999); target them to particular tissues in the body (Friedman et al., 1999); and modulate biological activity, degradation, and clearance (Macdougall, 2003; Rothman, Perussia, Herlyn, & Warren, 1989; Sola & Griebenow, 2009). Additionally, the human body can develop an immunogenic response to proteins with *N*-linked glycans that are from another organism (LaTemple, Abrams, Zhang, & Galili, 1999). Since *N*-linked glycosylation differs between eukaryotic and bacterial species, “human-like” therapeutic glycoproteins are produced with expensive mammalian cell expression systems that have a long development time, potential for viral contamination, and a tendency towards product heterogeneity that can lead to adverse effects on drug potency and pharmacokinetics (Sethuraman & Stadheim, 2006). In contrast, the bacterium *E. coli* is an ideal production host because it is genetically tractable, has simple nutritional requirements, and can be grown to high cell density (Baneyx, 1999; Hockney, 1994).

Engineering *E. coli* to efficiently produce human-like glycoproteins may lead the way towards the rapid development of safe, inexpensive, and uniform production platforms for existing and novel therapeutics. In addition to use as a large-scale production platform, engineering *E. coli* with bottom-up protein glycosylation pathways raises the prospect of synthesizing preparatory amounts of glycoproteins with designer glycan structures so that glycan structure-function relationships can be more easily interrogated. The use of *E. coli*-based cell-free protein synthesis and glycosylation platforms (Jaroentomeechai et al., 2018) is especially compelling in this latter use, due to increased modularity and faster design-build-test cycles over living systems.

N-linked glycosylation in bacteria was first discovered in the pathogenic ε-proteobacterium *Campylobacter jejuni* (Szymanski, Ruijin, Ewing, Trust, & Guerry, 1999). Shortly after its discovery, the *C. jejuni* glycosylation system was functionally transferred to *E. coli* (Wacker et al., 2002), giving the latter organism the unnatural ability to produce *N*-linked glycoproteins. Despite this success, there are several characteristic differences from eukaryotic organisms that present obstacles for producing human-like glycoproteins in *E. coli*. First, all known bacterial *N*-linked glycosylation pathways utilize a completely distinct set of glycan structures from their eukaryotic counterparts (Weerapana & Imperiali, 2006). Second, eukaryotic species transfer glycans onto the asparagine of a specific amino acid consensus sequence that differs from that of many bacteria (Jervis et al., 2012). Finally, eukaryotic glycans are modified by glycan-processing enzymes through a set of intricately compartmentalized processes in the endoplasmic reticulum and golgi apparatus that are absent in bacteria (Herscovics, 1999).

1.2.1 Biosynthesis of eukaryotic glycan structures in E. coli

Progress has been made to address differences in bacterial and eukaryotic glycan structures using bottom-up pathway engineering strategies. One of the first demonstrations of eukaryotic glycoprotein synthesis using an *E. coli* derived glycoprotein was demonstrated by Schwarz et al. (Schwarz et al., 2010). In this system, protein glycosylation with $Man_3GlcNAc_2$ was achieved by first producing proteins in *E. coli* containing heterologous glycosylation machinery from *C. jejuni*. These proteins were then purified, trimmed, and remodeled by a series of in vitro enzymatic steps to create a eukaryotic glycosylation profile. While this study demonstrated eukaryotic glycosylation of an *E. coli*-derived glycoprotein, it required in vitro reactions with an oxazolene $Man_3GlcNAc$ donor substrate (C. Li & Wang, 2018) and has limited scalability. More recently, the DeLisa lab engineered a synthetic pathway in *E. coli* to glycosylate proteins with mannose$_3$-*N*-acetylglucosamine$_2$ ($Man_3GlcNAc_2$), the core structure of all human *N*-linked glycans (Valderrama-Rincon et al., 2012). In this pathway, which can be divided into distinct stages of glycan synthesis, glycan flipping, and glycan targeting, genes from *Saccharomyces cerevisiae* and *C. jejuni* were inserted into a plasmid that complements native *E. coli* enzymes to encode for the $Man_3GlcNAc_2$ glycosylation pathway (**Fig. 1-1**). Despite successful protein glycosylation with the $Man_3GlcNAc_2$ glycan in this system, glycoprotein titers were low (~10 μg/L) and < 1% of target proteins were glycosylated.

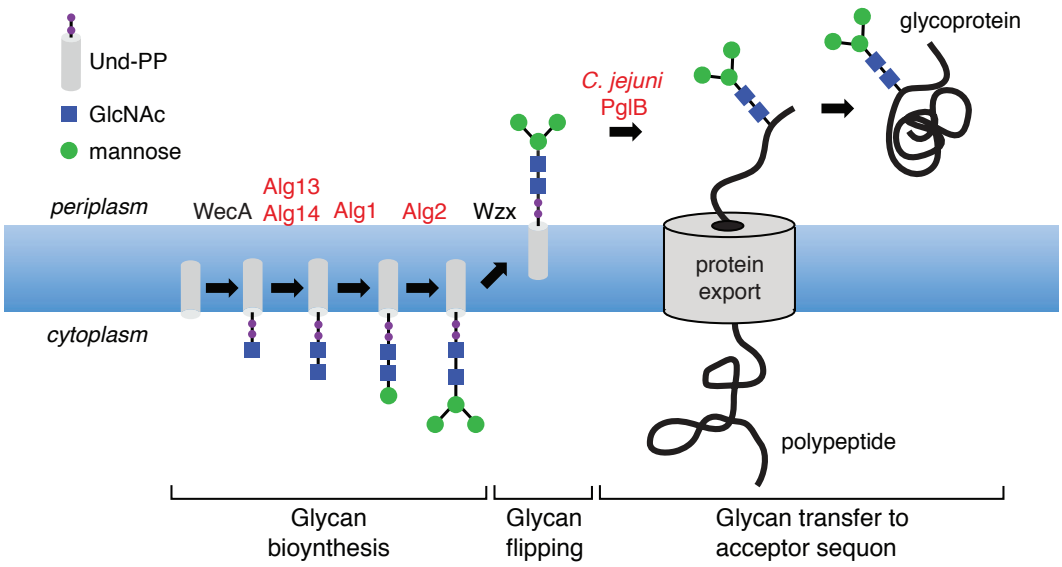


Figure 1-1. The pathway for eukaryotic *N*-linked protein glycosylation in *E. coli.*

The $Man_3GlcNAc_2$ protein glycosylation pathway can be divided into three distinct stages: glycan biosynthesis, membrane translocation of glycans, and glycan transfer onto polypeptide acceptor sequons. In the first stage, $Man_3GlcNAc_2$ glycans are assembled on the lipid carrier Und-PP on the cytoplasmic face of the inner membrane by yeast GTases Alg13, Alg14, Alg1, and Alg2, along with endogenous *E. coli* enzyme WecA. In the second stage, lipid-linked $Man_3GlcNAc_2$ glycans are flipped to the periplasmic side of the inner membrane by the action of the flippase Wzx. In the third stage, the OTase PglB transfers $Man_3GlcNAc_2$ glycans from Und-PP to asparagine residues of acceptor proteins that are exported into the periplasm. Protein glycosylation can occur in a co- or post-translational manner, involving acceptor site sequences having a consensus of D/E-X_1-N-X_2-S/T (where X_1 and X_2 can be any amino acid but proline). Figure adapted from Valderrama-Rincon et al., (Valderrama-Rincon et al., 2012).

1.2.3 Advances in cell-free glycoprotein synthesis for enzyme characterization and spatiotemporal reaction control

Finally, significant progress has been made in the development of *E. coli*-based cell-free platforms for glycoprotein synthesis (Guarino & Delisa, 2012; Jaroentomeechai et al., 2018). Cell-free glycoprotein synthesis platforms typically involve lysis of bacterial cells to release extracts with the enzymatic machinery necessary for transcription, translation, and glycosylation. These platforms promise to aid in enzyme characterization and the development of new bacterial glycosylation pathways as well as provide access to glycoproteins that are currently unattainable using conventional cell-based synthesis methods. In a report by Jaroentomeechai *et al*., the optimized pathway presented in Chapter 2 was used to create a bacterial extract capable of *N*-glycosylating proteins with the eukaryotic $Man_3GlcNAc_2$ glycan (Jaroentomeechai et al., 2018). This platform could be used as the basis to further modify glycan structures to create truly human-like glycoproteins by enriching bacterial extracts with glycosyltransferase enzymes capable of the requisite transformations. This would be particularly powerful in combination with microfluidic technologies that enable spatiotemporal enzyme arrangement to mimic glycan processing in the eukaryotic golgi. Cell-free protein synthesis and glycosylation has also been used in compelling demonstrations to rapidly characterize the sequence specificity of oligosaccharyltransferase (Schoborg et al., 2018) and glycosyltransferase enzymes (Kightlinger et al., 2018). Progress made in cell-free protein synthesis and glycosylation provides compelling justification for the continued development of bacterial glycan synthesis and protein glycosylation pathways.

1.3 Biomanufacturing of natural products in microorganisms

Natural product molecules from plants, fungi, and bacteria are a significant source of modern medicines, such as antibiotics (e.g. penicillin, tetracycline, erythromycin), anti-cancer drugs (e.g. taxol, doxorubicin), and anti-malarials (e.g. quinine, artemisinin), among numerous others (J. W. H. Li & Vederas, 2009). In addition to their frequent use as drug molecules, natural products are heavily used in agriculture (e.g. fungicides, insecticides, and herbicides) (Yan, Liu, Jacobsen, & Tang, 2018) and have potential for use as biofuels and commodity chemicals (Beller, Lee, & Katz, 2015). However, they are often produced in trace amounts in their native hosts and their harvesting can be costly and environmentally unsustainable, often leading to extinction of wild species and requiring large land use for farming (Kotopka, Li, & Smolke, 2018). Reconstitution of natural product pathways into industrial microorganisms could help alleviate these problems and is a major goal of biomanufacturing and metabolic engineering. Prominent examples of natural product pathway reconstitution in microorganisms include production of the anti-malarial compound artemisinin in *S. cerevisiae* (Ro et al., 2006), a precursor of the anti-cancer compound Taxol in *E. coli* (Ajikumar et al., 2010), the opioid compounds thebaine and hydrocodone in *S. cerevisiae* (Galanie, Thodey, Trenchard, Interrante, & Smolke, 2015), and recently, the production of cannabinoid compounds in *S. cerevisiae* (Luo et al., 2019). In addition to increasing the scalability, cost-effectiveness, and sustainability of their production, these efforts could also provide a means to generate tailored analogs of natural products with greater therapeutic efficacy or otherwise improved properties (J. W. H. Li & Vederas, 2009).

1.4 Metabolic engineering strategies for scalable biomanufacturing

The above examples represent only a subset of the diversity of molecules and materials that can be produced using biomanufacturing, with many more targets of interest in the areas of therapeutics, foods, feeds, bulk chemicals, fuels, materials, and beyond (Nielsen & Keasling, 2016). In all cases, strategies for optimizing their titer, rate, yield, and purity through genetic engineering are necessary to create scalable production platforms. Towards this end, many strategies have been developed that can be generally categorized into three categories (**Fig. 1-2**): (1) selection and engineering of enzymes that can efficiently produce the desired molecules and intermediates, (2) efforts to engineer host metabolism in ways that maximize flux towards the pathway of interest, and (3) efforts to tune expression of pathway enzymes to minimize the burden of heterologous pathway enzymes and toxic intermediates while simultaneously maximizing flux through the desired pathway. Many such strategies, specifically for the workhorse microbe *E. coli,* are described in this section.

In the first category of pathway design (**Fig. 1-2B)**, many strategies have been pursued to discover and engineer appropriate pathway enzymes. These include porting complete pathways from a non-model organism into genetically tractable organisms, such as *E. coli* (Wacker et al., 2002)*,* synthetic metagenomic screening of enzymes to identify optimal enzyme homologs that improve natural pathways (Bayer et al., 2009), and establishing *de novo* pathways of enzymes from a variety of organisms (Sheppard, Kunjapur, Wenck, & Prather, 2014). Additional strategies include rational design of enzymes to increase their catalytic rate and improve or modify their selectivity (Bonk, Tarasova, Hicks, Tidor, & Prather, 2018; Dietrich et al., 2009) as well as directed evolution

of pathway enzymes for improved properties (Fasan, Chen, Crook, & Arnold, 2007; Ollis, Zhang, Fisher, & Delisa, 2014).

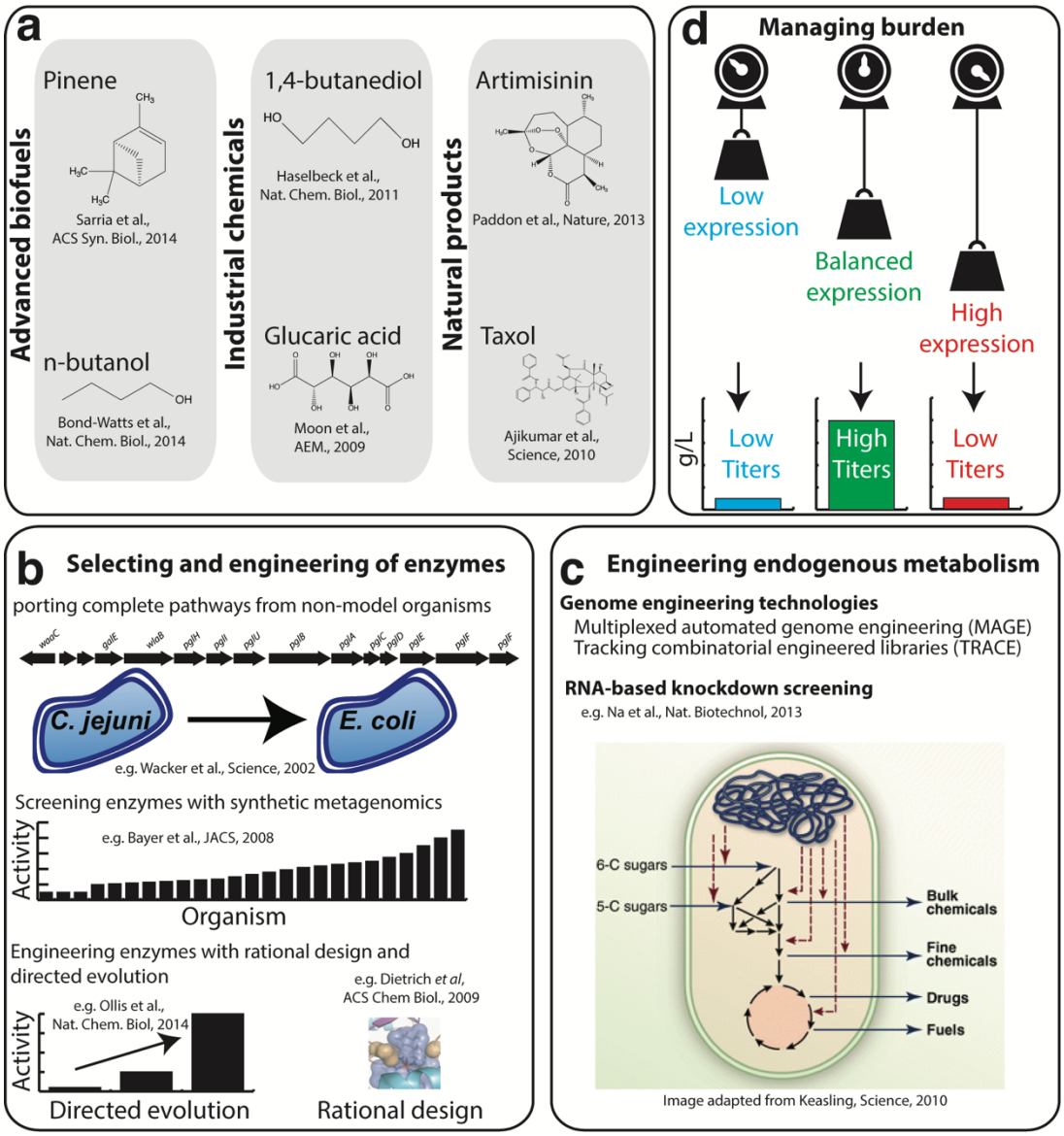


Figure 1-2. Strategies and tools for engineering biomanufacturing strains.

(A) Metabolic engineering holds great promise for the biomanufacturing of numerous compounds from biofuels and industrial chemicals to natural products. Strategies to engineer biomanufacturing strains broadly include three main categories. **(B)** Selection and engineering of appropriate enzymes, which can involve porting pathways from non-

model organisms (e.g. (Wacker et al., 2002)), screening of enzymes with synthetic metagenomics (e.g. (Bayer et al., 2009)), as well as directed evolution and rational design of enzymes for appropriate function (e.g. Ollis et al. (Ollis et al., 2014) and Dietrich et al. (Dietrich et al., 2009)). **(C)** Engineering endogenous metabolism using genome engineering tools such as MAGE (H. H. Wang et al., 2009) and TRACE (Zeitoun et al., 2015) or RNA-based knockdown screening approaches (Na et al., 2013). (**D)** A third strategy for improving biomanufacturing strain performance is to balance pathway enzyme expression levels to avoid overaccumulation of toxic intermediates and minimize the burden of overexpressing heterologous enzymes.

In the second category of engineering host metabolism (**Fig. 1-2C**), many genome engineering and screening technologies have been developed to facilitate the process of engineering endogenous metabolism to maximize flux towards the pathway of interest. In many cases, this involves determining genes that may be directing flux away from the pathway and downregulation or deletion of those genes; or determination of enzymes that produce precursors to the pathway that need to be overexpressed. The determination of appropriate genes often involves simple selection by hand through an understanding of the host's metabolism, an approach that was used in Chapter 2 to optimize the pathway for eukaryotic protein glycosylation (Glasscock et al., 2018), computational methods to predict flux through various knockouts in central carbon metabolism, an approach used in a co-authored manuscript to predict knockouts in a bacterial glycan biosynthesis pathway (Wayman, Glasscock, Mansell, DeLisa, and Varner, 2019), and high-throughput screening of gene knockout or knockdown libraries. In the final case, rapid and high-

throughput technologies are often utilized, such as multiplexed automated genome engineering (MAGE) (H. H. Wang et al., 2009), tracking combinatorial engineered libraries (TRACE) (Zeitoun et al., 2015), or RNA-based knockdown screening technologies (Kim et al., 2016; Na et al., 2013).

In the third category of managing burden in heterologous hosts (**Fig. 1-2D)**, many efforts have been made to engineer and optimize the expression levels of pathway enzymes. The concentration levels of these enzymes are a particularly important factor in determining pathway throughput that can be modulated by tuning the basic gene expression processes within a cell. This is important because flux imbalances in the cell can result in the accumulation of toxic gene products and intermediates, depletion of metabolites required for normal cell growth, and ultimately cause undue burden on the host cell. Fortunately, a focus of many synthetic biology efforts has been to develop genetic control elements that facilitate the search for balance between pathway expression and cell viability in metabolic engineering. These elements include synthetic promoters that vary the amount of mRNA transcribed from DNA (Alper, Fischer, Nevoigt, & Stephanopoulos, 2005; Jensen & Hammer, 1998), an approach used in Chapter 2 to optimize the pathway for eukaryotic protein glycosylation (Glasscock et al., 2018); 5' untranslated regions (UTR's) that differentially modulate transcription termination, mRNA degradation, and translation initiation (Pfleger, Pitera, Smolke, & Keasling, 2006); or ribosome-binding site (RBS) sequences that directly control translation initiation (Barrick et al., 1994; Mutalik et al., 2013; Salis, Mirsky, & Voigt, 2009). In recent years, dynamic control of pathway expression has become a popular area of research, which can allow a greater balance between cell growth and enzyme expression using timing control and

feedback regulation of enzyme expression (**Fig. 1-3).** A few examples of dynamic control in biomanufacturing include quorum-sensing-based enzyme control (Doong, Gupta, & Prather, 2018; Gupta, Reizman, Reisch, & Prather, 2017) and expression systems responsive to pathway intermediates and cellular stress (Dahl et al., 2013; Zhang, Carothers, & Keasling, 2012).

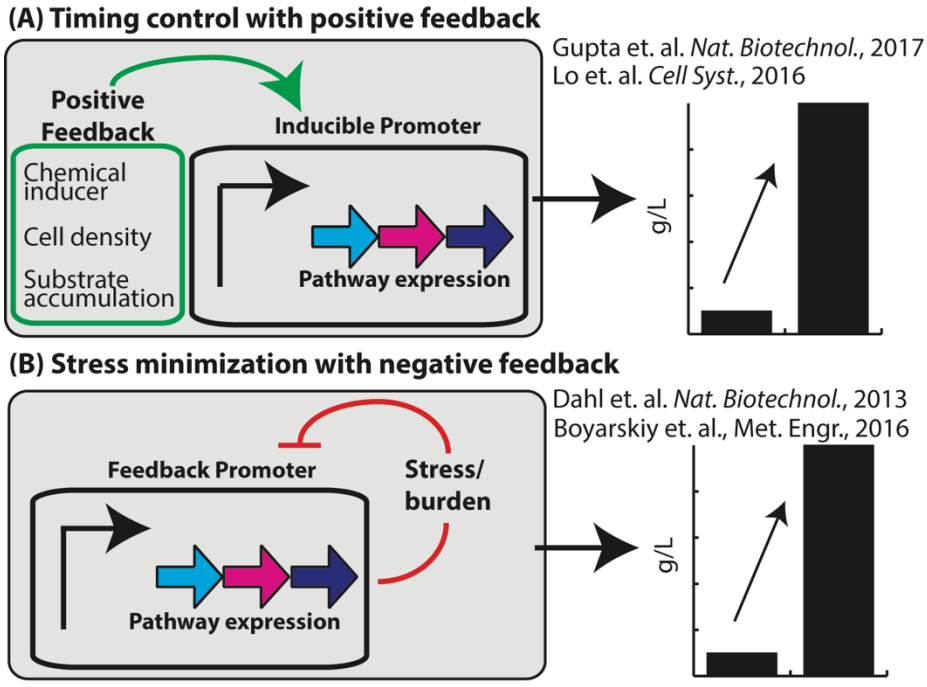


Figure 1-3. Dynamic metabolic pathway regulation.

Strategies for dynamic metabolic pathway regulation can be broadly categorized into two categories: **(A)** timing control with positive feedback regulation (examples include Gupta *et al*., 2017 and Lo *et al.,* 2016 (Gupta et al., 2017; Lo, Chng, Teo, Cho, & Chang, 2016) and **(B)** adaptive regulation by negative feedback, including representation examples by Boyarskiy *et al*., 2016 and Dahl *et al.,* 2013 (Boyarskiy, Davis López, Kong, & Tullman-Ercek, 2016; Dahl et al., 2013). These systems enable introduction or removal of pathway enzymes at optimal times and adaptive feedback regulation of enzymes to produce them at the optimal concentration in the fermentation that can change over time.

1.5 Controlling gene expression with RNA regulators

RNA has become an increasingly versatile substrate for controlling gene expression with metabolic engineering applications in many organisms, especially in bacteria (Chappell et al., 2013; Chappell, Watters, Takahashi, & Lucks, 2015; Cho, Haning, & Contreras, 2015). This is due to its ability to control nearly all aspects of gene expression from mRNA degradation to transcription and translation (**Fig. 1-4**). These points of control can be mediated through both cis- and trans-acting regulators of (1) mRNA stability, such as ribozymes (Collins, Irnov, Baker, & Winkler, 2007) and small RNAs (sRNAs) (Storz, Opdyke, & Zhang, 2004); (2) transcription, such as transcriptional riboswitches (Baker et al., 2012) and sRNA-regulated intrinsic transcription terminators (Brantl & Wagner, 2002); and (3) translation, such as translational riboswitches (J. X. Wang, Lee, Morales, Lim, & Breaker, 2008) and sRNAs (Desnoyers, Bouchard, & Massé, 2013) that often cause structural changes in the 5' UTR to modulate accessibility of the shine-dalgarno sequence to the ribosome. Recently, many synthetic RNA regulators have been designed for versatile control of these aspects of gene expression. These include synthetic RNAs based on natural sequences (Chappell, Takahashi, & Lucks, 2015; Lucks, Qi, Mutalik, Wang, & Arkin, 2011; Mutalik, Qi, Guimaraes, Lucks, & Arkin, 2012; Takahashi & Lucks, 2013; Westbrook & Lucks, 2017) and *de novo* designed RNAs that can activate and repress transcription and translation (Carlson, Glasscock, & Lucks, 2018; Chappell, Westbrook, Verosloff, & Lucks, 2017; Green, Silver, Collins, & Yin, 2014).

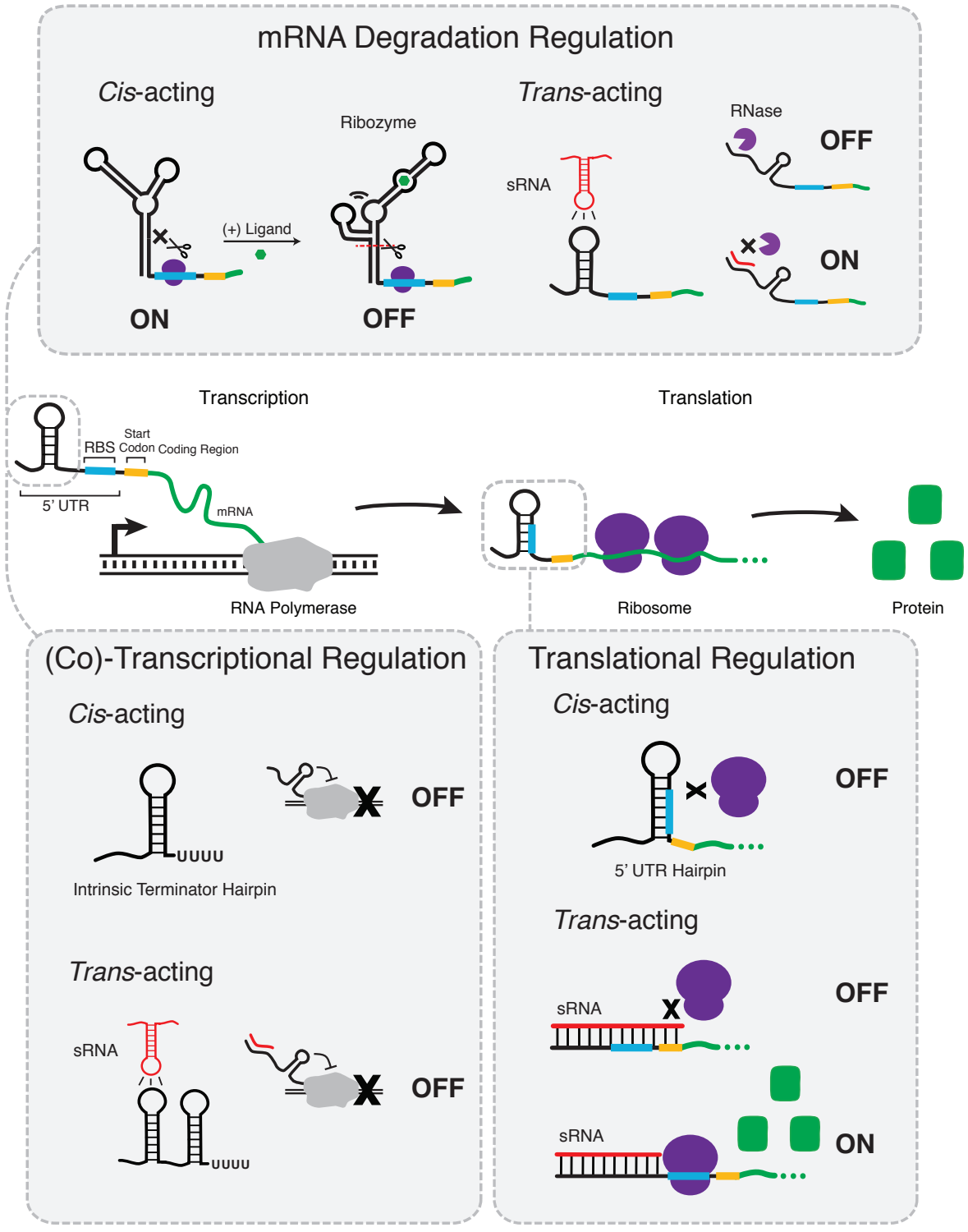


Figure 1-4. RNA regulation of gene expression.

A schematic showing how naturally occurring bacterial RNAs govern gene expression by regulating transcription, mRNA degradation, and translation. Figure adapted from works by Chappell et al. (Chappell et al., 2013; Chappell, Watters, et al., 2015).

1.6 Genetic tools and approaches for cellular engineering of metabolism and metabolic pathways

Motivated by the continuing need for new tools that optimize endogenous metabolism and metabolic pathways for biomanufacturing, this work sought to translate recent advances in synthetic biology into new techniques for metabolic pathway engineering in a variety of contexts. Novel technologies for metabolic engineering are particularly important because of the painstakingly long development times and costly efforts required for development of proof-of-principle biomanufacturing strains with modern approaches; this is exemplified by prominent examples such as the anti-malarial drug artemisinin (10 years, USD50M) and 1,3-propanediol (15 years, USD130M) (Nielsen & Keasling, 2016).

1.6.1 Chapter 2: A flow cytometric approach to engineering E. coli for improved eukaryotic protein glycosylation

Chapter 2 describes efforts to optimize a pathway for eukaryotic protein glycosylation that has relevance in the production of therapeutic glycoproteins. To do this, classical tools in metabolic engineering and a high-throughput flow cytometry screen for glycan biosynthesis were utilized to alleviate bottlenecks in pathway precursor supply and tune the expression levels of pathway enzymes (Glasscock et al., 2018). These efforts

resulted in significantly improved pathway performance in terms of glycan biosynthesis and protein glycosylation efficiency and provided important perspective on some of the most difficult aspect of cellular engineering for biomanufacturing. In addition to what is presented in Chapter 2, several offshoots of this work was contributed to a publication towards cell-free synthesis of eukaryotic glycoproteins (Jaroentomeechai et al., 2018), a review of metabolic engineering of glycoprotein biosynthesis in bacteria (Natarajan, Jaroentomeechai, Li, Glasscock, & DeLisa, 2018), and a first-author review on engineered protein machines with potential applications in bacterial protein glycosylation (Glasscock, Lucks, & DeLisa, 2016).

1.6.2 Chapter 3: High-throughput strain engineering with RNA regulators of gene expression

Chapter 3 describes efforts to utilize RNA translational repressors in development of a new technology for high-throughput strain engineering by rapid knockdown of endogenous gene expression. These efforts would ultimately provide a platform for screening endogenous genes for bottlenecks in metabolic pathways and enable identification of smaller sets of genes for combinatorial genome engineering, with technologies such as MAGE (Wang et al., 2009) or TRACE (Zeitoun et al., 2015). While this work was not fully published, several elements were published in a co-authored manuscript by Carlson *et al.,* (Carlson et al., 2018) and a blueprint for future efforts in development of this platform is provided.

1.6.3 Chapter 4: Dynamic control of pathway expression with riboregulated switchable feedback promoters

Engineered of microbial hosts for biomanufacturing often causes cellular toxicity due to intermediate metabolite accumulation and burdensome enzyme levels, presenting a need for strategies that balance pathway expression and cell fitness. Dynamic pathway regulation has emerged as a promising strategy in metabolic engineering for improved productivity and yield and continues to grow in sophistication (Brockman & Prather, 2015; Holtz & Keasling, 2010; Tan & Prather, 2017). Bacterial stress-response promoters allow dynamic gene regulation using the host's natural transcriptional networks (Boyarskiy et al., 2016; Dahl et al., 2013) but lack the flexibility to control the expression timing and overall magnitude of pathway genes.

In this chapter, a strategy was developed that uses RNA transcriptional regulators to introduce another layer of control over the output of natural stress-response promoters. This new class of gene expression cassette, called a riboregulated switchable feedback promoter (rSFP), can be modularly activated using a variety of mechanisms, from manual induction to quorum sensing. rSFPs were applied to enable optimally timed pathway expression with feedback regulation in a system for production of the Taxol precursor taxadiene-5α-ol by regulating expression of a toxic P450 enzyme. It is anticipated that this approach will enable the construction of more efficient biomanufacturing hosts and become a valuable addition to the synthetic biology and biomanufacturing toolkits.

1.6.4 Chapter 5: Engineering riboregulated switchable stabilized promoters for constant gene expression at any copy number in bacteria

Inspired by the ability of rSFPs (Chapter 4) to controllably gate transcription from stress-response promoters, Chapter 5 describes efforts to extend the rSFP concept towards other engineered promoter systems, specifically a recently published system, called a stabilized promoter, that enables constant gene expression regardless of copy number (Segall-Shapiro, Sontag, & Voigt, 2018). Based on concepts from control theory, this system uses an incoherent feed forward loop to buffer transcription from changes in copy number, such as moving a gene from plasmid-based expression to the genome. This is particularly valuable when it is convenient to prototype and tune metabolic pathway expression on a plasmid and later integrate into the genome without altering optimal expression. While a powerful tool, the architecture of stabilized promoters prohibits their control with inducible promoter systems commonly used in metabolic engineering. Here, the concept of gating promoter output with small-transcription activating RNAs (STARs) was applied to create novel riboregulated switchable stabilized promoters (rSSPs).

1.6.5 Chapter 6: Conclusions and perspectives

As demonstrated in this work, new tools from synthetic biology hold great promise in the development of novel approaches for engineering metabolism and metabolic pathways. This work describes successful efforts in improving a pathway for eukaryotic protein glycosylation in *E. coli* using conventional tools from metabolic engineering. Inspired by this early success, efforts were made to develop several novel technologies

that hold promise in engineering cells for biomanufacturing, including progress towards an RNA-based platform for high-throughput strain engineering, riboregulated switchable feedbacks promoters (rSFPs) for developing stress-responsive expression systems, and riboregulated switchable stabilized promoters (rSSPs) that introduce inducible functionality to previously developed stabilized promoters. It is expected that future research will further develop and enhance these technologies to aid in the development of sustainable and cost-effective biomanufacturing platforms.

1.7 References

Ajikumar, P. K., Xiao, W. H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., … Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science*. https://doi.org/10.1126/science.1191652

Alper, H., Fischer, C., Nevoigt, E., & Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0504604102

Baker, J. L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R. B., & Breaker, R. R. (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*. https://doi.org/10.1126/science.1215063

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*. https://doi.org/10.1016/S0958-1669(99)00003-8

Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., … Stormo, G. D. (1994). Quantitative analysis of ribosome binding sites in E.coli. *Nucleic Acids Research*. https://doi.org/10.1093/nar/22.7.1287

Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., & Voigt, C. A. (2009). Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja809461u

Beller, H. R., Lee, T. S., & Katz, L. (2015). Natural products as biofuels and bio-based chemicals: Fatty acids and isoprenoids. *Natural Product Reports*. https://doi.org/10.1039/c5np00068h

Bonk, B. M., Tarasova, Y., Hicks, M. A., Tidor, B., & Prather, K. L. J. (2018). Rational design of thiolase substrate specificity for metabolic engineering applications. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.26737

Boyarskiy, S., Davis López, S., Kong, N., & Tullman-Ercek, D. (2016). Transcriptional feedback regulation of efflux protein expression for increased tolerance to and production of n-butanol. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2015.11.005

Brantl, S., & Wagner, E. G. H. (2002). An antisense RNA-mediated transcriptional attenuation mechanism functions in Escherichia coli. *Journal of Bacteriology*. https://doi.org/10.1128/JB.184.10.2740-2747.2002

Brockman, I. M., & Prather, K. L. J. (2015). Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400422

Carlson, P. D., Glasscock, C. J., & Lucks, J. B. (2018). De novo Design of Translational RNA Repressors. *BioRxiv*. https://doi.org/10.1101/501767

Chappell, J., Takahashi, M. K., & Lucks, J. B. (2015). Creating small transcription activating RNAs. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1737

Chappell, J., Takahashi, M. K., Meyer, S., Loughrey, D., Watters, K. E., & Lucks, J. (2013). The

centrality of RNA for engineering gene expression. *Biotechnology Journal*. https://doi.org/10.1002/biot.201300018

Chappell, J., Watters, K. E., Takahashi, M. K., & Lucks, J. B. (2015). A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2015.05.018

Chappell, J., Westbrook, A., Verosloff, M., & Lucks, J. B. (2017). Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. *Nature Communications*. https://doi.org/10.1038/s41467-017-01082-6

Cho, S. H., Haning, K., & Contreras, L. M. (2015). Strain engineering via regulatory noncoding RNAs: Not a one-blueprint-fits-all. *Current Opinion in Chemical Engineering*. https://doi.org/10.1016/j.coche.2015.07.008

Collins, J. A., Irnov, I., Baker, S., & Winkler, W. C. (2007). Mechanism of mRNA destabilization by the glmS ribozyme. *Genes and Development*. https://doi.org/10.1101/gad.1605307

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., … Keasling, J. D. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2689

Desnoyers, G., Bouchard, M. P., & Massé, E. (2013). New insights into small RNA-dependent translational regulation in prokaryotes. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.10.004

Dietrich, J. A., Yoshikuni, Y., Fisher, K. J., Woolard, F. X., Ockey, D., McPhee, D. J., … Keasling, J. D. (2009). A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450BM3. *ACS Chemical Biology*.

https://doi.org/10.1021/cb900006h

Dimitrov, D. S. (2012). Therapeutic proteins. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-921-1_1

Doong, S. J., Gupta, A., & Prather, K. L. J. (2018). Layered dynamic regulation for improving metabolic pathway productivity in Escherichia coli . *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1716920115

Fasan, R., Chen, M. M., Crook, N. C., & Arnold, F. H. (2007). Engineered alkane-hydroxylating cytochrome P450BM3 exhibiting nativelike catalytic properties. *Angewandte Chemie - International Edition*. https://doi.org/10.1002/anie.200702616

Friedman, B., Vaddi, K., Preston, C., Mahon, E., Cataldo, J. R., & McPherson, J. M. (1999). A comparison of the pharmacological properties of carbohydrate remodeled recombinant and placental-derived beta-glucocerebrosidase: implications for clinical efficacy in treatment of Gaucher disease. *Blood*.

Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., & Smolke, C. D. (2015). Complete synthesis of opiods in yeast. *Science*. https://doi.org/10.1126/science.aac9373

Glasscock, C. J., Lucks, J. B., & DeLisa, M. P. (2016). Engineered Protein Machines: Emergent Tools for Synthetic Biology. *Cell Chemical Biology*. https://doi.org/10.1016/j.chembiol.2015.12.004

Glasscock, C. J., Yates, L. E., Jaroentomeechai, T., Wilson, J. D., Merritt, J. H., Lucks, J. B., & DeLisa, M. P. (2018). A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2018.04.014

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: De-novo-designed

regulators of gene expression. *Cell*. https://doi.org/10.1016/j.cell.2014.10.002

Guarino, C., & Delisa, M. P. (2012). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology*. https://doi.org/10.1093/glycob/cwr151

Gupta, A., Reizman, I. M. B., Reisch, C. R., & Prather, K. L. J. (2017). Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3796

Helenius, A., & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science*. https://doi.org/10.1126/science.291.5512.2364

Herscovics, A. (1999). Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochimica et Biophysica Acta - General Subjects*. https://doi.org/10.1016/S0304-4165(99)00171-3

Hockney, R. C. (1994). Recent developments in heterologous protein production in Escherichia coli. *Trends in Biotechnology*. https://doi.org/10.1016/0167-7799(94)90021-3

Holtz, W. J., & Keasling, J. D. (2010). Engineering Static and Dynamic Control of Synthetic Pathways. *Cell*. https://doi.org/10.1016/j.cell.2009.12.029

Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., … Delisa, M. P. (2018). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nature Communications*. https://doi.org/10.1038/s41467-018-05110-x

Jensen, P. R., & Hammer, K. (1998). The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and Environmental Microbiology*.

Jervis, A. J., Butler, J. A., Lawson, A. J., Langdon, R., Wren, B. W., & Linton, D. (2012). Characterization of the structurally diverse N-linked glycans of Campylobacter species.

Journal of Bacteriology. https://doi.org/10.1128/JB.00042-12

Joseph A. Wayman, Cameron J. Glasscock, Thomas J. Mansell, Matthew P. DeLisa, J. D. V. (2019). Improving designer glycan production in Escherichia coli through model-guided metabolic engineering. *Metabolic Engineering Communications*. https://doi.org/https://doi.org/10.1016/j.mec.2019.e00088

Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science*. https://doi.org/10.1126/science.1193990

Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., Delisa, M. P., Mrksich, M., & Jewett, M. C. (2018). Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases article. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-018-0051-2

Kim, S. K., Han, G. H., Seong, W., Kim, H., Kim, S. W., Lee, D. H., & Lee, S. G. (2016). CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2016.08.006

Kotopka, B. J., Li, Y., & Smolke, C. D. (2018). Synthetic biology strategies toward heterologous phytochemical production. *Natural Product Reports*. https://doi.org/10.1039/c8np00028j

LaTemple, D. C., Abrams, J. T., Zhang, S. Y., & Galili, U. (1999). Increased immunogenicity of tumor vaccines complexed with anti-gal: Studies in knockout mice for α1,3galactosyltransferase. *Cancer Research*.

Li, C., & Wang, L. X. (2018). Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chemical Reviews*. https://doi.org/10.1021/acs.chemrev.8b00238

Li, J. W. H., & Vederas, J. C. (2009). Drug discovery and natural products: End of an era or an endless frontier? *Science*. https://doi.org/10.1126/science.1168243

Lo, T. M., Chng, S. H., Teo, W. S., Cho, H. S., & Chang, M. W. (2016). A Two-Layer Gene

Circuit for Decoupling Cell Growth from Metabolite Production. *Cell Systems*. https://doi.org/10.1016/j.cels.2016.07.012

Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., & Arkin, A. P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1015741108

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., … Keasling, J. D. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*. https://doi.org/10.1038/s41586-019-0978-9

Macdougall, I. C. (2003). Optimizing the use of erythropoietic agents - pharmacokinetic and pharmacodynamic considerations. *Nephrology Dialysis Transplantation*. https://doi.org/10.1093/ndt/17.suppl_5.66

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2019). Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-018-0168-3

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., … Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*. https://doi.org/10.1038/nmeth.2404

Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., & Arkin, A. P. (2012). Rationally designed families of orthogonal RNA regulators of translation. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.919

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., & Lee, S. Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2461

Natarajan, A., Jaroentomeechai, T., Li, M., Glasscock, C. J., & DeLisa, M. P. (2018). Metabolic engineering of glycoprotein biosynthesis in bacteria. *Emerging Topics in Life Sciences*. https://doi.org/10.1042/ETLS20180004

Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*. https://doi.org/10.1016/j.cell.2016.02.004

Ollis, A. A., Zhang, S., Fisher, A. C., & Delisa, M. P. (2014). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1609

Pfleger, B. F., Pitera, D. J., Smolke, C. D., & Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*. https://doi.org/10.1038/nbt1226

Ro, D. K., Paradise, E. M., Quellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., … Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*. https://doi.org/10.1038/nature04640

Rothman, R. J., Perussia, B., Herlyn, D., & Warren, L. (1989). Antibody-dependent cytotoxicity mediated by natural killer cells is enhanced by castanospermine-induced alterations of IgG glycosylation. *Molecular Immunology*. https://doi.org/10.1016/0161-5890(89)90055-2

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1568

Schoborg, J. A., Hershewe, J. M., Stark, J. C., Kightlinger, W., Kath, J. E., Jaroentomeechai, T., … Jewett, M. C. (2018). A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnology and Bioengineering*.

https://doi.org/10.1002/bit.26502

Schwarz, F., Huang, W., Li, C., Schulz, B. L., Lizak, C., Palumbo, A., … Wang, L. X. (2010). A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.314

Segall-Shapiro, T. H., Sontag, E. D., & Voigt, C. A. (2018). Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4111

Sethuraman, N., & Stadheim, T. A. (2006). Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2006.06.010

Sheppard, M. J., Kunjapur, A. M., Wenck, S. J., & Prather, K. L. J. (2014). Retro-biosynthetic screening of a modular pathway design achieves selective route for microbial synthesis of 4-methyl-pentanol. *Nature Communications*. https://doi.org/10.1038/ncomms6031

Sola, R. J., & Griebenow, K. (2009). Effects of glycosylate on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*. https://doi.org/10.1002/jps.21504

Storz, G., Opdyke, J. A., & Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Current Opinion in Microbiology*. https://doi.org/10.1016/j.mib.2004.02.015

Szymanski, C. M., Ruijin, Y., Ewing, C. P., Trust, T. J., & Guerry, P. (1999). Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.1999.01415.x

Takahashi, M. K., & Lucks, J. B. (2013). A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt452

Tan, S. Z., & Prather, K. L. (2017). Dynamic pathway regulation: recent advances and methods of construction. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2017.10.004

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., … DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.921

Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., … Aebi, M. (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science*. https://doi.org/10.1126/science.298.5599.1790

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. https://doi.org/10.1038/nature08187

Wang, J. X., Lee, E. R., Morales, D. R., Lim, J., & Breaker, R. R. (2008). Riboswitches that Sense S-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2008.01.012

Weerapana, E., & Imperiali, B. (2006). Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology*. https://doi.org/10.1093/glycob/cwj099

Westbrook, A. M., & Lucks, J. B. (2017). Achieving large dynamic range control of gene expression with a compact RNA transcription-translation regulator. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx215

Wormald, M. R., & Dwek, R. A. (1999). Glycoproteins: Glycan presentation and protein-fold stability. *Structure*. https://doi.org/10.1016/S0969-2126(99)80095-1

Yan, Y., Liu, Q., Jacobsen, S. E., & Tang, Y. (2018). The impact and prospect of natural product

discovery in agriculture. *EMBO Reports*. https://doi.org/10.15252/embr.201846824

Zeitoun, R. I., Garst, A. D., Degen, G. D., Pines, G., Mansell, T. J., Glebes, T. Y., … Gill, R. T. (2015). Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3177

Zhang, F., Carothers, J. M., & Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2149

CHAPTER 2 - *A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation*[1]

2.1 Abstract

A synthetic pathway for production of the eukaryotic trimannosyl chitobiose glycan (mannose$_3$-*N*-acetylglucosamine$_2$, $Man_3GlcNAc_2$) and its transfer to specific asparagine residues in target proteins was previously engineered in *Escherichia coli*, providing this simple microbe with the ability to perform a complex post-translational protein modification. Here, we leveraged a flow cytometric fluorescence-based assay to improve $Man_3GlcNAc_2$ glycan biosynthesis in *E. coli* cells. Specifically, pathway improvements were identified, including reducing pathway enzyme expression levels and overexpressing nucleotide sugar biosynthesis genes, which enhanced production of lipid-linked $Man_3GlcNAc_2$ by nearly 50-fold to 13.9 μg/L. In turn, cells producing higher levels of the $Man_3GlcNAc_2$ substrate yielded up to 14 times more glycosylated acceptor protein (to ~14 mg/L) than their non-optimized counterparts. These results demonstrate the use of flow cytometry screening as a powerful tool for interrogating the surfaces of glyco-engineered bacteria and identifying meaningful improvements in glycan biosynthesis. We anticipate this approach will enable further optimization of bacterial glycan biosynthesis pathways using new strain engineering tools from metabolic engineering and synthetic biology.

---

[1] The results presented in this chapter have been published in the journal Metabolic Engineering:

Glasscock, C. J., Yates, L. E., Jaroentomeechai, T., Wilson, J. D., Merritt, J. H., Lucks, J. B., & DeLisa, M. P. (2018). A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation. *Metabolic Engineering*, *47*, 488–495. http://doi.org/10.1016/j.ymben.2018.04.014

2.2 Introduction

Asparagine-linked (*N*-linked) glycosylation is a widespread post-translational modification that affects the fold and function of proteins (Helenius & Aebi, 2001; Wormald & Dwek, 1999; Xu & Ng, 2015). *N*-linked glycosylation is also important for therapeutic protein development, since many proteins must be glycosylated to function as intended (Dalziel, Crispin, Scanlan, Zitzmann, & Dwek, 2014). Therapeutic glycoproteins are typically produced with eukaryotic expression systems that natively possess the machinery for human-like *N*-linked glycosylation, but can be difficult to manipulate (Sethuraman & Stadheim, 2006). This difficulty is further compounded by the fact that, despite many impressive achievements (S. R. Hamilton et al., 2003; Meuris et al., 2014; Yang et al., 2015), most *N*-linked glycoprotein expression platforms produce heterologous products containing multiple glycoforms (Mimura et al., 2018; Rich & Withers, 2009; Rudd & Dwek, 1997; Zhang et al., 2016). Glycosylation is not limited to eukaryotes, however, and *N*-linked glycosylation pathways can be found in many proteobacteria (Abu-Qarn, Eichler, & Sharon, 2008) and functionally transferred into Escherichia coli(Wacker et al., 2002). This latter development has paved the way for studying and engineering the protein glycosylation mechanism in this genetically tractable host (Feldman et al., 2005) and has potential for unlocking the use of robust prokaryotic molecular biology tools to achieve greater control over N-linked glycosylation (Baker, Çelik, & DeLisa, 2013; Keys & Aebi, 2017; Merritt, Ollis, Fisher, & Delisa, 2013; Terra et al., 2012).

Previous glycoengineering efforts in *E. coli* led to the development of a synthetic pathway that enables site-specific glycosylation of proteins with mannose$_3$-*N*-acetylglucosamine$_2$ ($Man_3GlcNAc_2$) (Valderrama-Rincon et al., 2012) a structure that

comprises the core of all human N-linked glycans (Song, Aldredge, & Lebrilla, 2015). This heterologous pathway is comprised of multiple glycosyltransferases (GTases) from yeast and the oligosaccharyltransferase (OTase), PglB, from *Campylobacter jejuni,* an archetype for bacterial *N*-linked glycosylation. The $Man_3GlcNAc_2$ protein glycosylation pathway can be divided into three distinct stages: glycan biosynthesis, membrane translocation of glycans, and glycan transfer onto polypeptide acceptor sequons (**Fig. 2-1**). In the first stage, $Man_3GlcNAc_2$ glycans are sequentially assembled in the form of undecaprenyl-pyrophosphate (Und-PP) lipid-linked oligosaccharides (LLOs) on the cytoplasmic face of the inner membrane. This involves four heterologous GTase enzymes from *S. cerevisiae* (*ALG13*, *ALG14*, *ALG1* and *ALG2*) that catalyze the sequential addition of nucleotide-activated sugars and are encoded on the arabinose-inducible plasmid pYCG. In the subsequent glycan-flipping stage, the native *E. coli* flippase enzyme Wzx translocates newly synthesized LLOs to the periplasmic face of the inner membrane. Finally, for glycan transfer, the OTase enzyme PglB from *C. jejuni* transfers the glycan from its lipid anchor to an asparagine residue on the periplasmically expressed substrate protein.

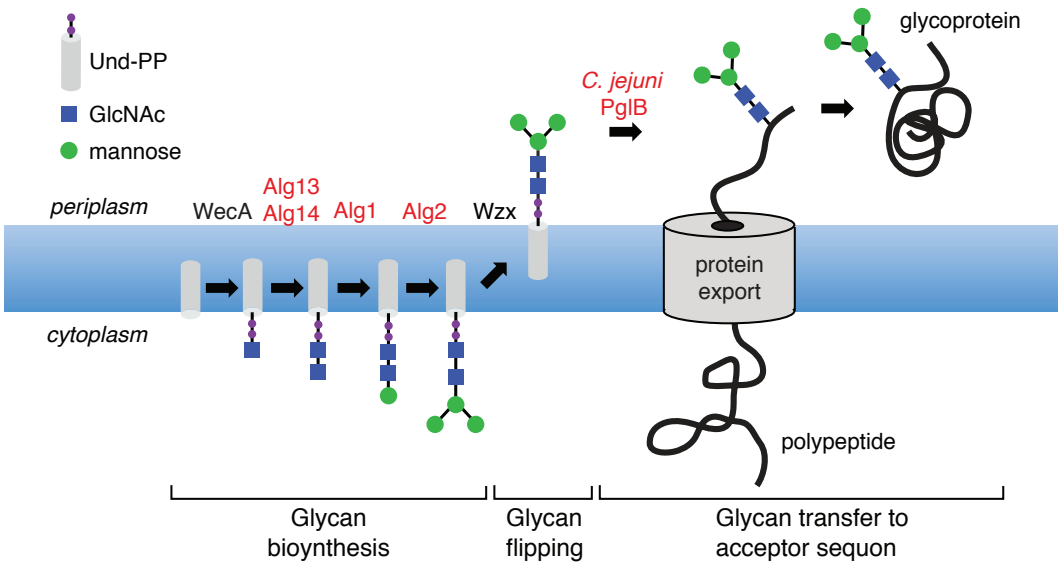


Figure 2-1. The pathway for eukaryotic N-linked protein glycosylation in *E. coli.*

The $Man_3GlcNAc_2$ protein glycosylation pathway can be divided into three distinct stages: glycan biosynthesis, membrane translocation of glycans, and glycan transfer onto polypeptide acceptor sequons. In the first stage, $Man_3GlcNAc_2$ glycans are assembled on the lipid carrier Und-PP on the cytoplasmic face of the inner membrane by yeast GTases Alg13, Alg14, Alg1, and Alg2, along with endogenous *E. coli* enzyme WecA. In the second stage, lipid-linked $Man_3GlcNAc_2$ glycans are flipped to the periplasmic side of the inner membrane by the action of the flippase Wzx. In the third stage, the OTase PglB transfers $Man_3GlcNAc_2$ glycans from Und-PP to asparagine residues of acceptor proteins that are exported into the periplasm. Protein glycosylation can occur in a co- or post-translational manner, involving acceptor site sequences having a consensus of D/E-X_1-N-X_2-S/T (where X_1 and X_2 can be any amino acid but proline).

Despite successful eukaryotic protein glycosylation, this initial report resulted in a low fraction of glycosylated protein (<1%) and yield of 50 μg/L that was difficult to detect

without extensive purification. We suspected that this inefficient glycosylation (relative to >50% glycosylation efficiency often observed in other prokaryotic N-linked glycosylation systems (Mills et al., 2016; Ollis et al., 2015)) was due in part to relatively poor accumulation of the lipid-linked $Man_3GlcNAc_2$ substrate during the glycan biosynthesis stage of the engineered pathway. Here, we sought to relieve this bottleneck and improve overall protein glycosylation levels by leveraging a cell-based fluorescence assay to screen for improvements made possible through manipulating enzyme expression and host metabolism. Specifically, we reduced the expression levels of the four heterologous yeast GTases and overexpressed two endogenous biosynthetic enzymes for GDP-mannose precursor synthesis. The combined effect of these changes was a 4-fold increase in cell surface $Man_3GlcNAc_2$ levels that was found to correlate with a nearly 50-fold increase in intracellular UndPP-linked $Man_3GlcNAc_2$. In cells producing higher levels of the $Man_3GlcNAc_2$ substrate, protein glycosylation was also significantly enhanced with yields of glycosylated acceptor protein that were more than 5 times greater than the yields achieved in non-optimized cells. Overall, we expect that our results will provide a robust engineering framework for future efforts to develop efficient eukaryotic protein glycosylation pathways in *E. coli* cells.

2.3 Results

2.3.1 Cell-surface glycan display for screening $Man_3GlcNAc_2$ levels in living cells.

To probe the abundance of $Man_3GlcNAc_2$ LLOs produced in living cells, we employed a flow cytometric assay for cell-surface display of $Man_3GlcNAc_2$ glycans (Valderrama-Rincon et al., 2012) (**Fig. 2-2A**). This assay takes advantage of the fact that

Gram-negative bacterial cell surfaces can have engineered oligosaccharides in their lipopolysaccharide (LPS) layer (Fisher et al., 2011; Ilg, Yavuz, Maffioli, Priem, & Aebi, 2010). Such remodeled LPS depends on the O-antigen ligase WaaL, which transfers periplasmic Und-PP-linked oligosaccharides onto lipid A. These oligosaccharides are shuttled to the cell surface via the LPS transport system, after which they can be conveniently labeled and detected by flow cytometry (Fisher et al., 2011; Ilg et al., 2010).

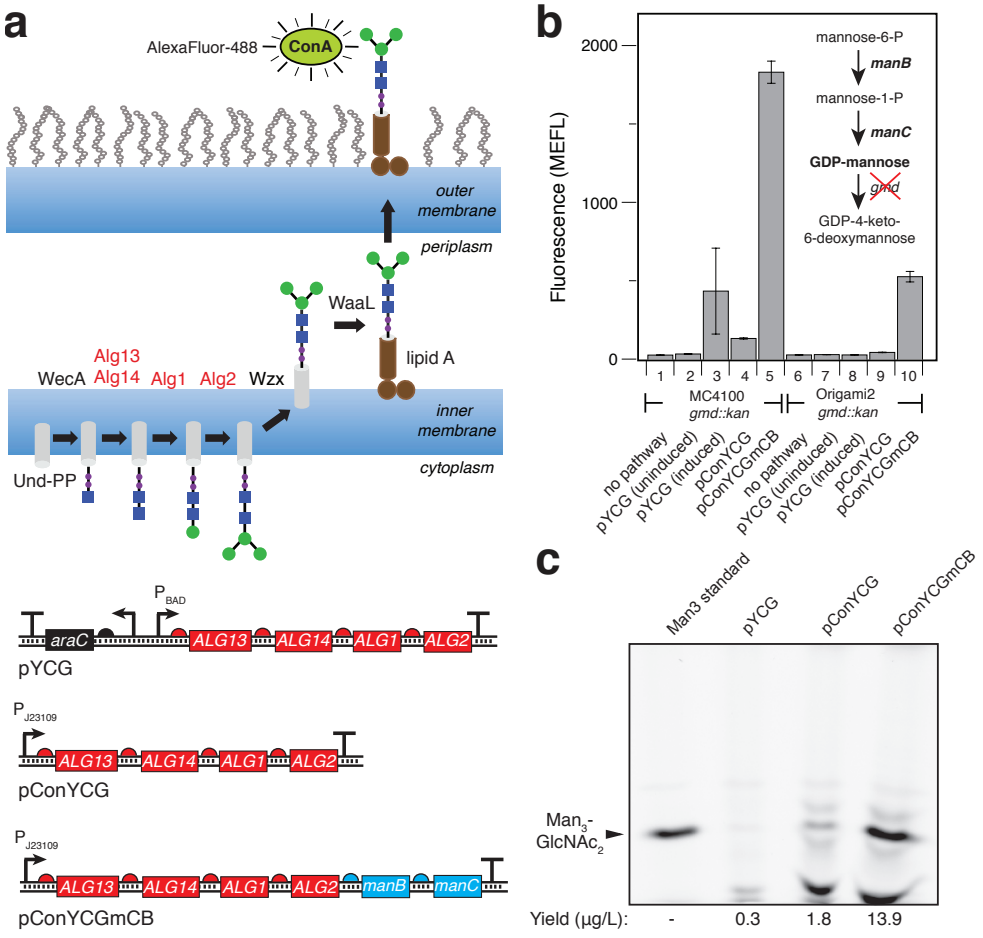


Figure 2-2. Screening of $Man_3GlcNAc_2$ pathway variants with glycan display.

(A) (upper panel) Scheme for flow cytometric analysis of glycan cell-surface display. Cytoplasmic LLOs are a substrate for Wzx-mediated translocation across the inner membrane into the periplasm. Glycans are subsequently transferred to lipid A by the

endogenous O-antigen ligase WaaL and shuttled to the cell surface where they become available for labeling with AlexaFluor-488 conjugated ConA. ConA labeled cells are readily analyzed by flow cytometry. (lower panel) Plasmid architecture for pYCG, pConYCG, and pConYCGmCB. For pYCG, expression of the artificial *ALG13-ALG14-ALG1-ALG2* operon is driven by an arabinose-inducible P_{BAD} promoter. For pConYCG, the P_{BAD} promoter and *araC* repressor gene in pYCG are replaced by a low-strength constitutive promoter J23109 from the Registry of Standard Biological Parts (partsregistry.org). For pConYCGmCB, genes encoding *manB* and *manC* were PCR amplified from *E. coli* MG1655 and cloned into pYCG, just after the *ALG13-ALG14-ALG1-ALG2* pathway operon. **(B)** Mean molecules of equivalent fluorescein (MEFL) for MC4100 *gmd::kan* or Origami2 *gmd::kan* cells containing: no pathway control (strains 1,6); pYCG uninduced (strains 2,7); pYCG induced (strains 3,8); pConYCG (strains 4,9); or pConYCGmCB (strains 5,10). All strains were grown in LB and labeled with ConA before flow cytometric analysis. Strains carrying pYCG were induced at inoculation with 0.2% L-arabinose. Error bars represent standard error of at least three biological replicates. Inset pathway details GDP-mannose biosynthesis by phosphoglucoisomerase (PGI), mannose-6-phosphate isomerase (ManA), phosphomannomutase (ManB), and mannose-1-phosphate guanylyltransferase (ManC). A knockout of GDP-mannose 4,6-dehydratase (GMD) prevents conversion of GDP-mannose into GDP-4-keto-6-deoxymannose. **(C)** FACE gel analysis of $Man_3GlcNAc_2$ glycans produced by MC4100 *gmd::kan* Δ*waaL* cells carrying pYCG, pConYCG, or pConYCGmCB as indicated. Glycans were isolated by extraction of LLOs from each strain followed by acid hydrolysis to release the glycan moiety. An equivalent volume was added to each lane. A total of 0.5

μg of commercial $Man_3GlcNAc_2$ was used as a standard for quantification. Yields of each glycan were estimated by densitometry analysis of individual bands using Image Lab 6.0.

Consistent with previous observations(Valderrama-Rincon et al., 2012), MC4100 *gmd::kan* cells carrying pYCG but not control cells carrying an empty vector were fluorescent upon labeling with fluorophore-conjugated ConA (**Fig. 2-2B**), which preferentially binds internal and non-reducing terminal α-mannose in oligosaccharides. Fluorescence microscopy revealed that ConA binding was clearly localized on the cell surface (**SI Fig. A.7-1**). During these studies, we also noticed that MC4100 *gmd::kan* cells expressing pYCG exhibited variable mean fluorescence intensities across biological replicates and a bimodal fluorescence distribution (**Fig. 2-2B** and **SI Fig. A.7-2A**), indicating genetic instability in the pYCG construct, as well as poor growth following arabinose induction (**SI Fig. A.7-3**). In addition, we noticed that cells carrying pYCG underwent a change in cell morphology (**SI Fig. A.7-2B**)**,** as evidenced by a comparison of forward scatter (FSC) and side scatter (SSC) profiles with cells containing an empty vector, which was indicative of cellular stress.

We speculated that this observed instability was due to high-level expression from the arabinose-inducible P_{BAD} promoter in pYCG that drives expression of the four *S. cerevisiae* GTases (**Fig. 2-2A**). To remedy this issue, we replaced the P_{BAD} promoter in pYCG with the low-strength constitutive promoter pJ23109 from the Anderson promoter library from the Registry of Standard Biological Parts (partsregistry.org) (Kelly et al., 2009), yielding a new plasmid for $Man_3GlcNAc_2$ biosynthesis called pConYCG (**Fig. 2-2A**). It should be noted that we were unable to clone stronger constitutive promoters from

the Registry of Standard Biological Parts without the accumulation of mutations in the pathway. Flow cytometric analysis of cells carrying pConYCG exhibited a lower mean fluorescence intensity compared to cells carrying pYCG; however, these cells displayed a more stable fluorescence phenotype as evidenced by reduced error between samples and a uniform fluorescence distribution (**Fig. 2-2B** and **SI Fig. A.7-2A**). In addition, cells expressing pConYCG showed virtually no change in cell morphology by FSC/SSC profile (**SI Fig. A.7-2B**) and achieved much higher cell densities than their pYCG counterparts (**SI Fig. A.7-3**). Based on the improvements in growth and phenotypic stability, we chose to proceed with pConYCG.

2.3.2 Increasing $Man_3GlcNAc_2$ LLO levels by ManBC overexpression.

Previously, sufficient availability of the Alg1/2 substrate precursor nucleotide-activated sugar guanosine diphosphate-mannose (GDP-mannose) was ensured by knockout of the chromosomal copy of GDP-mannose-4,6-dehydratase (GMD). GMD normally shunts GDP-mannose to GDP-4-keto-6-deoxymannose in the first step of GDP-L fucose synthesis (**Fig. 2B**). In cells lacking GMD (*e.g.*, MC4100 *gmd::kan*), levels of GDP-mannose are sufficiently elevated for producing detectable levels of $Man_3GlcNAc_2$-containing LLOs (Valderrama-Rincon et al., 2012). However, while detectable, these cells produced overall low levels of $Man_3GlcNAc_2$ (~0.3 mg/L) and accumulated incompletely synthesized glycans in the form of $ManGlcNAc_2$ and $Man_2GlcNAc_2$ intermediates (Valderrama-Rincon et al., 2012). We hypothesized here that these deficiencies were caused by suboptimal GDP-mannose levels resulting from low expression of the GDP-mannose biosynthetic enzymes phosphomannomutase (ManB) and mannose-1-

phosphate guanylyltransferase (ManC). The low expression would be expected since the *manB* and *manC* genes are naturally encoded on the heavily-regulated *cps* operon for capsular polysaccharide synthesis and may not be very highly expressed under the conditions used in our study (Gottesman, Trisler, & Torres-Cabassa, 1985).

To alleviate this potential bottleneck, the operon from *E. coli* MG1655 encoding the *manB* and *manC* genes was cloned into the pConYCG pathway to create plasmid pConYCGmCB (**Fig. 2-2A**). MC4100 *gmd::kan* cells transformed with this plasmid were subjected to ConA labeling and flow cytometric screening. In line with our hypothesis, MC4100 *gmd::kan* cells carrying pConYCGmCB exhibited very strong cell fluorescence, which was ~14 times greater than the fluorescence detected for MC4100 *gmd::kan* cells carrying pConYCG (**Fig. 2B**). In addition, we noticed that cells exhibited: (i) a uniform fluorescence distribution that was localized on the cell surface (**SI Figs. A-1 and A-2A**); (ii) a FSC/SSC profile that was nearly identical to negative control cells (**SI Fig. A-2B**); and (iii) robust growth (**SI Fig. A-3**). To investigate the influence of strain background on glycan biosynthesis with pYCG plasmid derivatives, we also considered the strain Origami2 *gmd::kan*, which has been used as a source of eukaryotic $Man_3GlcNAc_2$-containing LLOs for further *in vitro* editing (B. S. Hamilton et al., 2017). Interestingly, flow cytometric analysis of Origami2 *gmd::kan* cells carrying pYCG or pConYCG resulted in no detectable fluorescence above background (**Fig. 2-2B**). However, Origami2 *gmd::kan* cells carrying pConYCGmCB generated a strong fluorescence signal that was comparable to that of MC4100 *gmd::kan* cells with the same plasmid (**Fig. 2-2B**).

To verify that the increase in cell-surface fluorescence corresponded to greater intracellular $Man_3GlcNAc_2$ accumulation, LLOs were extracted from MC4100 *gmd::kan*

Δ*waaL* and Origami2 *gmd::kan* Δ*waaL* cells and analyzed by FACE analysis (Gao, 2005; Gao & Lehrman, 2006). The *waaL* gene encodes an O-antigen ligase that catalyzes the transfer of Und-PP-linked oligosaccharides to lipid A; hence, deletion of this gene is expected to prevent Und-PP-linked $Man_3GlcNAc_2$ from shuttling to the outer membrane, as was shown previously (Valderrama-Rincon et al., 2012). In line with the ConA labeling experiments, the levels of lipid-linked $Man_3GlcNAc_2$ produced by MC4100 *gmd::kan* Δ*waaL* cells carrying pConYCGmCB were ~8- and 46-fold greater than those in MC4100 *gmd::kan* Δ*waaL* cells carrying pConYCG and pYCG, respectively, as estimated by densitometry analysis of individual bands in the FACE gel (**Fig. 2-2C**). Interestingly, LLO levels produced in cells containing pConYCG were ~6-fold higher than cells containing pYCG, in contradiction with flow cytometry results. This may be explained by the substantial difference in cell fitness between the two strains and the variability of pYCG fluorescence observed by flow cytometry (SI Fig. B-3).

To verify the glycan structure produced by these strains, LLOs were extracted and characterized by MALDI-MS analysis. Consistent with previous MALDI-MS spectra (Valderrama-Rincon et al., 2012), MC4100 *gmd::kan* Δ*waaL* cells carrying pYCG produced $Hex_3HexNAc_2$, consistent with the expected $Man_3GlcNAc_2$ glycan, but also accumulated significant amounts of incompletely synthesized glycans in the form of $Hex_1HexNAc_2$ and $Hex_2HexNAc_2$ (**Fig. 2-3A**). In line with the improved glycan levels, MC4100 *gmd::kan* Δ*waaL* cells carrying pConYCG or pConYCGmCB produced $Hex_3HexNAc_2$ as the primary oligosaccharide with only background levels of $Hex_1HexNAc_2$ and $Hex_2HexNAc_2$ glycoforms (**Fig. 2-3B,C**). Hence, pathway expression from the pConYCG or pConYCGmCB plasmids yielded significant levels of $Man_3GlcNAc_2$

glycan at the expense of the undesired intermediate products. Moreover, these findings confirm that the fluorescence screening identified *bona fide* improvements in $Man_3GlcNAc_2$ glycan biosynthesis and reveal the potential of flow cytometry for optimizing the first stage of *N*-linked glycosylation in bacteria.

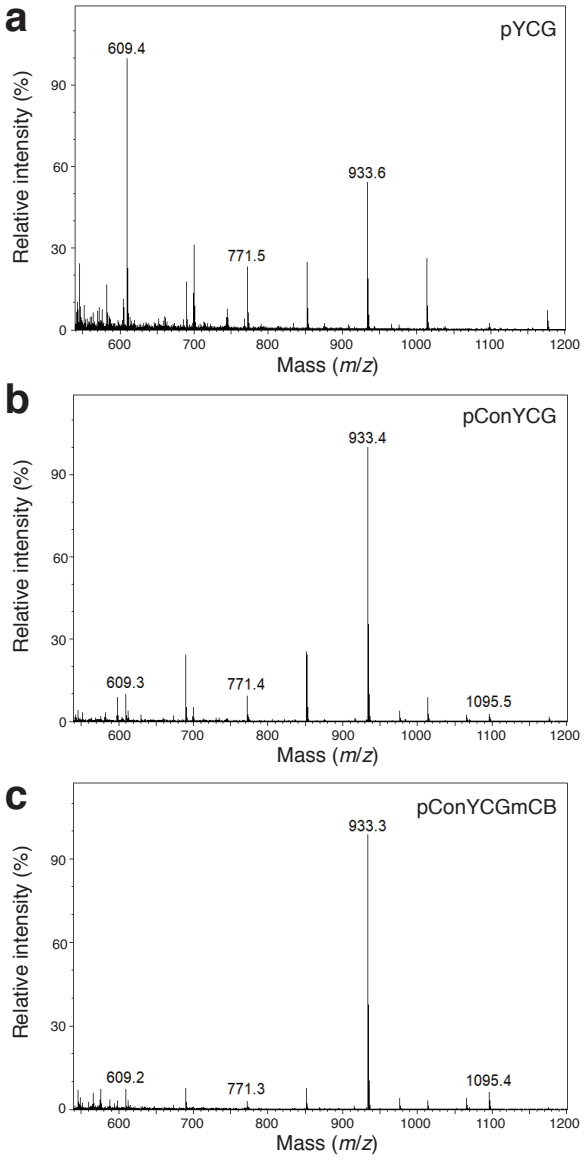


Figure 2-3. MS analysis of glycans isolated from glycoengineered E. coli strains.

MALDI-MS profiles of *N*-glycans derived from *E. coli* MC4100 *gmd::kan* Δ*waaL* cells carrying **(A)** pYCG, **(B)** pConYCG, or **(C)** pConYCGmCB as indicated. Glycans were obtained by first extracting LLOs from each of the different strains followed by separating the glycans from the lipid carrier by mild acid hydrolysis. The signals at *m*/*z* of 609, 771, 933, and 1095 correspond to $Hex_1HexNAc_2$, $Hex_2HexNAc_2$, $Hex_3HexNAc_2$, and $Hex_4HexNAc_2$, respectively.

2.3.3 Higher $Man_3GlcNAc_2$ levels correlate with increased protein glycosylation.

To determine whether transfer of $Man_3GlcNAc_2$ glycans to secretory glycoproteins *in vivo* was also improved, we focused our attention on PglB from *C. jejuni* (*Cj*PglB) because it was previously reported to recognize Und-PP-linked $Man_3GlcNAc_2$ as substrate (Valderrama-Rincon et al., 2012). For the glycoprotein target, we made use of a model fusion protein comprised of *E. coli* MBP and human GCG (residues 1-29) that was modified with an N-terminal co-translational export signal from *E. coli* DsbA and a C-terminal GlycTag having the sequence DQNAT (Fisher et al., 2011). When this engineered acceptor protein was expressed in MC4100 *gmd*::kan Δ*waaL* or Origami2 *gmd*::kan Δ*waaL* cells carrying pYCG and pMAF10 encoding *Cj*PglB, glycosylation was clearly achieved as evidenced by the mobility shift of the fusion protein from the unmodified (g0) to the monoglycosylated (g1) form (**Fig. 2-4**) in Western blots probed with anti-polyhistidine (anti-His) antibody. These results were corroborated by detection of the fusion proteins with ConA (**Fig. 2-4**), which confirmed attachment of mannose-terminal glycans. Importantly, when similar experiments were performed with strains carrying the pConYCG or pConYCGmCB plasmids, we detected highly efficient glycosylation (nearly

100%) with no visible bands corresponding to the g0 form of the acceptor protein (Fig. 4). In addition, striking improvements in the amount of monoglycosylated product were observed for both MC4100 *gmd*::kan Δ*waaL* and Origami2 *gmd*::kan Δ*waaL* carrying pConYCGmCB (**Fig. 2-4**). The yield of glycosylated fusion protein produced by these latter strains was determined to be 14.1 and 8.0 mg/L of cell culture, respectively, which is competitive with other engineered *N*-linked glycosylation pathways in *E. coli* of 25 mg/L (Ihssen et al., 2010). These values represent yield enhancements of ~8- and ~10-fold, respectively, compared to the corresponding strains carrying the pYCG plasmid. Taken together, these results confirm that the improvements in $Man_3GlcNAc_2$ biosynthesis directly translated to significantly greater accumulation of glycosylated acceptor proteins that rival yields of *E. coli*-based production pathways with bacterial polysaccharides.

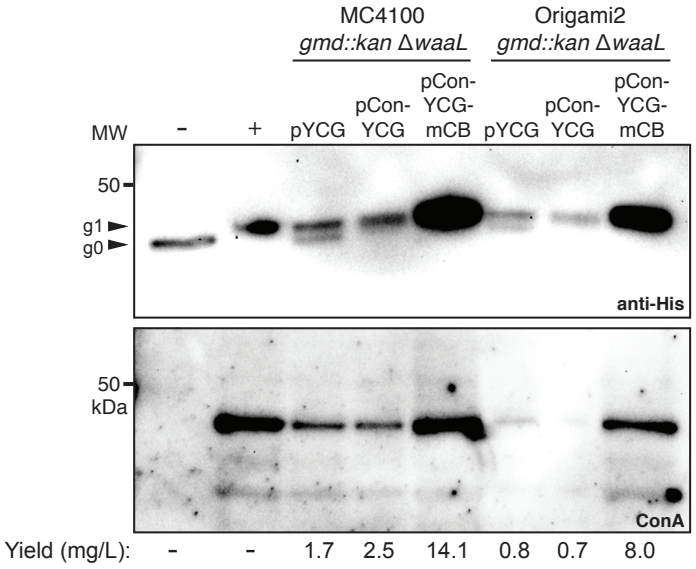


Figure 2-4. Efficient glycosylation of acceptor protein with $Man_3GlcNAc_2$.

Western blot analysis of MBP-GCG^{DQNAT} in periplasmic fractions isolated from *E. coli* MC4100 *gmd*::kan Δ*waaL* or Origami2 *gmd::kan* Δ*waaL* cells carrying pYCG, pConYCG, or pConYCGmCB as indicated. An equivalent volume was added to each lane.

Polyhistidine tags on the proteins were detected using anti-His antibodies while mannose glycans on the proteins were detected using ConA. Molecular weight (MW) markers are indicated on left side of each blot. Blots are representative of results obtained for biological triplicates. A total of 2.0 µg of purified aglycosylated or glycosylated MBP-GCG^{DQNAT} protein was added to lanes 1 and 2, respectively, and glycosylated MBP-GCG^{DQNAT} was used as a standard for quantification. Yields of each glycoprotein were estimated by densitometry analysis of individual bands using Image Lab 6.0.

2.4 Discussion

The engineering of *N*-linked glycosylation in bacteria holds great promise for the cost-effective biomanufacturing of conjugate vaccines and therapeutic glycoproteins(Baker et al., 2013; Merritt et al., 2013; Terra et al., 2012). However, in the seminal report describing N-linked glycosylation of acceptor proteins with the eukaryotic $Man_3GlcNAc_2$ glycan in *E. coli*, the yield of glycosylated proteins was reported to be ~50 µg/L, which amounted to only a small fraction (<1%) of each expressed protein under the conditions tested(Valderrama-Rincon et al., 2012). It was proposed that increasing these levels would likely require, among other things, strategies for relieving enzymatic and metabolic bottlenecks and/or optimizing the glycosylation enzymes. Indeed, we showed here that overexpressing enzymes involved in GDP-mannose biosynthesis and decreasing the expression of the yeast Alg enzymes resulted in enhanced production of lipid-linked $Man_3GlcNAc_2$. In turn, the yield of glycosylated acceptor proteins produced by these optimized strains appeared to be very efficient, with nearly 100% of the acceptor protein undergoing conversion to the glycosylated form. Moreover, yields of ~14 mg/L were

achieved in the best cases, representing a two-orders-of-magnitude improvement compared to our earlier report and rivaling the yield (25 mg/L) reported for *E. coli*-based production of carrier proteins glycosylated with bacterial polysaccharides(Ihssen et al., 2010).

To improve the glycan biosynthesis stage, we leveraged a flow cytometric assay for cell-surface displayed glycans. Historically, interrogation of this step has relied on a limited repertoire of low-throughput methods, such as Western blotting, FACE, and MS analysis, which are time consuming and labor intensive. In contrast, the use of flow cytometry for optimizing glycan biosynthesis (which can be performed over 3 days from transformation to flow cytometry with a total bench time < 5 hrs) adds a complementary tool that should streamline the pace at which pathways for *N*-linked glycosylation in *E. coli* can be studied and engineered. It is also worth pointing out that our approach is entirely compatible with new tools from metabolic engineering and synthetic biology for optimizing enzymes, pathways, and strains, or creating entirely new ones from scratch. This is significant because techniques for bacterial strain engineering such as multiplex automated genome engineering (MAGE) (Wang et al., 2009) or sRNA-based metabolic engineering (Na et al., 2013) often require high-throughput screens or selections to be effectively utilized. Although a potentially powerful approach, one concern with coupling high-throughput strain engineering techniques with our cell-surface display assay is the possibility of selecting for strains that are highly fluorescent, but exhibit growth impairment, as is the case for MC4100 Δ*gmd* cells containing pYCG. One could imagine overcoming this issue by coupling flow cytometry screening with a simple screen for fitness.

It is interesting to note that while the levels of $Man_3GlcNAc_2$ were increased by nearly 50x-fold, the yield of glycosylated acceptor protein produced in these optimized host strains only increased by 8-10x-fold. This discrepancy suggests the existence of additional bottlenecks after glycan biosynthesis, such as glycan flipping and glycan transfer to protein targets, that impact overall system efficiency. Fortuitously, in addition to the use of flow cytometry for engineering the first stages of *N*-linked glycosylation, several new tools have become available for engineering the later stages of protein glycosylation. These include a high-throughput assay called glycoSNAP (glycosylation of secreted *N*-linked acceptor proteins) that was used with great success to engineer *Cj*PglB variants with relaxed acceptor site specificity (Ollis, Zhang, Fisher, & Delisa, 2014), as well as glycophage display systems that have been used to screen acceptor-site libraries for optimal sequons (Çelik, Fisher, Guarino, Mansell, & DeLisa, 2010; Drr, Nothaft, Lizak, Glockshuber, & Aebi, 2010). We anticipate in the future that the flow cytometric approach described here could be combined with these and/or other engineering tools to systematically uncover further improvements in our eukaryotic protein glycosylation pathway.

Here, we found that overexpression of enzymes involved in GDP-mannose biosynthesis improved the production of lipid-linked $Man_3GlcNAc_2$. We also found that tuning the expression of the Alg GTases from yeast increased biosynthesis of the $Man_3GlcNAc_2$ glycan. Further manipulation of *E. coli* metabolism by genetic techniques is likely to result in additional improvements and may become necessary for the conversion of $Man_3GlcNAc_2$ to more complex human glycan structures (*i.e.*, by addition or substitution of heterologous GTase enzymes). We expect that the glycan biosynthesis

within these elaborated pathways could be rapidly optimized in a similar fashion using flow cytometry in combination with fluorophore-conjugated lectins or antibodies with different glycan specificity. It could also be used to more efficiently guide bioprocess optimization, which has so far been limited to one-at-a-time manipulations that must be manually characterized by low-throughput methods for product detection (Ihssen et al., 2010; Kämpf et al., 2015). Based on these possibilities, the future looks bright for the development of cost-effective and modular *E. coli*-based platforms for efficient biosynthesis of important glycomolecules.

2.5 Material and Methods

2.5.1 Bacterial strains and growth conditions

E. coli strains MC4100 *gmd::kan* and MC4100 *gmd::kan* Δ*waaL* (Valderrama-Rincon et al., 2012) as well as strains Origami2(DE3) *gmd::kan* and Origami2(DE3) *gmd::kan* Δ*waaL* (B. S. Hamilton et al., 2017) were all developed previously and used here. Chemically competent versions of these *E. coli* strains were transformed with plasmid combinations, plated on Luria-Bertani (LB)-agar (BD Difco) containing 100 μg/mL carbenicillin, and/or 34 μg/mL chloramphenicol. Following overnight incubation at 37°C, plates were taken out of the incubator and left at room temperature for approximately 8 h. Three colonies were picked and used to inoculate 500 μL of LB containing appropriate antibiotics at the concentrations above in a 2-mL 96-well block (Costar), and grown approximately 17 h overnight at 37°C and 1,000 rpm in a Vortemp 56 benchtop shaker (Labnet International). Ten μL of each overnight culture was then added to separate wells

on a new block containing 490 µL (1:50 dilution) of LB containing antibiotics and grown for 18 h at 30°C and 1,000 rpm in a Vortemp 56 benchtop shaker (Labnet International). Cells containing plasmid pYCG were induced by the addition of 0.2% L-arabinose at inoculation. 200 µL of each sample was spun down and washed once with 200 µL PBS buffer before resuspension in 200 µL PBS. Ten µL of each well were then transferred into 96-well plates (Costar) with 90 µL of PBS containing 3 µg/mL AlexaFluor-488 conjugated concanavalin A (ConA) lectin (Sigma) and incubated at 37°C and 1,000 rpm for 30 min in a Vortemp 56 benchtop shaker (Labnet International). Absorbance at 600 nm (Abs_{600}) was measured for each well on a plate reader (BioTek Synergy).

2.5.2 Plasmid construction

Plasmid pYCG is described elsewhere (Valderrama-Rincon et al., 2012). Plasmids pConYCG and pConYCGmCB were constructed using Gibson assembly in a manner analogous to the construction of pYCG (Valderrama-Rincon et al., 2012). Important DNA sequences in these plasmids are listed in **Supplementary Table 1**. The constitutive promoter sequences were from the Anderson promoter library from the Registry of Standard Biological Parts (partsregistry.org) (Kelly et al., 2009). The genes encoding ManB and ManC were PCR amplified from the *E. coli* MG1655 *cps* operon. Plasmid pMAF10, which encodes *C. jejuni* PglB, is described elsewhere (Feldman et al., 2005). Plasmid pTrc-spDsbA-MBP-GCG^{DQNAT} was generated using standard homologous recombination in *S. cerevisiae* as previously described (Shanks, Caiazza, Hinsa, Toutain, & O'Toole, 2006). Briefly, DNA encoding the DsbA signal peptide and *E. coli* maltose-binding protein (MBP; encoded by the *malE* gene) were amplified with primers containing homology to vector pTrc99Y (Valderrama-Rincon et al., 2012). Similarly, human glucagon

(GCG; residues 1-29) was amplified from a synthetic oligonucleotide with primers that introduced DNA encoding a C-terminal DQNAT glycosylation tag (GlycTag) (Çelik et al., 2010) and 6x-His tag followed by DNA with homology to pTrc99Y. These PCR products were used with linearized pTrc99Y to co-transform *S. cerevisiae* for cloning by homologous recombination to generate plasmid pTrc-spDsbA-MBP-GCG^{DQNAT}.

2.5.3 Flow cytometric analysis and microscopy

Five µL from each 96-well plate was diluted into 195 µL PBS in a new FACS round-bottom 96-well plate. The plate was read on an Accuri C6 Plus flow cytometer (BD Biosciences). Data for the following parameters were collected on the flow cytometer: forward scatter (FSC), side scatter (SSC), and AlexaFluor-488 fluorescence (488 nm excitation, 525 nm emission). Three to ten µL of each sample was measured. All samples were collected with 10,000 to 50,000 counts. Counts were gated in FSC versus SSC by choosing a window surrounding the largest cluster of cells. Fluorescence values were recorded in relative channel number (1-262,144 corresponding to 18-bit data) and the geometric mean over the gated data was calculated for each sample. Data analysis and FACS calibration was performed as described previously (Lucks, Qi, Mutalik, Wang, & Arkin, 2011). Rainbow calibration particles (Spherotech) were used to obtain a calibration curve to convert fluorescence intensity (geometric mean, relative channel number) into molecules of equivalent fluorescein (MEFL) units. Mean MEFL values were calculated over replicates. All flow cytometry experiments involved three biological replicates. Extraction of Flow Cytometry Standard (FCS) files, traditional gating, analysis of calibration beads data, standard curve generation, transformation to MEFL, and generation of histograms was performed using FlowCal software(Castillo-Hair et al.,

2016). For fluorescence microscopy, ConA-AlexaFluor labeled cells were pelleted, washed, and resuspended in 100 µL PBS buffer. Agarose pads were prepared by adding 2% low-melt agarose to culture medium and heating until melted. Thirty µL of agarose solution were then added onto a depression slide and covered with another flat slide (Fisher Scientific). Following pad solidification (~10 min), 5 µL of cells treated with ConA-AlexaFluor were added to the pad, and covered with a cover slip. Slides were viewed on an inverted fluorescent Leica DM-IL LED microscope with a Leica HCX PL APO 100x/1.40 PH CS oil-immersion objective lens and high-resolution cooled Q-imaging CCD.

2.5.4 Isolation and analysis of LLOs

The methods for extraction of LLOs and release of the glycan from the lipid were followed as described elsewhere (Gao, 2005; Gao & Lehrman, 2006). Briefly, dried *E. coli* cell pellets from 250 mL culture were resuspended in 2:1 chloroform:methanol, sonicated, and the remaining solids collected by centrifugation. This pellet was sonicated in water and collected by centrifugation. The resulting pellet was sonicated in 10:10:3 chloroform:methanol:water to isolate the LLOs from the inner membrane. The LLOs were purified using acetate-converted DEAE anion exchange chromatography as they bind to the anion exchange resin via the phosphates that link the lipid and glycan. The resulting compound was dried and treated by mild acid hydrolysis to release glycans from the lipids (Gao, 2005; Gao & Lehrman, 2006). The released glycans were then separated from the lipid by a 1:1 butanol:water extraction, wherein the water layer contains the glycans. The glycans were then further purified with a graphitized carbon column using a 0-50% water:acetonitrile gradient. Isolated $Man_3GlcNAc_2$ glycans were analyzed by fluorophore-

assisted carbohydrate electrophoresis (FACE) as described (Gao, 2005). Briefly, 20 µl of purified glycan was dried and resuspended in equal volumes of $NaCNBH_3$ and 7-amino-1,3-naphthalenedisulfonic acid (ANDS) solution. The reaction was incubated for 18 h at 37°C, dried, and resuspended in thorin I loading dye. Gels were prepared according to the published protocol(Gao, 2005) and run at 4°C with a constant current of 10 mAmp. Gels were visualized using a Bio-Rad ChemiDoc MP. Isolated $Man_3GlcNAc_2$ glycans were also analyzed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was performed as described. Briefly, after desalting with Dowex 50WX8 hydrogen form, 200-400 mesh (Sigma) and AG1-X8 formate form, 200-400 mesh (Bio-Rad) resins, glycans were concentrated and MALDI-MS profiles were acquired using 2,5-dihydroxybenzoic acid (Alfa Aesar) as matrix on a 5800 MALDI-TOF/TOF (SciEx) in positive ion reflectron mode (500-3000 Da mass range). MS traces were generated and analyzed using mMass (Strohalm, Hassman, Košata, & Kodíček, 2008).

2.5.5 Glycoprotein expression and isolation

Cells carrying one of the glycan biosynthesis plasmids along with pMAF10 and pTrc-spDsbA-MBP-GCG^{DQNAT} were grown in 100 mL of LB at 37°C until Abs_{600} reached ≈ 1.5. Culture temperature was reduced to 30°C (and L-arabinose was added to a final concentration of 0.2% (w/v) to induce *alg* genes in pYCG) and allowed to grow overnight at 30°C. The next day, cells were induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to initiate synthesis of the MBP-GCG^{DQNAT} acceptor protein. Protein expression proceeded for 8 h at 30°C. Cells were then harvested and subjected to subcellular fractionation, which was performed as described elsewhere

(Karlsson et al., 2012). Briefly, we pelleted and washed 100 mL of IPTG-induced culture with subcellular fractionation buffer (0.2 M Tris-Ac (pH 8.2), 0.25 mM EDTA, and 0.25 M sucrose, and 160 µg/mL lysozyme). Cells were resuspended in 1.5 mL subcellular fractionation buffer and then incubated for 5 min on ice and spun down. After addition of 60 µL of 1 M $MgSO_4$, cells were incubated for 10 min on ice. Cells were spun down, and the supernatant was taken as the periplasmic fraction. To isolate glycoproteins, periplasmic fractions were subjected to affinity chromatography with a Qiagen Ni-NTA kit. Eluates were collected, solubilized in Laemmli sample buffer containing 5% β-mercaptoethanol, and resolved on SDS-polyacrylamide gels.

2.5.6 Western blot analysis.

Purified protein samples were separated using 8% SDS-PAGE gels and transferred to PVDF membranes. Proteins that harbored 6x-His affinity tags were detected with a monoclonal anti-polyhistidine-horse radish peroxidase (HRP) conjugate (Abcam) per manufacturers' instructions. Protein-conjugated glycans were detected with 5 µg/mL ConA-HRP conjugate (Sigma).

2.6 Competing financial interests

M.P.D. has a financial interest in Glycobia, Incorporated and Versatope, Incorporated. M.P.D.'s interests are reviewed and managed by Cornell University in accordance with their conflict of interest policies.

2.7 Author Contributions

C.J.G. designed research, performed research, analyzed data, and wrote the paper. L.E.Y., T.J., J.D.W. and J.H.M. designed research, performed research, and analyzed data. J.B.L. and M.P.D. designed research, analyzed data, and wrote the paper.

2.8 Acknowledgements

The authors would like to thank members of the Lucks and DeLisa labs for helpful discussions during the process of preparing this manuscript as well as Prof. Josh Leonard and Taylor Dolberg for assistance with fluorescence microscopy. This research was supported by National Science Foundation Awards CBET-1605242 (to M.P.D.), CBET-1159581 (to M.P.D.), CBET-1402843 (to J.B.L. and M.P.D.), a National Science Foundation Graduate Research Fellowship (to C.J.G.), and a Royal Thai Government Fellowship (to T.J.).

2.9 References

Abu-Qarn, M., Eichler, J., & Sharon, N. (2008). Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2008.06.010

Baker, J. L., Çelik, E., & DeLisa, M. P. (2013). Expanding the glycoengineering toolbox: The rise of bacterial N-linked protein glycosylation. *Trends in Biotechnology*. https://doi.org/10.1016/j.tibtech.2013.03.003

Castillo-Hair, S. M., Sexton, J. T., Landry, B. P., Olson, E. J., Igoshin, O. A., & Tabor, J. J. (2016). FlowCal: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synthetic*

Biology. https://doi.org/10.1021/acssynbio.5b00284

Çelik, E., Fisher, A. C., Guarino, C., Mansell, T. J., & DeLisa, M. P. (2010). A filamentous phage display system for N-linked glycoproteins. *Protein Science*. https://doi.org/10.1002/pro.472

Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N., & Dwek, R. A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. *Science*. https://doi.org/10.1126/science.1235681

Drr, C., Nothaft, H., Lizak, C., Glockshuber, R., & Aebi, M. (2010). The Escherichia coli glycophage display system. *Glycobiology*. https://doi.org/10.1093/glycob/cwq102

Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., … Aebi, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0500044102

Fisher, A. C., Haitjema, C. H., Guarino, C., Çelik, E., Endicott, C. E., Reading, C. A., … DeLisa, M. P. (2011). Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01901-10

Gao, N. (2005). Fluorophore-assisted carbohydrate electrophoresis: A sensitive and accurate method for the direct analysis of dolichol pyrophosphate-linked oligosaccharides in cell cultures and tissues. *Methods*. https://doi.org/10.1016/j.ymeth.2004.10.003

Gao, N., & Lehrman, M. A. (2006). Non-Radioactive Analysis of Lipid-Linked Oligosaccharide Compositions by Fluorophore-Assisted Carbohydrate Electrophoresis. *Methods in Enzymology*. https://doi.org/10.1016/S0076-6879(06)15001-6

Gottesman, S., Trisler, P., & Torres-Cabassa, A. (1985). Regulation of capsular polysaccharide synthesis in Escherichia coli K-12: Characterization of three regulatory genes. *Journal of Bacteriology*. https://doi.org/10.1111/j.1365-2958.1991.tb01906.x

Hamilton, B. S., Wilson, J. D., Shumakovich, M. A., Fisher, A. C., Brooks, J. C., Pontes, A., … Delisa, M. P. (2017). A library of chemically defined human N-glycans synthesized from microbial oligosaccharide precursors. *Scientific Reports*. https://doi.org/10.1038/s41598-017-15891-8

Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., … Gerngross, T. U. (2003). Production of complex human glycoproteins in yeast. *Science*. https://doi.org/10.1126/science.1088166

Helenius, A., & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science*. https://doi.org/10.1126/science.291.5512.2364

Ihssen, J., Kowarik, M., Dilettoso, S., Tanner, C., Wacker, M., & Thöny-Meyer, L. (2010). Production of glycoprotein vaccines in Escherichia coli. *Microbial Cell Factories*. https://doi.org/10.1186/1475-2859-9-61

Ilg, K., Yavuz, E., Maffioli, C., Priem, B., & Aebi, M. (2010). Glycomimicry: Display of the GM3 sugar epitope on Escherichia coli and Salmonella enterica sv Typhimurium. *Glycobiology*. https://doi.org/10.1093/glycob/cwq091

Kämpf, M. M., Braun, M., Sirena, D., Ihssen, J., Thöny-Meyer, L., & Ren, Q. (2015). In vivo production of a novel glycoconjugate vaccine against Shigella flexneri 2a in recombinant Escherichia coli: Identification of stimulating factors for in vivo glycosylation. *Microbial Cell Factories*. https://doi.org/10.1186/s12934-015-0195-7

Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., … Endy,

D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*. https://doi.org/10.1186/1754-1611-3-4

Keys, T. G., & Aebi, M. (2017). Engineering protein glycosylation in prokaryotes. *Current Opinion in Systems Biology*. https://doi.org/10.1016/j.coisb.2017.05.016

Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., & Arkin, A. P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1015741108

Merritt, J. H., Ollis, A. A., Fisher, A. C., & Delisa, M. P. (2013). Glycans-by-design: Engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.24885

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., … Callewaert, N. (2014). GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2885

Mills, D. C., Jervis, A. J., Abouelhadid, S., Yates, L. E., Cuccui, J., Linton, D., & Wren, B. W. (2016). Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria. *Glycobiology*. https://doi.org/10.1093/glycob/cwv111

Mimura, Y., Katoh, T., Saldova, R., O'Flaherty, R., Izumi, T., Mimura-Kimura, Y., … Rudd, P. M. (2018). Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein and Cell*. https://doi.org/10.1007/s13238-017-0433-3

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., & Lee, S. Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2461

Ollis, A. A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., … DeLisa, M. P. (2015). Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Scientific Reports*. https://doi.org/10.1038/srep15237

Ollis, A. A., Zhang, S., Fisher, A. C., & Delisa, M. P. (2014). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1609

Rich, J. R., & Withers, S. G. (2009). Emerging methods for the production of homogeneous human glycoproteins. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.148

Rudd, P. M., & Dwek, R. A. (1997). Glycosylation: Heterogeneity and the 3D structure of proteins. *Critical Reviews in Biochemistry and Molecular Biology*. https://doi.org/10.3109/10409239709085144

Sethuraman, N., & Stadheim, T. A. (2006). Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2006.06.010

Shanks, R. M. Q., Caiazza, N. C., Hinsa, S. M., Toutain, C. M., & O'Toole, G. A. (2006). Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.00682-06

Song, T., Aldredge, D., & Lebrilla, C. B. (2015). A Method for In-Depth Structural Annotation of Human Serum Glycans That Yields Biological Variations. *Analytical Chemistry*. https://doi.org/10.1021/acs.analchem.5b01340

Strohalm, M., Hassman, M., Košata, B., & Kodíček, M. (2008). mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass*

Spectrometry. https://doi.org/10.1002/rcm.3444

Terra, V. S., Mills, D. C., Yates, L. E., Abouelhadid, S., Cuccui, J., & Wren, B. W. (2012). Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design. *Journal of Medical Microbiology*. https://doi.org/10.1099/jmm.0.039438-0

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., … DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.921

Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., … Aebi, M. (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science*. https://doi.org/10.1126/science.298.5599.1790

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. https://doi.org/10.1038/nature08187

Wormald, M. R., & Dwek, R. A. (1999). Glycoproteins: Glycan presentation and protein-fold stability. *Structure*. https://doi.org/10.1016/S0969-2126(99)80095-1

Xu, C., & Ng, D. T. W. (2015). Glycosylation-directed quality control of protein folding. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm4073

Yang, Z., Wang, S., Halim, A., Schulz, M. A., Frodin, M., Rahman, S. H., … Clausen, H. (2015). Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3280

Zhang, P., Woen, S., Wang, T., Liau, B., Zhao, S., Chen, C., … Rudd, P. M. (2016). Challenges of glycosylation analysis and control: An integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today*.

https://doi.org/10.1016/j.drudis.2016.01.006

CHAPTER 3 - *High-throughput strain engineering with RNA regulators of gene expression*[2]

3.1 Abstract

Genetic manipulation of the host's endogenous metabolism is a common approach to re-direct carbon flux from the production of biomass to desired products. This often requires high-throughput experimental approaches to identify genes, gene combinations, and their optimal expression levels that can influence and optimize production titers of desired products. In this chapter, we explored the of use RNA regulators, namely synthetic small regulatory RNAs (sRNAs) and de novo designed translational RNA repressors, called looped antisense oligonucleotides (LASOs) to down-regulate pathways in the cell that compete with $Man_3GlcNAc_2$ glycan biosynthesis. To do this, we selected gene targets for down-regulation by studying sugar biosynthesis pathways in *E. coli* and analyzed the effect of each modification by flow-cytometry analysis of $Man_3GlcNAc_2$ glycan production in multiple host strains. Finally, we propose a standardized approach for high-throughput strain engineering by screening gene knockdown combinations that could rapidly guide construction of genome-engineered strains for industrial use.

[2] Results contained in this chapter (**Fig. 3A-D**) have been submitted for publication:

Carlson, P. D., **Glasscock, C. J.**, and Lucks, J. B. (2019) De Novo Design of Translational RNA Repressors. *In revision (Nature Communications)*. DOI 10.1101/501767

3.2 Introduction

One goal of synthetic biology and metabolic engineering has been to develop tools for rapidly modifying endogenous gene expression in microbial hosts. Synthetic small regulatory RNAs (sRNAs) are a technology that has been developed to repress target gene expression at the translational level (Chae, Kim, Choi, Park, & Lee, 2015; Na et al., 2013; Noh, Yoo, Kim, & Lee, 2017). These sRNAs contain a reverse-complement sequence that binds the translation initiation region (TIR) of a target mRNA (**Fig. 3-1A**), either blocking the ribosome from translating the protein coding sequence or facilitating increased mRNA transcript degradation. sRNAs that naturally regulate gene expression in bacteria are typically composed of an additional scaffold sequence, which is known to interact with the chaperone protein Hfq thought to facilitate sRNA-mRNA hybridization.

Na *et al.* first reported the use of sRNAs for metabolic engineering and achieved ~55% improvements in flux through a cadaverine producing pathway in *E. coli* by targeting natural genes in *E. coli* metabolism (Na et al., 2013). In addition, this report showed that synthetic sRNAs can be multiplexed to screen the effect of combination knockdowns and that synthetic sRNAs can be easily ported into different *E. coli* strains, eliminating the need to individually create gene knockouts in each strain. The approach in this report was to repurpose natural *E. coli* sRNAs (namely the MicC sRNA) by replacing the binding sequence with the reverse complement of the TIR of a desired target mRNA. Na *et al*. showed strong repression of fluorescent reporter genes in this study. However, little work has been done to fully characterize the effectiveness of the these synthetic sRNAs for repression of arbitrary endogenous gene targets. In addition, the role of the Hfq scaffold sequence is poorly understood in the context of synthetic sRNAs and

studies have failed to show a general positive affect of the Hfq sequence in synthetic sRNAs (Meyer, Chappell, Sankar, Chew, & Lucks, 2016). One recent study established a predictive model for repression efficiency of endogenous *E. coli* genes with synthetic sRNAs, however repression efficiency was indirectly validated for only two endogenous targets (Lee, Kim, Amrofell, & Moon, 2019).

More recently, Carlson *et al.* (Carlson, Glasscock, & Lucks, 2018), developed an entirely *de novo* approach for designing translational RNA repressors that do not require the use of a natural sRNA scaffold sequence. These translational RNA repressors, called looped antisense oligonucleotides (LASOs), exhibit many of the best qualities of synthetic sRNAs but avoid the murkiness arising from modifying a natural sRNA scaffold and use a unique design approach to avoid off-target repression. In addition, the effect of many sequence parameters on repression efficiency were thoroughly characterized in terms of the LASO sequence and the structural effects of the target mRNA. This characterization included the effectiveness of LASOs against a set of 6 endogenous gene targets with various structural properties.

3.3 Results

3.3.1 Design and use of synthetic small regulatory RNAs for optimizing $Man_3GlcNAc_2$ glycan biosynthesis

We hoped to use sRNAs to down regulate expression of genes in *E. coli* that might be competing with $Man_3GlcNAc_2$ glycan biosynthesis. Hence, we constructed a plasmid for expressing the natural MicC sRNA sequence and modified the targeting sequence to bind a gene encoding red fluorescent protein (RFP) that was integrated into the *E. coli*

genome. We found that expression of the anti-RFP sRNA resulted in ~90% knockdown of RFP fluorescence (**Fig. 3-1B**). Encouraged, we designed sRNAs against *E. coli* genes *wecB* and *wecG*, which we expected might reduce flux to $Man_3GlcNAc_2$ glycan biosynthesis by consuming UDP-GlcNAc (*wecB*) and UDP-GlcNAc-UndPP (*wecG*). We observed that both sRNAs resulted in glycan synthesis improvements over pConYCG or pConYCGmCB (**Fig. 3-1C**).

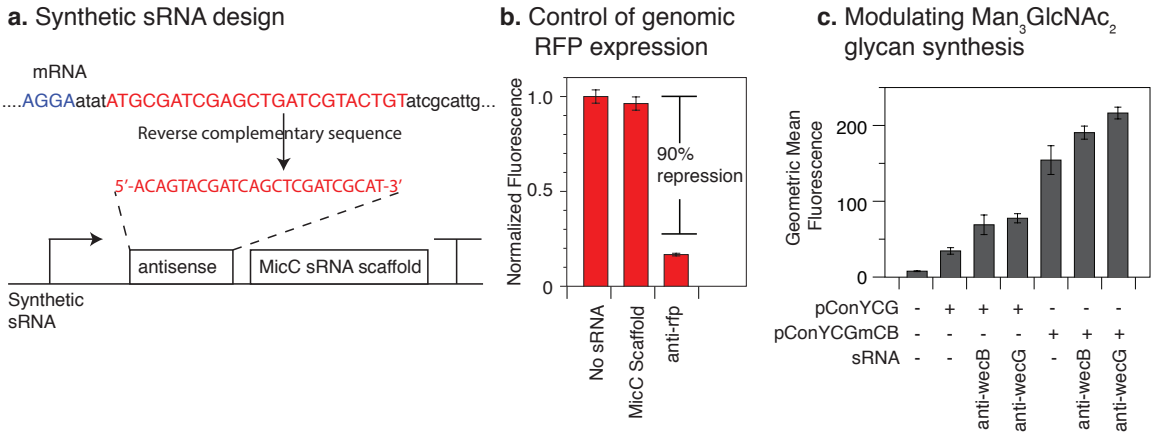


Figure 3-1. Metabolic engineering with synthetic sRNAs

(A) Design of synthetic sRNAs. sRNAs contain an antisense sequence that binds the target mRNA at the translation initiation region (TIR) to block ribosome binding or facilitate mRNA degradation. An additional scaffold sequence from the *E. coli* sRNA MicC is thought to recruit sRNA binding protein *Hfq,* which facilitates sRNA-mRNA hybridization. **(B)** sRNAs designed against genome-integrated red fluorescent protein (RFP) expression result in ~90% knockdown over no sRNA or MicC scaffold sRNA controls. Error bars mean ± standard error. **(C)** sRNAs designed against *wecB* or *wecG* result in improvements in $Man_3GlcNAc_2$ glycan synthesis over pConYCG or pConYCGmCB alone. Error bars mean ± standard error.

3.3.2 Screening gene knockdowns for $Man_3GlcNAc_2$ glycan biosynthesis in multiple strains with synthetic small regulatory RNAs.

One difficulty in strain engineering is the process of making genome modifications with conventional methods, such as lambda-red based recombination (Datsenko & Wanner, 2002), which becomes increasingly difficult with the number of modifications that are desired and when it is beneficial to observe the effect of these modifications in multiple host strains. Technologies such as multiplexed automated genome engineering (MAGE) (Wang et al., 2009) and trackable recursive multiplex recombineering (TRMR) (Glebes et al., 2012) increase the speed by which these modifications can be constructed. However, modifications must still be made one strain at time making it difficult to rapidly observe the effect of genome modifications in different hosts. sRNAs circumvent this issue by regulating gene expression in *trans* so that one sRNA construct can be used to screen a knockdown in multiple strains simultaneously. We hoped to use this feature of sRNA regulation to screen gene knockdown targets for improvements in $Man_3GlcNAc_2$ glycan biosynthesis while thoroughly exploring the effect of these knockdowns in different strains. Hence, we chose a panel of 10 genes in *E. coli* that were involved in glycan biosynthesis (**Fig. 3-2A**) and designed sRNAs to repress these targets. We then chose a set of 6 candidate strains for glycan biosynthesis resulting in a total of 60 strain-sRNA pairs, which would be time-intensive and laborious to construct with other methods. We transformed each strain with the sRNA panel along with plasmid pConYCGmCB. Flow cytometry analysis was then used to analyze cell-surface displayed glycan and find the best performing variants (**Fig. 3-2B**). Results show that MG1655 performs significantly better than other tested strains and the anti-*murA* sRNA in MG1655 resulted in ~7.5 fold

improvement over wildtype MC4100, the previous standard strain for $Man_3GlcNAc_2$ glycan synthesis. Most sRNAs resulted in strain-dependent phenotypes. For example, anti-*murA* in MC4100 resulted in only a slight improvement over wildtype, while significantly improving glycan synthesis in MG1655. These findings highlight the advantages of a multi-strain metabolic engineering approach streamlined by sRNA regulation.

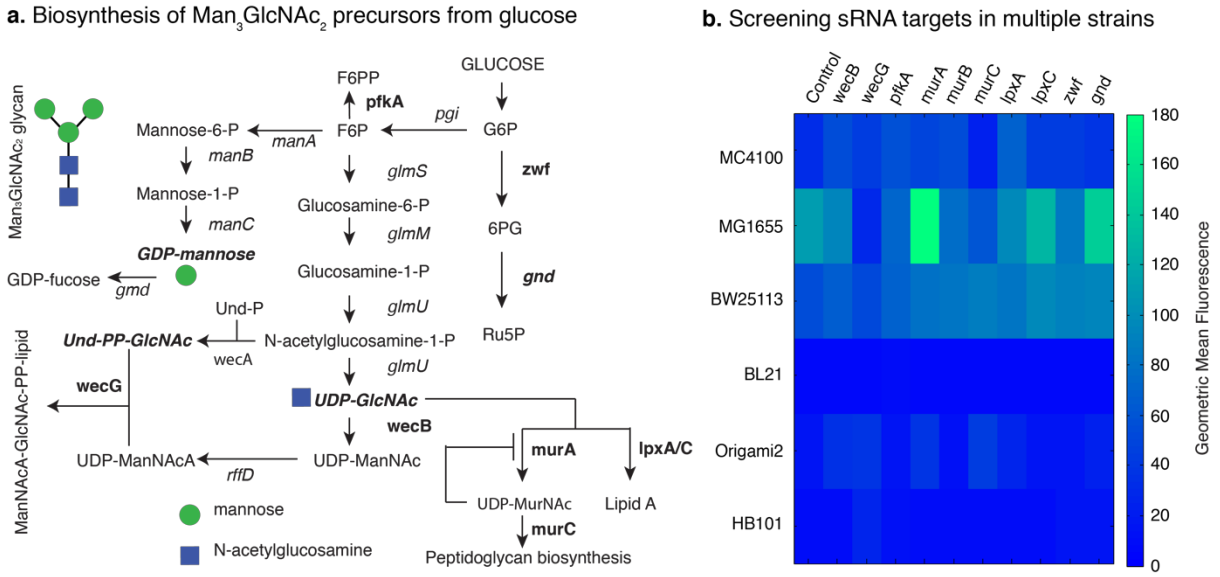


Figure 3-2. Screening sRNA targets for improvements in $Man_3GlcNAc_2$ glycan biosynthesis in multiple strains.

(A) Biosynthesis of $Man_3GlcNAc_2$ precursors (Bold italic) from glucose in *E. coli.* Ten sRNA targets (bold) were selected based on potential competition with biosynthesis of GDP-mannose, UDP-GlcNAc, or Undd-PP-GlcNAc precursors. **(B)** Screening sRNA targets in multiple wildtype strains with flow cytometry. MG1655 performed significantly better than other tested strains. The anti-*murA* sRNA in MG1655 resulted in ~2x fold improvement over wildtype MG1655 and ~7.5x fold improvement over wildtype MC4100.

3.3.3 LASO designs can predictably repress endogenous mRNA sequences to improve $Man_3GlcNAc_2$ glycan biosynthesis

In our work using synthetic sRNAs, we observed several limitations in their ability to predictably down-regulate endogenous gene expression. In particular, the role of *Hfq* scaffold sequences are poorly understood and the design principles involved in swapping the antisense sequence in natural sRNAs have yet to be fully elucidated. In our hands, these factors led to variable and unpredictable performance for targeting arbitrary endogenous genes for strain engineering. During this work, Carlson et al. (Carlson et al., 2018), developed a new approach for *de novo* design of translational RNA repressors, called looped antisense oligonucleotides (LASOs), with several unique advantages over synthetic sRNAs. In particular, LASOs do not require modifying a natural sRNA sequence and their repression efficiency was shown to be predictably correlated to an updated definition of the free energy of the mRNA/LASO complex (ΔG_{CF}) that accounts for intramolecular structures formed by the mRNA target (Carlson et al., 2018), providing a simple heuristic to screen the effectiveness of LASOs for repression of arbitrary endogenous genes *a priori*. To demonstrate the ability of LASOs to repress sequences derived from natively expressed mRNAs, we created fusion proteins where the 5' UTR and initial coding region of several *E. coli* genes were fused to sfGFP (**Fig. 3-3A**). The selected *E. coli* genes were a subset of those selected for analysis with synthetic sRNAs in Fig. 3-2B. LASOs were designed to target the translation start site of these fusion constructs (**Fig. 3-3B**), and the repression performance was measured by flow cytometry. Significant repression was observed in each case (Fig. 3C), although the strength of repression ranged from 42% to 94%. We hypothesized that structural inaccessibility of

the mRNA target may inhibit mRNA/LASO complex formation and cause this range of observed repression. As expected, based on characterization in Carlson *et al.*, repression strength decreased with increasing ΔG_{CF} (**Fig. 3-3D**). This suggests that the definition of ΔG_{CF} is a useful heuristic to predict the strength of repression of mRNA targets with varying degrees of structural inaccessibility and that, for this reason, LASOs might be a superior choice for effective and predictable knockdown of endogenous gene targets.

We next sought to test the ability of LASOs to modulate $Man_3GlcNAc_2$ glycan biosynthesis in *E. coli.* Similar to previous experiments with synthetic sRNAs, we transformed LASO plasmids into 3 strains of *E. coli* with pConYCGmCB and performed flow cytometry analysis to determine their effect on glycan biosynthesis. We observed a strain dependent improvement in glycan biosynthesis from targeting *E. coli* genes *gnd* and *lpxC* with LASOs only in *E. coli* MG1655 (**Fig. 3-3**). These same gene targets also improved $Man_3GlcNAc_2$ glycan biosynthesis levels when targeted with synthetic sRNAs (**Fig. 3-2**), providing validation that downregulation of these gene targets is responsible for improving glycan biosynthesis rather than a non-specific targeting of sRNA or LASO expression. Interestingly, in the case of *gnd,* glycan biosynthesis was significantly worsened upon expression of the same LASO in *E. coli* MC4100, demonstrating the importance of examining the effect of each knockdown in multiple strain backgrounds.

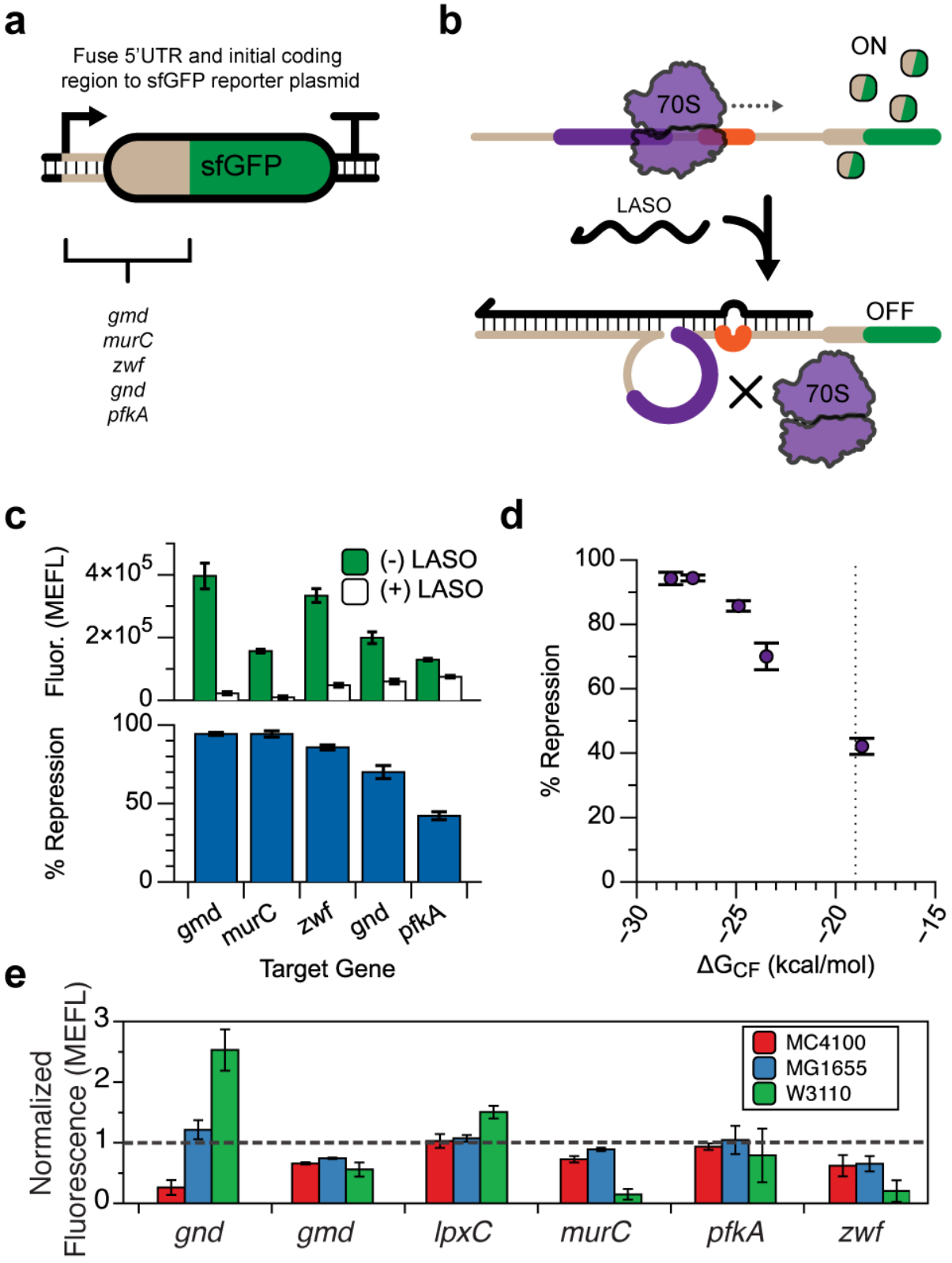


Figure 3-3. The LASO design can predictably target mRNA sequences from endogenous genes.

(A) The 5' UTR and initial coding region (51nt corresponding to 17AA) of several endogenous *E. coli* genes were translationally fused to an sfGFP coding sequence to create a chimeric expression cassette. **(B)** LASOs were designed to target the expressed chimeric mRNAs. **(C)** Flow cytometry characterization of sfGFP fluorescence. Each target was responsive to LASO expression, with repression efficiency ranging from 94% to 42%. **(D)** Differences in repression can be explained by calculated ΔG_{CF} values for each LASO/target pair, with stronger repression corresponding to lower values of ΔG_{CF} (-19 kcal/mol indicated by a dashed line). **(E)** Screening sRNA targets in multiple wildtype strains for improved production of $Man_3GlcNAc_2$ glycan biosynthesis by flow cytometry. Data for each strain were normalized to the fluorescence of cells with pConYCGmCB and a no-LASO control plasmid. MG1655 performed significantly better than other tested strains and LASOs targeting *gnd* and *lpxC* significantly improved glycan biosynthesis. Figure adapted from Carlson *et al*.

3.4 Discussion

In this chapter, we explored the use of RNA translational repressors, including both synthetic sRNAs and LASOs, for screening the effect of gene knockdowns on metabolic pathway performance in the context of $Man_3GlcNAc_2$ glycan biosynthesis. While we were able to demonstrate the design and application of single sRNA/LASO knockdown for improving pathway performance, we ultimately expected to develop a multiplexed platform for screening the effect of combinatorial gene knockdowns with varying repression levels to thoroughly map the effect of endogenous gene expression on pathway performance. During this work, a report published by Voigt and coworkers

developed a highly similar approach for balancing metabolic pathway enzymes (Ghodasara & Voigt, 2017). In this report, sRNAs were designed to target specific 5' UTRs that could regulate enzyme expression in metabolic pathways. Pools of sRNA constructs were created where up to six sRNAs were multiplexed onto a single plasmid with modulated repression by using various strength promoters to control each sRNA. These pools could be transformed into cells containing a metabolic pathway with each enzyme regulated by the specific target 5' UTRs and screened for improvements in titer. Ultimately, the best strains were then sequenced to identify the optimal sRNA repression strength for each pathway enzyme (based on the sRNA promoters in the optimal constructs). This allowed the final pathway to be hardcoded by modifying pathway enzyme promoters to match the optimal repression level achieved achieved with optimal sRNAs.

Although Ghodasara *et al*. focused on balancing heterologous pathway enzymes rather than regulating endogenous genes to re-direct carbon flux, their solution was highly similar to what we envisioned, and we ultimately decided to discontinue this work. However, we propose a similar workflow for rapid strain engineering with sRNA pools targeting endogenous genes (**Fig. 3-4).** In this proposed workflow, LASO RNAs can be designed against endogenous gene targets and multiplexed into a single plasmid with varying strength promoters. These multiplexed LASO constructs would then be pooled and screened for improvements in metabolic pathway titers in a similar manner. Ideally, each sRNA expression cassette in the pooled constructs would correspond to a specific barcode that facilitates sequencing of the top variants. Ultimately, if implemented, this technique could provide a rapid tool for informing hardcoded genome engineering

manipulations in production strains. In addition to the general workflow, we propose a restriction enzyme-based cloning scheme that would allow simple construction of pools of barcoded, multiplexed LASO constructs (**Fig. 3-5**).

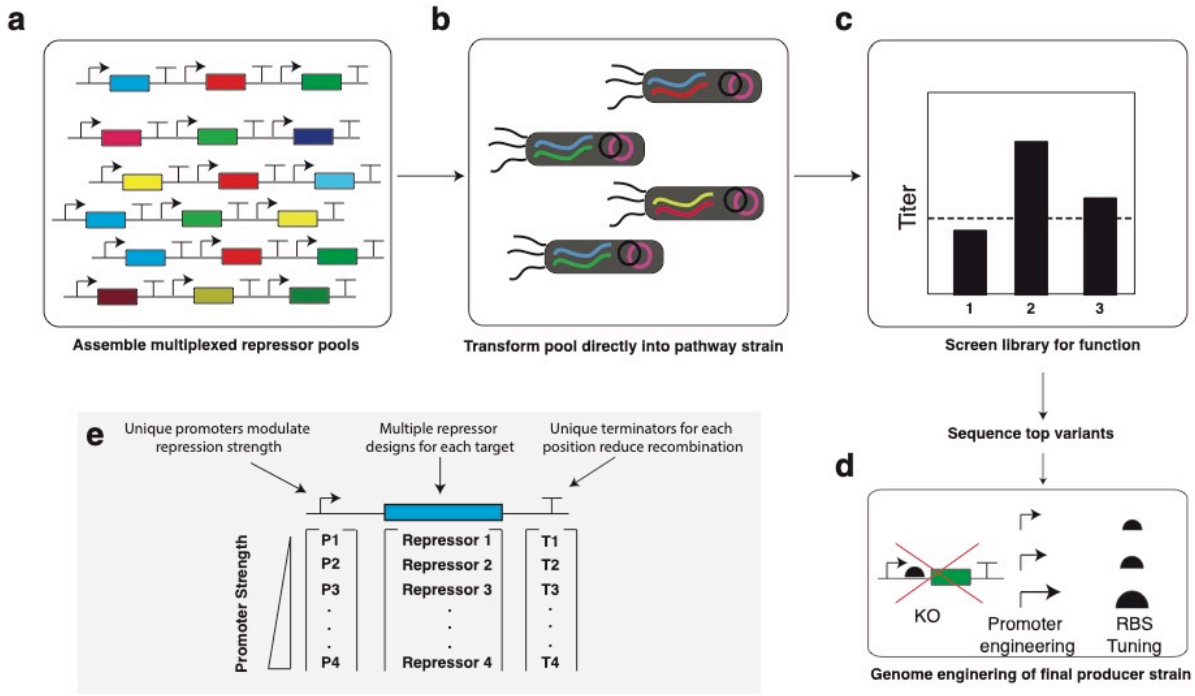


Figure 3-4. Proposed workflow for high-throughput strain engineering with LASOs.

(A) Pooled barcoded, multiplexed LASO constructs are assembled for modulated repression of target endogenous genes. **(B)** LASO pools are transformed into production strains so that each colony has a unique LASO construct expressing different LASOs at various strengths. **(C)** Variants are screened for the desired phenotype using an appropriate analytical technique (e.g. flow cytometry, HPLC, GCMS) and top hits are sequenced to determine optimal sRNA constructs. **(D)** Results from pooled LASO construct screening are used to inform genome engineering of a final producer strain by creating gene knockouts or modulating expression of endogenous genes by modifying the natural promoter or RBS. **(E)** Each LASO cassette in the barcoded, multiplexed LASO

constructs would include a LASO RNA with a unique promoter to modulate its expression level and a unique terminator to prevent genetic instability by unintended recombination.

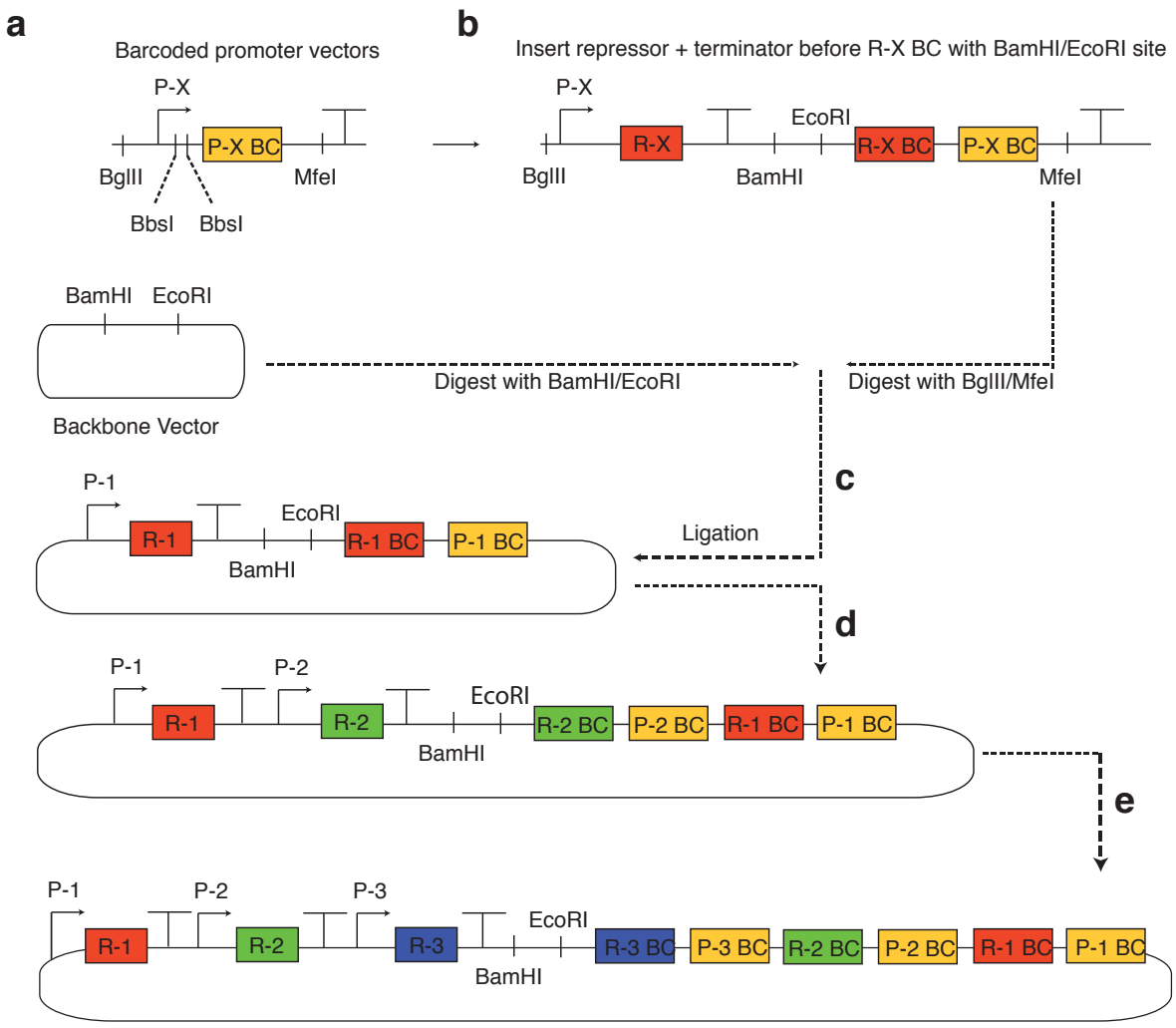


Figure 3-5. Proposed restriction enzyme cloning scheme for pools of barcoded, multiplexed LASO constructs.

(A) Backbone plasmids are constructs with a library of various strength promoters to modulate LASO expression. Each promoter is accompanied by a corresponding barcode sequence (P-X BC). The promoter and P-X are flanked by BglII and MfeI cut sites and

contain two unique BbsI cut sites between the promoter and P-X. **(B)** Promoter backbones are digested with BbsI to facilitate golden gate ligation of a LASO repressor with a terminator and a unique barcode sequence (R-X BC) between the promoter and P-X. Sequences are designed so that R-X BC is adjacent to P-X BC at the 3' junction. BamHI and EcoRI cut sites are designed between the terminator and R-X BC. **(C)** Individual barcoded promoter-LASO cassettes are digested with BglII/MfeI for ligation into a new backbone plasmid digested with BamHI/EcoRI. **(D)** The new single LASO construct is then used as the backbone for subsequent ligations of additional barcoded promoter-LASO cassettes. This is facilitated by the new BamHI/EcoRI cut sites between the repressor terminator and R-X BC. This second ligation creates a multiplexed construct with barcodes located at the 3' junction corresponding the promoter and repressor in each position. **(E)** The process in C and D can be cycled repeatedly to create pooled constructs with increasing repressors with corresponding barcodes for each position at the 3' junction.

3.5 Material and Methods

3.5.1 Plasmid construction

All sRNA and LASO sequences were constructed by inverse PCR (iPCR). sRNA/LASO sequences were cloned into ColE1 backbones harbouring carbenecillin resistance.

3.5.2 Bacterial strains and growth conditions for $Man_3GlcNAc_2$ glycan production

E. coli strains MC4100 *gmd::kan* (Valderrama-Rincon *et al.*, 2012) was developed previously and used here. In addition, *E. coli* MC4100, MG1655, BW25113, BL21, Origami2, and HB101 were used in Fig. 2. Chemically competent versions of these *E. coli* strains were transformed with plasmid combinations (pConYCG or pConYCGmCB from Chapter 2 and an sRNA/LASO plasmid), plated on Luria-Bertani (LB)-agar (BD Difco) containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol. Following overnight incubation at 37°C, plates were taken out of the incubator and left at room temperature for approximately 8 h. Three colonies were picked and used to inoculate 500 µL of LB containing appropriate antibiotics at the concentrations above in a 2-mL 96-well block (Costar), and grown approximately 17 h overnight at 37°C and 1,000 rpm in a Vortemp 56 benchtop shaker (Labnet International). Ten µL of each overnight culture was then added to separate wells on a new block containing 490 µL (1:50 dilution) of LB containing antibiotics and grown for 18 h at 30°C and 1,000 rpm in a Vortemp 56 benchtop shaker (Labnet International). Cells containing plasmid pYCG were induced by the addition of 0.2% L-arabinose at inoculation. 200 µL of each sample was spun down and washed once with 200 µL PBS buffer before resuspension in 200 µL PBS. Ten µL of each well were then transferred into 96-well plates (Costar) with 90 µL of PBS containing 3 µg/mL AlexaFluor-488 conjugated concanavalin A (ConA) lectin (Sigma) and incubated at 37°C and 1,000 rpm for 30 min in a Vortemp 56 benchtop shaker (Labnet International). Absorbance at 600 nm (Abs_{600}) was measured for each well on a plate reader (BioTek Synergy).

3.5.3 Flow cytometry data collection and analysis for $Man_3GlcNAc_2$ glycan production

Five µL from each 96-well plate was diluted into 195 µL PBS in a new FACS round-bottom 96-well plate. The plate was read on an Accuri C6 Plus flow cytometer (BD Biosciences). Data for the following parameters were collected on the flow cytometer: forward scatter (FSC), side scatter (SSC), and AlexaFluor-488 fluorescence (488 nm excitation, 525 nm emission). Three to ten µL of each sample was measured. All samples were collected with 10,000 to 50,000 counts. Counts were gated in FSC versus SSC by choosing a window surrounding the largest cluster of cells. Fluorescence values were recorded in relative channel number (1-262,144 corresponding to 18-bit data) and the geometric mean over the gated data was calculated for each sample. Data analysis and FACS calibration was performed as described previously(Lucks, Qi, Mutalik, Wang, & Arkin, 2011). Rainbow calibration particles (Spherotech) were used to obtain a calibration curve to convert fluorescence intensity (geometric mean, relative channel number) into molecules of equivalent fluorescein (MEFL) units. Mean MEFL values were calculated over replicates. All flow cytometry experiments involved three biological replicates. Extraction of Flow Cytometry Standard (FCS) files, traditional gating, analysis of calibration beads data, standard curve generation, transformation to MEFL, and generation of histograms was performed using FlowCal software(Castillo-Hair et al., 2016).

3.5.4 Flow cytometry data collection and analysis for sfGFP fluorescence analysis

All flow cytometry experiments were performed in *E. coli* strain TG1 (*F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB)*). Cells were transformed with the relevant combination of target and trigger/LASO plasmids or target plasmid with pJBL002 (a blank sRNA/LASO plasmid). An autofluorescence control was included by transforming blank target and trigger plasmids (pJBL001 and pJBL002, respectively) which were both taken from Lucks *et al*.(Lucks et al., 2011). Plasmid combinations were transformed into chemically competent *E. coli* TG1 cells, plated on Difco LB+Agar plates containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol, and grown overnight at 37 °C. Following overnight incubation, plates were left at room temperature for approximately 9 h. Individual colonies were grown overnight in LB, then diluted 1:50 into M9 minimal media. After 6 h, cells were diluted 1:100 in 1x Phosphate Buffered Saline (PBS) containing 2 mg/mL kanamycin. A BD Accuri C6 Plus flow cytometer fitted with a high-throughput sampler was then used to measure sfGFP fluorescence. Measurements were taken for at least 6 biological replicates collected over at least two days.

Flow cytometry data analysis was performed using FlowJo (v10.4.1). Cells were gated by FSC-A and SSC-A, and the same gate was used for all samples. Samples exhibiting non-unimodal fluorescence distributions were excluded from further analysis, and the geometric mean fluorescence was calculated for each sample. All fluorescence measurements were converted to Molecules of Equivalent Fluorescein (MEFL) using CS&T RUO Beads (BD cat#661414), which were run on each day of data collection. The

average fluorescence (MEFL) over replicates of cells expressing empty plasmids (pJBL001 and pJBL002)[37] was then subtracted from each measured fluorescence value.

3.5.5 In vivo bulk fluorescence data collection and analysis

Bulk fluorescence experiments were performed in E. coli strain MC4100 gmd::kan (Valderrama-Rincon et al., 2012). E. coli strain MC4100 transformed with pJBL001 and pJBL002 served as the autofluorescence control. Transformed cells were plated on Difco LB+Agar plates containing 100 µg/mL carbenicillin and grown overnight at 37 °C. Following overnight incubation on plates, individual colonies were grown overnight in LB, then diluted 1:50 into M9 minimal media. After 4 h, mRFP fluorescence, and optical density (OD600) were measured using a Biotek Synergy plate reader. Measurements were taken for at least 3 biological replicates.

Bulk florescence data for mRFP were analyzed in a similar manner to the flow cytometry experiments, except that all fluorescence values were normalized to absorbance (OD_{600}). First, the OD_{600} of a blank well (containing only M9 media) was subtracted from the OD_{600} of each well. Then, FL/OD_{600} was calculated for each well containing cells; the average FL/OD_{600} of the autofluorescence control was then subtracted from each experimental well to determine the final, normalized value of FL/OD_{600} for each condition. Following calculations of average and standard deviation for each experimental condition, data were re-normalized such that the ON fluorescence was set to 1.

3.6 Acknowledgements

Thank you to Paul Carlson for help designing and collecting data for repression of endogenous genes with LASOs.

3.7 References

Carlson, P. D., Glasscock, C. J., & Lucks, J. B. (2018). De novo Design of Translational RNA Repressors. *BioRxiv*. https://doi.org/10.1101/501767

Castillo-Hair, S. M., Sexton, J. T., Landry, B. P., Olson, E. J., Igoshin, O. A., & Tabor, J. J. (2016). FlowCal: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.5b00284

Chae, T. U., Kim, W. J., Choi, S., Park, S. J., & Lee, S. Y. (2015). Metabolic engineering of Escherichia coli for the production of 1,3-diaminopropane, a three carbon diamine. *Scientific Reports*. https://doi.org/10.1038/srep13040

Datsenko, K. A., & Wanner, B. L. (2002). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.120163297

Ghodasara, A., & Voigt, C. A. (2017). Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx530

Glebes, T. Y., Aucoin, H. R., Kim, J. Y. H., Gill, R. T., Warner, J. R., Sandoval, N. R., & Reeder, P. J. (2012). Strategy for directing combinatorial genome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1206299109

Lee, Y. J., Kim, S. J., Amrofell, M. B., & Moon, T. S. (2019). Establishing a Multivariate Model for Predictable Antisense RNA-Mediated Repression. *ACS Synthetic*

Biology. https://doi.org/10.1021/acssynbio.8b00227

Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., & Arkin, A. P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1015741108

Meyer, S., Chappell, J., Sankar, S., Chew, R., & Lucks, J. B. (2016). Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. *Biotechnology and Bioengineering*. https://doi.org/10.1002/bit.25693

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., & Lee, S. Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2461

Noh, M., Yoo, S. M., Kim, W. J., & Lee, S. Y. (2017). Gene Expression Knockdown by Modulating Synthetic Small RNA Expression in Escherichia coli. *Cell Systems*. https://doi.org/10.1016/j.cels.2017.08.016

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., … DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.921

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. https://doi.org/10.1038/nature08187

CHAPTER 4 - *Dynamic control of pathway expression with riboregulated switchable feedback promoters*[3]

4.1 Abstract

Dynamic pathway regulation has emerged as a promising strategy in metabolic engineering for improved system productivity and yield and continues to grow in sophistication. Bacterial stress-response promoters allow dynamic gene regulation using the host's natural transcriptional networks but lack the flexibility to control the expression timing and overall magnitude of pathway genes. Here, we report a strategy that uses RNA transcriptional regulators to introduce another layer of control over the output of natural stress-response promoters. This new class of gene expression cassette, called a riboregulated switchable feedback promoter (rSFP), can be modularly activated using a variety of mechanisms, from manual induction to quorum sensing. We develop and apply rSFPs to regulate a toxic cytochrome P450 enzyme in the context of a Taxol precursor biosynthesis pathway and show this leads to 2.4x fold higher titers than from the best reported strain. We envision that rSFPs will become a valuable tool for flexible and dynamic control of gene expression in metabolic engineering, protein and biologic production, and many other applications.

[3] The results presented in this chapter have been submitted for publication:

Glasscock, C. J., Lazar, J. T., Biggs, B. W., Arnold, J. A., Kyoung-Kang, M., Tullman-Ercek, D., Tyo, E. J., Lucks, J. B. (2019) Dynamic control of pathway expression with riboregulated switchable feedback promoters. *In revision (Nature Biotechnology).* DOI 10.1101/529180

4.2 Introduction

Sustainable production of chemicals and materials in microbes through metabolic engineering (Keasling, 2010; Nielsen & Keasling, 2016) is a long-standing focus of synthetic biology. A primary challenge in metabolic engineering is the burden and toxicity on engineered cells owing to heterologous enzyme expression and unnecessary intracellular accumulation of toxic pathway intermediates (Biggs, De Paepe, Santos, De Mey, & Kumaran Ajikumar, 2014; Sun, Jeffryes, Henry, Bruner, & Hanson, 2017). This deleterious effect to the host often results in a loss to productivity and yield, creating a pressing need for strategies that can alleviate or avoid these pitfalls. This is nontrivial because each pathway can present unique cellular stresses, making it difficult to find generalizable solutions.

Synthetic biologists have sought to alleviate pathway toxicity by using dynamic pathway regulation to precisely tune the level and timing of enzyme expression (Brockman & Prather, 2015; Farmer & Liao, 2000; S. Z. Tan & Prather, 2017; Venayak, Anesiadis, Cluett, & Mahadevan, 2015). These systems are designed to adaptively adjust enzyme expression in response to changes in growth phase, cellular stress, fermentation conditions, and pathway intermediate concentrations so that they maintain an optimal concentration of enzymes that can vary over time. In order to implement these designs, synthetic biologists have created synthetic feedback networks that dynamically control gene expression using regulatory parts, such as engineered transcription factors (Moser et al., 2018; Xu, Li, Zhang, Stephanopoulos, & Koffas, 2014; Zhang, Carothers, & Keasling, 2012) or ligand-induced ribozymes (Carothers, Goler, Juminaga, & Keasling, 2011), that respond to relevant cues.

While these systems represent important advances, synthetic feedback networks are often difficult to construct because the sensors required for specific inputs are hard to design or source from nature, and the added burden of expressing regulatory components can itself negatively impact the host (C. Tan, Marguet, & You, 2009). Ultimately, this means that synthetic feedback networks require considerable additional engineering to match the specific requirements of every application. On the other hand, nature has evolved stress-responsive feedback networks that are already compatible with host cells. This is a result of the fact that microbial cells persist and thrive in changing environments due to their ability to sense and respond to stresses and environmental conditions (Aertsen & Michiels, 2004). Much of this ability is encoded within regulatory elements called stress-response promoters that integrate signals from complex and interconnected transcriptional networks to modulate mRNA synthesis in response to specific cellular stresses (Belliveau et al., 2018). This creates the possibility of using stress-response promoters to regulate heterologous pathway expression as a means to implement genetic feedback networks that lead to improvements in productivity and yield. In fact, synthetic biologists have used stress-response promoters to control pathway expression, leading to notable improvements to productivity and yield for protein expression (Ceroni et al., 2018) and industrially important pathways, such as the artemisinin precursor amorphadiene (Dahl et al., 2013) and n-butanol (Boyarskiy, Davis López, Kong, & Tullman-Ercek, 2016). However, stress-response promoters have not been widely adopted, as their complexity makes it difficult to fine-tune their behavior for specific applications. As their function is determined by the complex topologies of natural genetic networks, there are no simple methods to tune either the timing or overall magnitude of

their transcriptional outputs – two key parameters that are important for optimizing metabolic pathway productivity and yield (Jones et al., 2015).

To address this limitation, we sought to create a new regulatory motif called a switchable feedback promoter (SFP) that combines the feedback properties of natural stress-response promoters with regulators that offer control of the timing and overall magnitude of transcriptional outputs (**Figure 4-1A-C**). The SFP concept is general, and can be implemented in several ways including engineering transcription factor operator sites within the stress-response promoter region (Lutz & Bujard, 1997). However, the architecture of many stress-response promoters is still unknown, making the rational design of transcription factor-based SFPs difficult. Instead, we utilized *trans*-acting synthetic RNA regulators (Chappell, Takahashi, & Lucks, 2015; Green, Silver, Collins, & Yin, 2014), which can be configured to control transcription in the case of small transcription activating RNAs (STARs) (Chappell et al., 2015), or translation in the case of toehold switches (Green et al., 2014). The key feature of both STARs and toehold switches is that they have well-defined composition rules such that they can be inserted into a gene expression construct without modification or disruption of the desired promoter sequence. In this way, riboregulated SFPs (rSFPs) can be created by inserting a STAR or toehold target binding region 3' of a natural stress-response promoter (Fig. 1B). By default, when transcribed, these RNA sequences fold into structures that block gene expression. However, binding of the trans-acting STAR or toehold trigger RNA changes these structures to allow gene expression. This enables the timing and overall magnitude of the rSFP output to be controlled with any strategy that can regulate the expression of the *trans*-acting RNA (**Fig. 4-1C**). These regulators have additional

advantages that make inclusion in rSFPs promising, including the availability of large libraries of orthogonal STARs and toehold switches with a range of functional properties (ON/OFF levels) (Chappell, Westbrook, Verosloff, & Lucks, 2017; Green et al., 2014) and their having compact DNA footprints (<100nts). Furthermore, they have been shown to control gene expression in a variety of contexts, including within metabolic pathways (Chappell et al., 2017).

Here we report the creation and characterization of a library of STAR-mediated rSFPs, and their application to optimizing the yield of a metabolic pathway that produces an oxygenated taxane precursor to the anticancer drug Taxol. We first show that we can create a library of 17 rSFPs by interfacing STARs with natural *Escherichia coli* stress-response promoters and placing *trans*-acting STAR production under control of an inducible promoter. We then applied rSFPs to control the expression of a plant cytochrome P450 that is known to cause envelope stress (Biggs et al., 2016) in the context of a pathway that produces an oxygenated Taxol precursor (Ajikumar et al., 2010). By screening rSFPs for oxygenated taxane production, we were able to find multiple rSFPs that showed improvement in both overall and oxygenated taxane titers compared to the previously reported best strain. We next used the external control of rSFPs to systematically optimize both timing and expression level to ultimately find pathway conditions that produce 25.4 mg/L of oxygenated taxanes and 39.0 mg/L of total taxanes, representing a 2.4x and a 3.6x fold improvement over the current state-of-the-art, respectively. To demonstrate the use of other control points for rSFPs, we next sought to interface them with a quorum sensing system and show that quorum sensing rSFPs

offer completely autonomous pathway expression regulation with yields similar to our fully optimized system without costly external inducers.

Overall, rSFPs are a novel and general strategy to achieve dynamic regulation of metabolic pathway enzymes and we envision them to be broadly useful for introducing controllable stress-response promoters in many synthetic biology applications.

4.3 Results

4.3.1 Riboregulated switchable feedback promoters (rSFPs) enable tunable outputs from stress-response promoters

We chose to build rSFPs with STARs because they exhibit low leak and high dynamic range comparable to exemplary protein-based regulators (Chappell et al., 2017). STARs activate transcription by disrupting the folding pathway of a terminator hairpin sequence, called a Target, that is placed upstream of the gene to be regulated (**Fig. 4-1B**). In the absence of a STAR, the Target region folds into an intrinsic terminator hairpin which stops transcription before reaching the downstream gene. When present, a STAR RNA can bind to the 5' portion of the terminator hairpin, preventing its formation, and allowing transcription. rSFPs are then created by inserting a Target sequence 3' of a candidate stress-response promoter. In this way, the introduction of the STAR/Target adds an additional layer of control to the stress-response promoter, effectively gating its transcriptional output through the additional regulation of STAR RNA expression, which can be controlled using a variety of mechanisms, including manual inducible promoters or quorum-sensing systems.

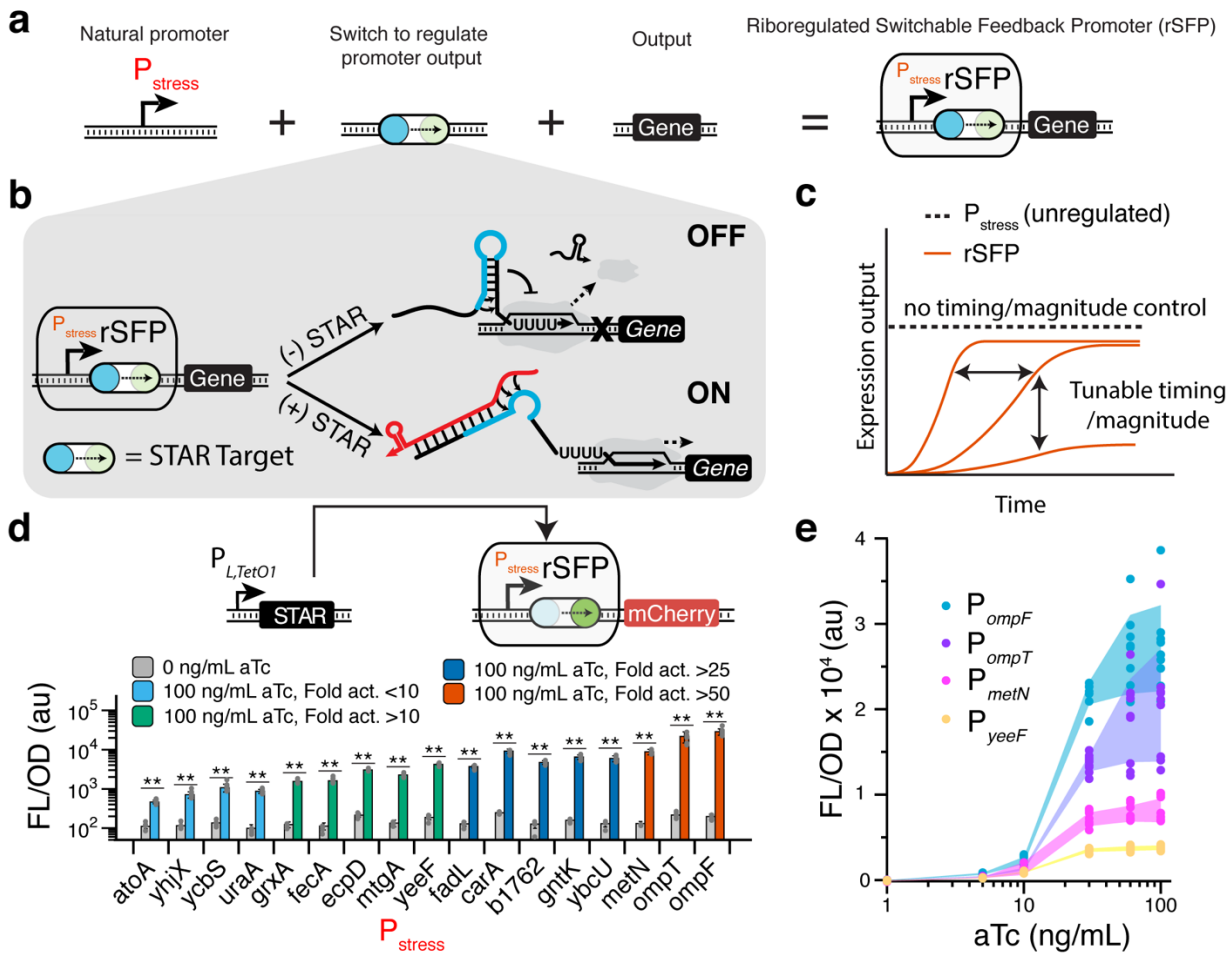


Figure 4-1. Riboregulated switchable feedback promoters.

(**A**) Riboregulated switchable feedback promoters (rSFPs) are composed of a natural stress-response promoter and an RNA transcriptional switch that allows control over the output of native stress-mediated transcriptional networks. (**B**) Schematic of the small transcription activating RNA (STAR) transcriptional switch mechanism used in rSFPs. A Target DNA sequence (switch symbol) is placed 3' of a stress-response promoter. The transcribed Target RNA is designed to fold into an intrinsic transcription terminator hairpin, composed of a hairpin structure followed immediately by a poly uracil sequence. The formation of this terminator hairpin causes RNA polymerase (RNAP) to terminate transcription upstream of the gene to be regulated (gene OFF). A separately transcribed

STAR RNA (colored red) can bind to both the linear region and the 5' half of the terminator hairpin (colored blue) of the Target RNA, preventing its formation and allowing transcription elongation of the downstream gene (gene ON). In this way the output of the stress-response promoter is controlled by STAR expression, which adds an additional layer of regulation to that present within the stress-mediated transcriptional network that governs expression at the stress-response promoter. (**C**) Illustration of expression control enabled by rSFPs. Natural stress-response promoters (dashed line) can exhibit dynamic behaviors in response to stress but are fixed with regards to user-defined timing and overall expression magnitude. rSFPs (red lines) use the additional layer of regulation to resolve this issue and allow control of timing and overall expression magnitude by gating transcriptional output with a *trans*-acting RNA regulatory switch. (**D**) Characterization of rSFP variants containing unique envelope stress-response promoters. $P_{L,TetO1}$ inducible STAR expression is used to activate rSFPs containing a natural stress-response promoter upstream of a STAR Target sequence, a ribosome binding site, and a red fluorescent protein (mCherry) coding sequence. Fluorescence characterization was performed on *E. coli* transformed with plasmids encoding each rSFP controlling mCherry expression in the absence and presence of 100 ng/mL aTc. (**E**) rSFPs enable titration of natural stress-response promoter output. Fluorescence characterization performed on *E. coli* cells containing rSFPs controlling mCherry expression under different levels of aTc induction. Data in **d** represent mean values in units of arbitrary fluorescence/optical density (FL/OD) and error bars represent s.d. of at least n = 7 biological replicates. * indicate a statistically significant difference in FL/OD by a two-tailed paired-sample t-test (* = $P < 0.05$, ** = $P < 0.005$). P values for each condition are reported in SI Table 8. Grey

points represent individual data points for each condition. Shaded area in **e** represents mean values in units of arbitrary fluorescence/optical density (FL/OD) +/- s.d. of at least n = 7 biological replicates. Colored points represent individual data points for each condition.

Our initial rSFP designs utilized a previously developed STAR (Chappell et al., 2017) under the well-characterized inducible system TetR/$P_{L,TetO1}$ (Lutz & Bujard, 1997) interfaced with a library of 17 putative membrane stress-responsive promoters (Boyarskiy et al., 2016; Dahl et al., 2013). These promoters were chosen as several had been previously identified to regulate a biofuel transporter protein in *E. coli* (Boyarskiy et al., 2016), and could be valuable for dynamic regulation of membrane proteins in metabolic pathways. To construct and characterize these rSFPs, a STAR Target sequence was cloned immediately 3' of each promoter to regulate expression of an mCherry reporter, and its cognate STAR was cloned in a second $P_{L,TetO1}$ plasmid. Plasmids were transformed into *E. coli* and fluorescence was measured with and without the presence of the $P_{L,TetO1}$ inducer anhydrotetracycline (aTc) at saturating levels (100 ng/mL). We found that induction of $P_{L,TetO1}$-STAR resulted in significant activation from all members of the stress-response promoter library (**Fig. 4-1D**), exemplifying the modularity of the rSFP concept. We also observed that 8 library members were activated by greater than 25x fold in the presence of aTc, with a maximum activation of nearly 150x fold (**SI Fig. B.8-1**). We next selected a set of high-performing rSFPs and characterized their transfer functions by titrating levels of aTc and found that all exhibited a similar transfer function shape, though with different maximal activation levels (**Fig. 4-1E**). This is evidence that

the transfer function of the $P_{L,TetO1}$ regulatory system can be overlaid on a range of stress response promoters through the STAR intermediate. Overall these results demonstrated that we can create a library of rSFPs that provide tunable control of gene expression level by selecting different stress-response promoters and manipulating inducer concentration.

4.3.2 rSFPs enhance production of an oxygenated Taxol precursor

We next tested the ability of rSFPs to regulate expression of a challenging metabolic pathway enzyme. As a model system, we chose a portion of the anticancer drug Paclitaxel's biosynthesis pathway that has been previously reconstituted in *E. coli* (Biggs et al., 2016). Specifically, we focused on the first P450-mediated step where taxadiene is oxygenated by the membrane anchored cytochrome P450 CYP725A4 (**Fig. 4-2A**). This system is an ideal test bed for the use of rSFPs because CYP725A4 expression causes membrane stress due to lipid anchoring of an N-terminal domain. This stress appears to reduce pathway productivity and makes pathway optimizations extremely difficult (Biggs et al., 2016). The sensitivity of product titers to expression level of CYP725A4 must be carefully managed, as too low expression will create a bottleneck in oxygenated taxane synthesis, but too high expression will also suppress synthesis due to stress. Previous reports to optimize this system required significant experimental effort(Biggs et al., 2016), exemplifying the importance of new pathway engineering strategies, but also providing a competitive benchmark for comparison with rSFPs. Furthermore, this problematic expression is not unique to CYP725A4, but extends to many P450's (Chang, Eachus, Trieu, Ro, & Keasling, 2007; Leonard & Koffas, 2007), along with other classes of proteins such as transporters (Boyarskiy et al., 2016) and

glycosylation enzymes (Glasscock et al., 2018; Valderrama-Rincon et al., 2012). Therefore, the system is a model challenge for testing the concept of using rSFPs to leverage external control of natural stress-response promoters to maintain pathway expression in a narrow optimal range.

Previous work has shown that expression level of a CYP725A4/tcCPR reductase fusion is critical to achieving high titers of oxygenated taxanes in *E. coli*(Biggs et al., 2016). A previously optimized low-copy expression vector (p5Trc-CYP725A4/tcCPR) (**Fig. 4-2B**) transformed into the *E. coli* Tax1 strain containing genomic modifications to maximize the synthesis of taxadiene, produces ~11 mg/L of oxygenated taxanes, albeit with a loss to total taxane production. However, increasing expression of the enzyme using a medium copy expression vector (p10Trc) does not increase titer, but causes a complete loss of pathway productivity (**Fig. 2C**), presumably due to the enzyme's membrane stress crossing a critical threshold and triggering a global response.

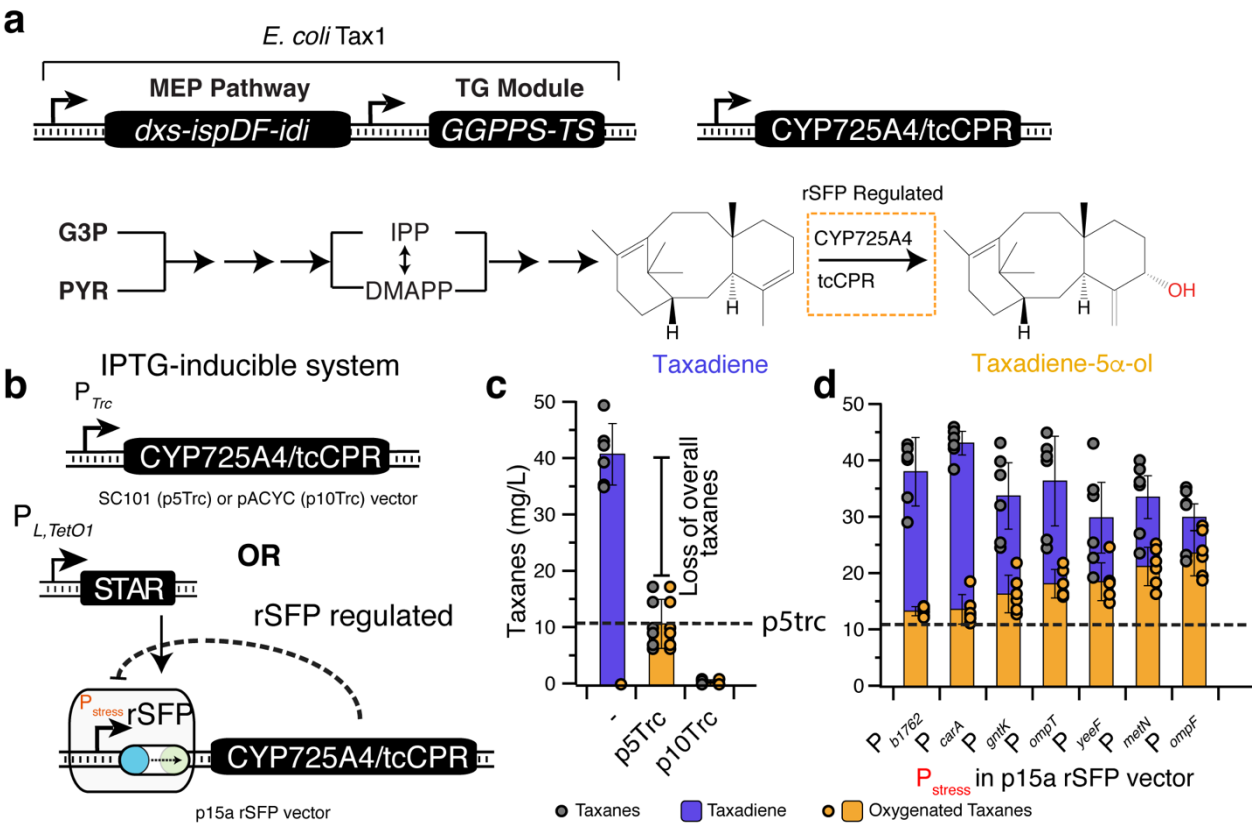


Figure 4-2. rSFPs enhance productivity of a Taxol precursor synthesis pathway in *E. coli*.

(**A**) Taxol biosynthesis schematic depicting an abbreviated overview of the Taxol precursor pathway involving the toxic cytochrome P450 725A4 (CYP725A4) enzyme. In the *E. coli* strain Tax1, the methylerythritol phosphate (MEP) pathway and taxadiene synthase/geryanlgeranyl diphosphate (GGPP) synthase (TG) module convert glyceraldehyde-3-phosphate (G3P) and pyruvate (PYR) into the 20-carbon backbone taxa-4 (5),11 (12)-diene (taxadiene). Taxadiene is oxygenated by the membrane-anchored CYP725A4 fused with its reductase partner (tcCPR) to form taxadiene-5α-ol. rSFPs utilizing envelope stress-response promoters are applied to control the expression of CYP725A4/tcCPR. IPP = isopentenyl diphosphate, DMAP = dimethylallyl diphosphate.

(B) Plasmids used for CYP725A4/tcCPR expression in *E. coli* Tax1. CYP725A4/tcCPR is expressed from a standard IPTG-inducible P_{Trc} promoter in either a low copy (p5Trc, SC101) or medium copy (p10Trc, pACYC) plasmid, or from rSFPs with various P_{stress} promoters encoded on a p15a plasmid. $P_{L,TetO1}$-STAR was used to activate rSFP expression. rSFPs allow CYP725A4/tcCPR expression to be controlled by externally supplied aTc and feedback-regulated by the natural stress response pathways. p5Trc is a gold standard CYP725A4/tcCPR expression system from previously reported optimization efforts. (**C**) Titers of fermentations with empty *E. coli* Tax1 and *E. coli* Tax1 containing low copy p5Trc or medium copy p10Trc expression of CYP725A4/tcCPR. Addition of p5Trc to *E. coli* Tax1 enables production of oxygenated taxanes at a cost to overall taxane production, while addition of p10Trc eliminates nearly all taxane production presumably due to the toxicity of CYP725A4/tcCPR expression. (**D**) Titers of fermentations with *E. coli* Tax1 containing CYP725A4/tcCPR under control of high performing rSFPs. Dashed line represents production of oxygenated taxanes from p5Trc. The P_{ompF} rSFP resulted in ~2.2x fold greater oxygenated taxanes (~23.5 mg/L) and ~2.8x fold greater overall taxanes (29.8 mg/L) than the p5Trc strain.Data in **c, d** represent mean values of fermentation titers after 96 hrs and error bars represent s.d. of at least n = 5 biological replicates. Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.

We hypothesized we could achieve greater pathway productivity over the p5Trc benchmark strain by identifying putative envelope stress rSFPs for control of CYP725A4/tcCPR. To test this, the CYP725A4/tcCPR coding sequence was introduced

into each one of the 17 rSFP constructs. *E. coli* Tax1 was transformed with each rSFP construct and the $P_{L,TetO1}$-STAR plasmid and each tested in the context of taxadiene oxygenation fermentations with addition of 100 ng/mL aTc at inoculation. Using this approach, we found that several performed well against the p5Trc benchmark strain (**SI Fig. B.8-2**). In particular, 7 of the rSFPs had greater titers of oxygenated taxanes than the p5Trc strain (**Fig. 4-2D**), with all also improving overall taxane production. Furthermore, the P_{ompF} rSFP resulted in ~2.2x fold greater oxygenated taxanes (~23.5 mg/L) and ~2.8x fold greater overall taxanes (29.8 mg/L) than the p5Trc strain, clearly showing the benefit of rSFP pathway regulation.

To confirm that rSFPs can indeed be feedback regulated by CYP725A4/tcCPR stress, we performed fluorescence analysis of *E. coli* cells containing plasmids for rSFP expression of an mCherry reporter with the top two performing stress-response promoters and the p10Trc plasmid separately expressing CYP725A4/tcCPR, in order to monitor changes in rSFP expression caused by membrane stress (SI Fig. B3A). We observed reduced expression from P_{ompF} when p10Trc was present in place of an empty vector (**SI Fig. B.8-3B**), suggesting that it is indeed responsive to CYP725A4/tcCPR induced stress. On the other hand, a constitutive promoter control had no response as expected. Interestingly, the P_{metN} rSFP did not exhibit a reduced expression in response to CYP725A4/tcCPR expression, indicating that not all rSFPs respond to stresses in the same way, as we expected for a diverse set of natural stress-response promoters.

Overall these results show that rSFPs can be effectively used to optimize overall pathway expression and that they can exhibit the dynamic feedback behaviors of incorporated stress-response promoters.

4.3.3 rSFPs allow further pathway optimization through the control of expression timing and overall magnitude

Having shown significant improvements in taxadiene oxygenation with rSFPs, we next sought to test how the external control offered by rSFPs can be used to further optimize induction level and timing of stress-response promoter activity. To test this, we selected the two best rSFP systems and performed a matrix of aTc induction at four levels (0, 16, 32, and 100 ng/mL aTc), which were added at six different induction times (0, 3, 6, 12, 24, 48 hrs) post fermentation inoculation (**Fig. 4-3A-B**). We found that oxygenated taxane production with both rSFPs was indeed sensitive to induction level and timing (**Fig. 4-3C,D**) and that late induction of P_{metN} and P_{ompF} rSFPs could improve final titers of oxygenated taxanes even further to 25.4 and 25.1 mg/L, respectively, and overall taxanes to 39.0 and 31.0 mg/L (**Fig. 4-3E,F**), representing an overall 2.4x and 2.3x fold improvement over the previous gold standard benchmark in terms of oxygenated taxanes, and 3.6x and 2.9x fold improvements in terms of overall taxanes. These results demonstrate that rSFPs can be implemented to enable rapid tuning of expression timing and overall magnitude of stress-response promoter output to further enhance fermentation titers.

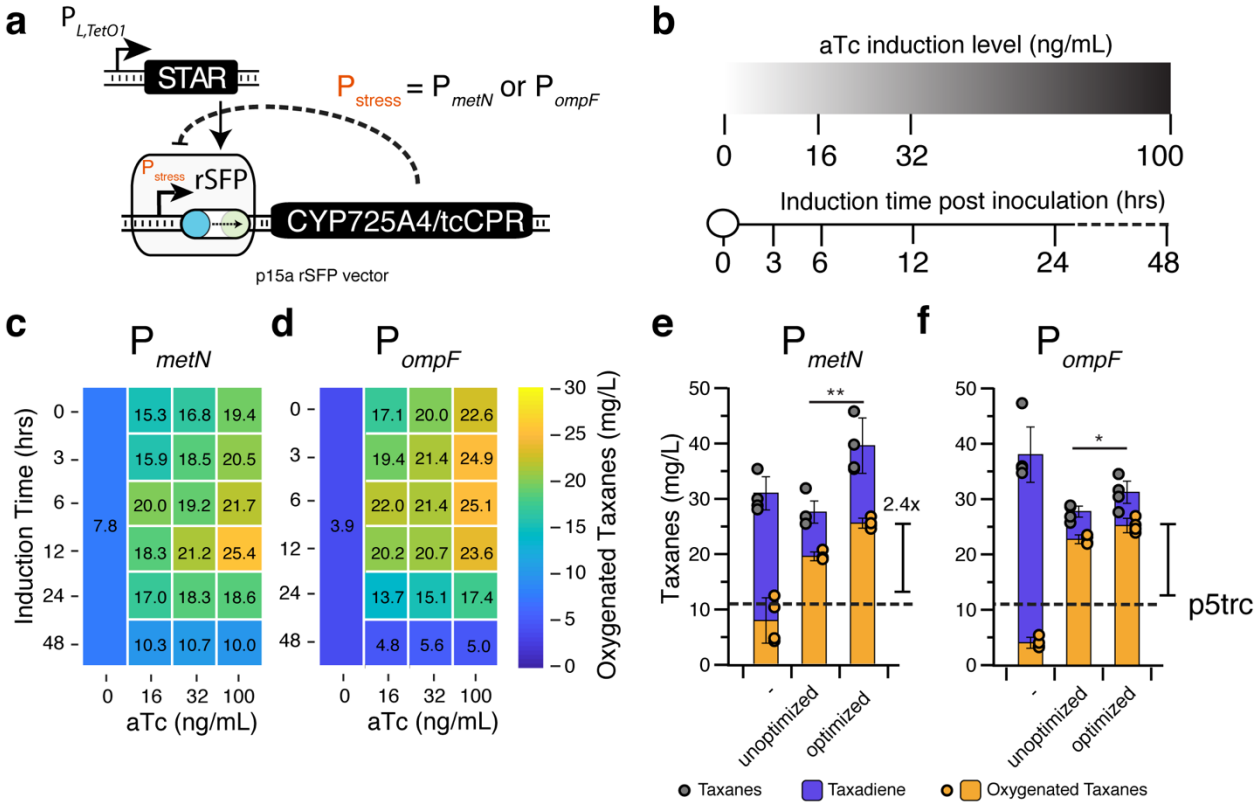


Figure 4-3. External control of rSFPs enable optimization of induction level and timing from stress-response promoters.

(A) Plasmids used for CYP725A4/tcCPR expression from rSFPs with P_{stress} promoters encoded on a p15a plasmid in *E. coli* Tax1. $P_{L,TetO1}$-STAR was used to activate rSFP expression. **(B)** Conditions of aTc induction level and timing used for rSFP CYP725A4/tcCPR expression optimization. (**C, D**) Induction level and timing optimization of fermentations with *E. coli* Tax1 containing the P_{metN} (**c**) or P_{ompF} (**d**) rSFP controlling CYP725A4/tcCPR. Heatmap shows the average oxygenated taxane titers for different combinations of aTc concentration and time of induction. (**E, F**) Titers of fermentations with *E. coli* Tax1 containing the P_{metN} (**e**) or P_{ompF} (**f**) rSFP before (100 ng/mL aTc at 0 hrs) and after induction optimization. Dashed line represents production of oxygenated taxanes from p5Trc. Data in **c, d, e, f** represent mean values of fermentation titers after

96 hrs and error bars represent s.d. of at least n = 3 biological replicates. * indicate a statistically significant difference in oxygenated taxane production by a two-tailed Welch's t-test (* P = 0.0211, ** P = 0.0000822). Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.

4.3.4 Quorum-sensing activated rSFPs allow autonomous regulation of pathway expression

Though inducible systems offer flexibility for screening of optimal induction timing, the cost of inducers can be prohibitive at an industrial scale (Van Dien, 2013; Weber & Fussenegger, 2007), and several efforts have been carried out to design autonomous means of induction. Quorum-sensing (QS) systems that are activated in a cell-density dependent manner offer one such route to this behavior (Papenfort & Bassler, 2016). QS systems have been used with great utility in metabolic engineering to create a separation of cell growth and pathway production phases without the need for a chemical inducer, and provide a natural means for balancing carbon utilization with biomass production (Gupta, Reizman, Reisch, & Prather, 2017; Kim et al., 2017; Tsao, Hooshangi, Wu, Valdes, & Bentley, 2010). We therefore sought to utilize this strategy within our model pathway by leveraging the modularity of rSFPs to be easily configured to utilize different input systems. Specifically, we chose the P_{Lux} promoter that is activated by the LuxR transcriptional activator upon sufficient production of the C6-homoserine lactone (HSL) signaling molecule (Engebrecht & Silverman, 1984). We cloned a STAR under control of P_{Lux} and integrated an operon with the EsaI HSL synthase (Minogue, Wehland-Von

Trebra, Bernhard, & Von Bodman, 2002) and LuxR into the genome of the *E. coli* Tax1 strain to create the Tax1-QS strain (**Fig. 4-4A**). When plasmids encoding the expression of P_{Lux}-STAR and the P_{metN} or P_{ompF} rSFPs controlling mCherry expression were transformed into *E. coli* Tax1 or Tax1-QS, we found that activation only occurred in the engineered Tax1-QS strain containing EsaI and LuxR (**Fig. 4-4B**). These QS-activated rSFPs produced comparable fold activation to manual induction with $P_{L,TetO1}$ and exhibited a time dependent activation.

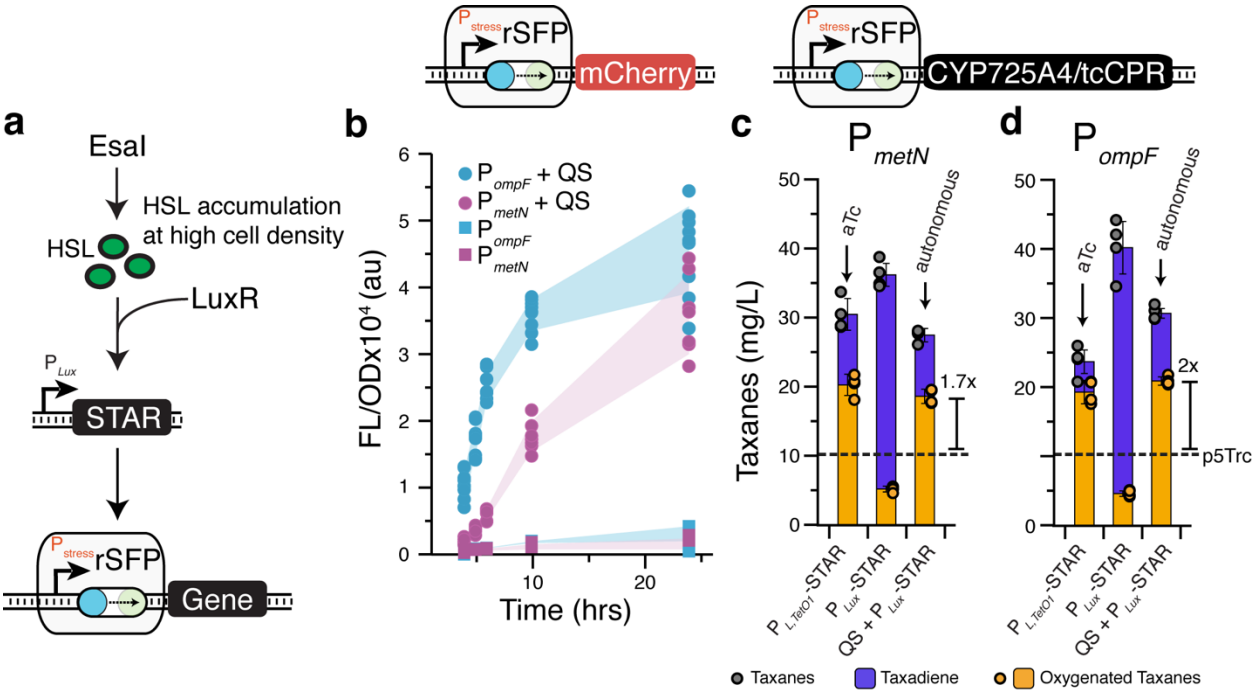


Figure 4-4. Quorum sensing activation of rSFPs allows autonomous control of CYP725A4 expression.

(**A**) Schematic showing quorum-sensing (QS) activation of rSFPs to allow autonomous control of pathway expression. LuxR is activated by C6-HSL produced by the EsaI HSL synthase upon sufficient accumulation due to an increase in cell density. LuxR activation results in STAR production from the P_{Lux} promoter, thereby activating rSFP expression.

(**B**) Fluorescence experiments showing autonomous activation of P_{ompF} and P_{metN} rSFPs, configured to control mCherry expression, over time. Fluorescence is only significant with the Tax1-QS strain containing a chromosomal LuxR/EsaI expression cassette, but not with the parent Tax1 strain. (**C, D**) Titers of oxygenated taxadiene fermentations in *E. coli* Tax1 strains containing the P_{metN} (**c**) or P_{ompF} (**d**) rSFP controlling CYP725A4/tcCPR expression. Left: in *E. coli* Tax1 (without QS insert) containing a $P_{L,TetO1}$-STAR plasmid induced by aTc; middle: in *E. coli* Tax1 containing a P_{Lux}-STAR plasmid; right: in *E. coli* Tax1-QS containing a P_{Lux}-STAR plasmid activated by EsaI produced HSL. QS-activated rSFPs obtained similar titers to the unoptimized $P_{L,TetO1}$-STAR activated rSFPs, but without any external interventions. Dashed line represents production of oxygenated taxanes from p5Trc Fig. 2C. Shaded area in **b** represents mean values in units of arbitrary fluorescence/optical density (FL/OD) +/- s.d. of at least n = 7 biological replicates. Colored points represent individual data points for each condition. Data in **c** represent mean values of fermentation titers after 96 hrs and error bars represent s.d. of at least n = 3 biological replicates. Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.

To demonstrate that QS-activated rSFPs could be used to autonomously control the expression of metabolic pathway enzymes, we applied the P_{metN} and P_{ompF} QS-activated rSFPs to control the expression of CYP725A4/tcCPR within the taxadiene oxygenation pathway. Fermentations were performed by inoculating cell cultures into medium without addition of exogenous inducer. Upon fermentation and analysis, we found that QS-based activation resulted in comparable titers of oxygenated taxanes to

those obtained from manual induction of rSFPs with aTc before optimization (**Fig. 4-4C,D**). Importantly, this represented 1.7x and 2x fold improvements, respectively, over the previous gold standard, and was achieved with a completely autonomous genetic feedback network without the need for costly inducers.

4.4 Discussion

Here we report the development, characterization and application of switchable feedback promoters that enable an additional synthetic layer of control over natural stress-response promoters. Stress-response promoters are a promising route to achieving dynamic control of heterologous metabolic pathways by acting as sensor-actuators to stresses caused by pathway expression, intermediate metabolites and other fermentation conditions(Boyarskiy et al., 2016; Dahl et al., 2013). While stress-response promoters can improve production of desired chemicals by regulating expression in response to toxic pathway intermediates and enzymes, they are constrained by their complexity, leading to a lack of control over the timing and overall magnitude of their transcriptional output, which is essential to achieving a separation of growth phase and production phase in large-scale fermentations (Malinowski, 2001). By design, the rSFP concept enables this control by introducing an additional regulatory layer within the natural stress-response pathway by gating stress-response promoter outputs with *trans*-acting RNA regulators. The use of an inducible promoter to control RNA regulator synthesis allows modification of the timing and overall magnitude of the natural stress-response promoter outputs. Furthermore, the use of QS systems allows the autonomous activation of rSFPs in a cell-density dependent manner. In this way, rSFPs have

modularity both at the lever of their inputs and outputs, and the types of stresses they can respond to through changing of the regulated stress-response promoter. This offers the flexible implementation of controllable stress-response networks in a single compact locus.

In this work, we design and implement rSFPs and demonstrate that they are both modular and tunable – the rSFP concept can be applied to many unique stress-response promoters in a plug-and-play fashion, activator inputs can be easily interchanged, and activated output levels can be modulated by titrating inducer concentrations. Notably, we found that all 17 of the stress-response promoters that were inserted into rSFPs were activated significantly, strongly suggesting that rSFPs can be used with new stress-response promoters as they are discovered, and potentially that the rSFP concept can be used to easily regulate engineered promoter systems as well. These features allow rapid screening of rSFP libraries within combinatorial strain engineering procedures (Smanski et al., 2014) that could be used by industry to identify effective implementations of dynamic control. In addition, the ability of rSFPs to naturally adapt to an optimal expression level may allow for rapid prototyping of potentially toxic enzymes and pathways without the requisite need to first balance expression levels with constitutive static regulators – speeding the pace of pathway construction for new chemical products.

To demonstrate their utility in the context of optimizing metabolic pathway production, we applied rSFPs to a synthetic Taxol precursor pathway in *E. coli* (Ajikumar et al., 2010) by regulating expression of a problematic cytochrome P450 enzyme that causes a membrane stress detrimental to productivity (Biggs et al., 2016). By screening through a library of envelope-stress-response promoters in rSFPs, we identified variants

that improved pathway productivity over a previous strain that had been optimized using a laborious trial-and-error approach. Furthermore, we showed that optimizing rSFP induction timing and magnitude in the fermentation enabled additional improvements, highlighting an advantage of the rSFP system to enable the control of pathway expression timing. We also showed that rSFPs can be controlled by QS systems that do not require addition of an external inducer, enabling fully autonomous control of pathway expression.

Dynamic pathway regulation is a promising strategy in metabolic engineering but can be difficult to implement. The rSFP strategy enables modular and tunable control of endogenous promoters that have evolved sophisticated transcriptional responses to a range of cellular stresses and fermentation conditions. Due to their simplicity, we envision that the rSFP concept will enable streamlined implementation of dynamic regulation into metabolic pathways. Furthermore, given their modularity, we imagine rSFPs will be useful for dynamic control in other applications, such as high-level expression of difficult or toxic proteins, living therapeutics (Isabella et al., 2018), and cellular diagnostics (Watstein, McNerney, & Styczynski, 2015) where endogenous promoters could be used as sensor-actuators for numerous environments.

4.5 Methods

4.5.1 Plasmid assembly

All plasmids used in this study can be found in **Supplementary Table 1** with key sequences provided in **Supplementary Tables 2 and 3**. Gibson assembly and inverse PCR (iPCR) was used for construction of all plasmids. All assembled plasmids were verified using DNA sequencing.

4.5.2 Integration of QS operon into the E. coli genome

Strains containing genomic insertions of the EsaI-LuxR operon were created using the clonetegration (St-Pierre et al., 2013) platform as summarized in **Supplementary Table 4**. The HK022 plasmid was used to integrate constructs into the *attB* site of the *E. coli* genome. Successful integrations were identified by antibiotic selection and colony PCR according to the published protocol.

4.5.3 Strains, growth media, in vivo bulk fluorescence measurements

Fluorescence characterization experiments for all envelope stress-response promoters were performed in *E. coli* strain Tax1 (Biggs et al., 2016) containing the synthetic pathway for taxadiene biosynthesis or modified Tax1-QS containing the QS operon. Experiments were performed for 7-9 biological replicates collected over three separate days. For each day of fluorescence measurements, plasmid combinations were transformed into chemically competent *E. coli* cells and plated on LB+Agar (Difco) plates containing combinations of 100 µg/mL carbenicillin, 34 µg/mL chloramphenicol and/or 50 µg/mL spectinomycin depending on plasmids used (see Appendix B SI Table 1 for plasmids used in each experiment), and incubated approximately 17 hours (h) overnight at 37 °C. Plates were taken out of the incubator and left at room temperature for approximately 7 h. Three colonies were used to inoculate three cultures of 300 µL of LB containing antibiotics at the concentrations described above in a 2 mL 96-well block (Costar), and grown for approximately 17 h overnight at 37 °C at 1,000 rpm in a VorTemp 56 (Labnet) bench top shaker. **Figures 1d, 1e**: 4 µL of each overnight culture were added to 196 µL (1:50 dilution) of supplemented M9 minimal media (1 × M9 minimal salts, 1 mM thiamine hydrochloride, 0.4 % glycerol, 0.2 % casamino acids, 2 mM $MgSO_4$, 0.1 mM

$CaCl_2$) containing the selective antibiotics and grown for 6 h at the same conditions as the overnight culture. Appropriate concentrations of anhydrotetracycline (Sigma) were added to culture media as indicated. **Figure 4b:** 20 µL of each overnight culture were added to 980 µL of M9 minimal media containing selective antibiotics and grown for 24 h at 37C. Periodic samples of 10-200 µL of culture were collected for characterization by bulk fluorescence measurements. **For all bulk fluorescence measurements**: 10-200 µL of sampled culture were transferred to a 96-well plate (Costar) containing 0-190 µL of phosphate buffered saline (PBS). Fluorescence (FL) and optical density (OD) at 600 nm were then measured using a Synergy H1 plate reader (Biotek). The following settings were used: mCherry fluorescence (560 nm excitation, 630 nm emission).

4.5.4 Bulk fluorescence data analysis

On each 96-well block there were two sets of controls; a media blank and *E. coli* Tax1 cells transformed with combination of control plasmids JBL002 and JBL644 (blank cells) and thus not expressing mCherry (**Appendix B Supplementary Table 1**). The block contained three replicates of each control. OD and FL values for each colony were first corrected by subtracting the corresponding mean values of the media blank. The ratio of FL to OD (FL/OD) was then calculated for each well (grown from a single colony) and the mean FL/OD of blank cells was subtracted from each colony's FL/OD value. Three biological replicates were collected from independent transformations, with three colonies characterized per transformation (9 colonies total). Occasional wells were discarded due to poor growth (OD < 0.1 at measurement), however, all samples contained at least 7 replicates over the three experiments. Means of FL/OD were calculated over replicates and error bars represent standard deviations (s.d).

4.5.5 Small-scale "Hungate" fermentation

Small-scale fermentation assays were used to quantify oxygenated taxanes and taxadiene production in *E. coli* Tax1 or Tax1-QS. Experiments were performed with six biological replicates collected over three independent experiments (**Figure 2c**, **2d**) or four biological replicates collected over two independent experiments (**Figure 3c**, **3d, 3e, 3f, 4c, 4d**). For each experiment, plasmid combinations (**Appendix B SI Table 1**) were transformed into chemically competent *E. coli* cells and plated on LB+Agar (Difco) plates containing appropriate antibiotics (100 μg/mL carbenicillin, 34 μg/mL chloramphenicol and/or 50 μg/mL spectinomycin). Plates were incubated approximately 17 hrs overnight at 30 °C. Individual colonies were inoculated into culture tubes containing LB and appropriate antibiotics and incubated at 30°C for roughly 16 hrs overnight to achieve an approximate OD600 of 3. For 2 mL batch fermentations, 50 μL of overnight cells were added to 1.95 mL of complete R-media (**Appendix B Supplementary Tables 5-7**) and appropriate antibiotics in glass hungate tubes (ChemGlass). 0.1 mM IPTG was added for induction of the upstream pathway enzymes and p5Trc/p10Trc expression. 16-100 ng/mL aTc was added, as indicated, to induce $P_{L,TetO1}$-STAR activated rSFPs. A 10% v/v dodecane layer (200 μL) was added in all fermentations. Hungate tubes were sealed with a rubber septa and plastic screw-cap (ChemGlass). PrecisionGlide 18G hypodermic needles (BD) were inserted into the rubber septa to allow for gas exchange. Hungate tubes were incubated at 22°C and 250 rpm for 96 hrs. After the fermentations were completed, the culture was centrifuged to collect the dodecane overlay. This overlay was subsequently diluted into hexane for analytical procedures described below.

4.5.6 GC-MS analysis

Dodecane samples collected from batch fermentations were diluted at a ratio of 1:40 in n-hexane containing 5 mg/L ß-caryophyllene. The 5 mg/L ß-caryophyllene was utilized as a standard to calculate titer of taxadiene and oxygenated taxanes. GC-MS analysis was performed with an Agilent 7890 GC and Agilent HP-5ms-UI column (Ultra Inert, 30 m, 0.25 mm, 025 μm, 7 in cage). Helium was utilized as a carrier gas at a flow rate of 1 mL/min and the sample injection volume was 1 μL. The splitless method begins at 50 °C hold for 1 minute followed by a 10°C/min ramp to 200 °C and a final 5°C/min ramp to 270 °C. Mass spectroscopy data was collected for 22.5 minutes with an 11-minute solvent delay with an Agilent 7000 QQQ in scan mode using Electron Ionization (EI). m/z values ranging from 40-500 were scanned with a scan time of 528ms. MassHunter Workstation Qualitative Analysis software (vB.06.00) was utilized to integrate peaks on the chromatograms and determine their respective mass spectrums. The ratio of peak area of taxadiene (m/z 272) to the standard ß-caryophyllene (m/z 204) was used to calculate titer of taxadiene, while the ratio of the sum of all peaks of oxygenated taxanes (m/z 288) to ß-caryophyllene was used to calculate titer of the oxygenated taxanes. Means of titers were calculated over replicates and error bars represent s.d.

4.6 Acknowledgements

The authors gratefully acknowledge Dr. Ryan Philippe for careful reading of the manuscript and the gift of *E. coli* Tax1 and plasmids p5Trc and p10Trc from Manus Bio. The pOSIP plasmid kit used for clonetegration was a gift from Drew Endy and Keith

Shearwin (Addgene kit # 1000000035). This work was supported by an NSF CAREER award (1452441 to J. B. L.), an NSF CBET award (1803747 to J. B. L., K. T. and D. T.-E.), an NSF Graduate Research Fellowship (DGE-1144153 to C. J. G.), and an NIH Biotechnology Training Grant (T32-GM008449-23 to B. W. B.).

4.7 Author Contributions

C.J.G., B.W.B., M.K.K., D.T.E, K.E.T., and J.B.L. designed the study. C.J.G., J.T.L., and J.A. performed experiments. All authors contributed to data analysis and preparation of the manuscript.

4.8 Competing Financial Interests

A utility patent application has been filed for some of the developments contained in this article (US 5369-00519).

4.9 References

Aertsen, A., & Michiels, C. W. (2004). Stress and how bacteria cope with death and survival. *Critical Reviews in Microbiology*. https://doi.org/10.1080/10408410490884757

Ajikumar, P. K., Xiao, W. H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., … Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science*. https://doi.org/10.1126/science.1191652

Belliveau, N. M., Barnes, S. L., Ireland, W. T., Jones, D. L., Sweredoski, M. J.,

Moradian, A., … Phillips, R. (2018). Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1722055115

Biggs, B. W., De Paepe, B., Santos, C. N. S., De Mey, M., & Kumaran Ajikumar, P. (2014). Multivariate modular metabolic engineering for pathway and strain optimization. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2014.05.005

Biggs, B. W., Lim, C. G., Sagliani, K., Shankar, S., Stephanopoulos, G., De Mey, M., & Ajikumar, P. K. (2016). Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1515826113

Boyarskiy, S., Davis López, S., Kong, N., & Tullman-Ercek, D. (2016). Transcriptional feedback regulation of efflux protein expression for increased tolerance to and production of n-butanol. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2015.11.005

Brockman, I. M., & Prather, K. L. J. (2015). Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400422

Carothers, J. M., Goler, J. A., Juminaga, D., & Keasling, J. D. (2011). Model-driven engineering of RNA devices to quantitatively program gene expression. *Science*. https://doi.org/10.1126/science.1212209

Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., … Ellis,

T. (2018). Burden-driven feedback control of gene expression. *Nature Methods*. https://doi.org/10.1038/nmeth.4635

Chang, M. C. Y., Eachus, R. A., Trieu, W., Ro, D. K., & Keasling, J. D. (2007). Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio875

Chappell, J., Takahashi, M. K., & Lucks, J. B. (2015). Creating small transcription activating RNAs. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1737

Chappell, J., Westbrook, A., Verosloff, M., & Lucks, J. B. (2017). Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. *Nature Communications*. https://doi.org/10.1038/s41467-017-01082-6

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., … Keasling, J. D. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2689

Engebrecht, J., & Silverman, M. (1984). Identification of genes and gene products necessary for bacterial bioluminescence. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.81.13.4154

Farmer, W. R., & Liao, J. C. (2000). Improving lycopene production in Escherichia coli by engineering metabolic control. *Nature Biotechnology*. https://doi.org/10.1038/75398

Glasscock, C. J., Yates, L. E., Jaroentomeechai, T., Wilson, J. D., Merritt, J. H., Lucks, J. B., & DeLisa, M. P. (2018). A flow cytometric approach to engineering

Escherichia coli for improved eukaryotic protein glycosylation. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2018.04.014

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: De-novo-designed regulators of gene expression. *Cell*. https://doi.org/10.1016/j.cell.2014.10.002

Gupta, A., Reizman, I. M. B., Reisch, C. R., & Prather, K. L. J. (2017). Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3796

Isabella, V. M., Ha, B. N., Castillo, M. J., Lubkowicz, D. J., Rowe, S. E., Millet, Y. A., … Falb, D. (2018). Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4222

Jones, J. A., Vernacchio, V. R., Lachance, D. M., Lebovich, M., Fu, L., Shirke, A. N., … Koffas, M. A. G. (2015). EPathOptimize: A combinatorial approach for transcriptional balancing of metabolic pathways. *Scientific Reports*. https://doi.org/10.1038/srep11301

Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science*. https://doi.org/10.1126/science.1193990

Kim, E. M., Woo, H. M., Tian, T., Yilmaz, S., Javidpour, P., Keasling, J. D., & Lee, T. S. (2017). Autonomous control of metabolic state by a quorum sensing (QS)-mediated regulator for bisabolene production in engineered E. coli. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2017.11.004

Leonard, E., & Koffas, M. A. G. (2007). Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by Escherichia coli. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01411-07

Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.6.1203

Malinowski, J. J. (2001). Two-phase partitioning bioreactors in fermentation technology. *Biotechnology Advances*. https://doi.org/10.1016/S0734-9750(01)00080-5

Minogue, T. D., Wehland-Von Trebra, M., Bernhard, F., & Von Bodman, S. B. (2002). The autoregulatory role of EsaR, a quorum-sensing regulator in Pantoea stewartii ssp. stewartii: Evidence for a repressor function. *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.2002.02987.x

Moser, F., Borujeni, A. E., Ghodasara, A. N., Cameron, E., Park, Y., & Voigt, C. A. (2018). Dynamic control of endogenous metabolism with combinatorial logic circuits. *Molecular Systems Biology*, *14*(e8605). https://doi.org/10.15252/msb.20188605

Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*. https://doi.org/10.1016/j.cell.2016.02.004

Papenfort, K., & Bassler, B. L. (2016). Quorum sensing signal-response systems in Gram-negative bacteria. *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro.2016.89

Smanski, M. J., Bhatia, S., Zhao, D., Park, Y. J., Woodruff, L. B. A., Giannoukos, G., …

Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3063

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology*. https://doi.org/10.1021/sb400021j

Sun, J., Jeffryes, J. G., Henry, C. S., Bruner, S. D., & Hanson, A. D. (2017). Metabolite damage and repair in metabolic engineering design. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2017.10.006

Tan, C., Marguet, P., & You, L. (2009). Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.218

Tan, S. Z., & Prather, K. L. (2017). Dynamic pathway regulation: recent advances and methods of construction. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2017.10.004

Tsao, C. Y., Hooshangi, S., Wu, H. C., Valdes, J. J., & Bentley, W. E. (2010). Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of E. coli. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2010.01.002

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., … DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.921

Van Dien, S. (2013). From the first drop to the first truckload: Commercialization of microbial processes for renewable chemicals. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2013.03.002

Venayak, N., Anesiadis, N., Cluett, W. R., & Mahadevan, R. (2015). Engineering metabolism through dynamic control. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2014.12.022

Watstein, D. M., McNerney, M. P., & Styczynski, M. P. (2015). Precise metabolic engineering of carotenoid biosynthesis in Escherichia coli towards a low-cost biosensor. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2015.06.007

Weber, W., & Fussenegger, M. (2007). Inducible product gene expression technology tailored to bioprocess engineering. *Current Opinion in Biotechnology*. https://doi.org/10.1016/j.copbio.2007.09.002

Xu, P., Li, L., Zhang, F., Stephanopoulos, G., & Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1406401111

Zhang, F., Carothers, J. M., & Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2149

CHAPTER 5 - *Engineering riboregulated switchable stabilized promoters for constant inducible gene expression at any copy number in bacteria*

5.1 Abstract

Inspired by the ability of rSFPs (Chapter 4) to controllably gate transcription from stress-response promoters, Chapter 5 describes efforts to extend the rSFP concept towards other engineered promoter systems, specifically a recently published system, called a stabilized promoter, that enables constant gene expression regardless of copy number (Segall-Shapiro, Sontag, & Voigt, 2018). Based on concepts from control theory, this system uses an incoherent feed forward loop to buffer transcription from changes in copy number, such as moving a gene from plasmid-based expression to the genome. This is particularly valuable when it is convenient to prototype and tune metabolic pathway expression on a plasmid and later integrate into the genome without altering optimal expression. While a powerful tool, the architecture of stabilized promoters prohibits their control with inducible promoter systems commonly used in metabolic engineering. Here, the concept of gating promoter output with small-transcription activating RNAs (STARs) was applied to create novel riboregulated switchable stabilized promoters (rSSPs).

5.2 Introduction

Metabolic engineering projects require balancing expression of enzymes to optimize flux to the desired product and reduce burden and toxicity to the host (Pfleger, Pitera, Smolke, & Keasling, 2006). Many genetic engineering parts have been developed to help engineers achieve this balance, including libraries of synthetic constitutive promoters (Kelly et al., 2009; Mutalik et al., 2013) for control of transcription level, inducible promoters (Lutz & Bujard, 1997a) that allow tuning of transcription level and timing, and RBS sequences that allow tuning of translation efficiency (Mutalik et al., 2013; Salis, Mirsky, & Voigt, 2009). Recently, more advanced parts for controlling gene expression have been developed, including *E. coli* strains containing 12 highly-optimized and orthogonal inducible promoter systems (Meyer et al., 2019) and highly effective RNA regulators of gene expression, such as small-transcription activating RNAs (STARs) (Chappell, Takahashi, & Lucks, 2015; Chappell, Westbrook, Verosloff, & Lucks, 2017) and toehold switches that regulate translation (Green, Silver, Collins, & Yin, 2014).

Altogether, these genetic parts provide effective means of tuning expression and timing of pathway enzymes in metabolic pathways. However, the nature of these genetic parts makes them sensitive to DNA copy number, especially in plasmid-based expression, which can change over time and between cells (Wong Ng, Chatenay, Robert, & Poirier, 2010), in different host strains (Lopilato, Bortner, & Beckwith, 1986), and in different growth conditions such as medium (Lin-Chao & Bremer, 1986; Wegrzyn, 1999), temperature (Lin-Chao, Chen, & Wong, 1992), and growth rate (Lin-Chao & Bremer, 1986). Furthermore, any changes to the host cell, such as adding new pathway enzymes or accumulating mutations during a bioprocess, can influence plasmid copy number

(Cheah, Weigand, & Stark, 1987; Corchero & Villaverde, 1998; Stueber & Bujard, 1982). Even integrating metabolic pathways into the genome does not fully insulate gene expression from the effects of copy number, as fast dividing bacteria have large enrichment of genes closer to the origin of replication that can differ up to eightfold (Block, Hussein, Liang, & Lim, 2012; Chandler & Pritchard, 1975). The dependence of gene expression on copy number is hugely frustrating for metabolic engineers when genetic systems are modified, moved from one plasmid to another, or integrated into the genome, because any prior optimization of expression level is disrupted and must be repeated in each new condition (Cardinale & Arkin, 2012; Kittleson, Wu, & Anderson, 2012).

A longstanding goal of synthetic biology has been to develop genetic parts that are predictable, reliable, and robust to perturbations such as changing or fluctuating copy number. Recently, Voigt and coworkers attempted to help solve the problem of DNA copy number dependence using the concept of an incoherent feedforward loop (iFFL) (Segall-Shapiro et al., 2018), which is an intuitively simple approach to achieve perfect adaption (Bleris et al., 2011; Ma, Trusina, El-Samad, Lim, & Tang, 2009). A feedforward loop is 'incoherent' when an input signal (in this case DNA copy number) is split, and it both positively and negatively controls the output (Mangan & Alon, 2003). Segall-Shapiro *et al*. used the iFFL concept to build a "stabilized promoter" by making the promoter responsive to a protein repressor (**Fig. 5-1A,B**), which itself is controlled by a constitutive promoter and placed onto the same plasmid as the stabilized promoter. In this way, increased copy number results in increased expression of the repressor, which interacts with the stabilized promoter to cancel out the change in expression of the gene(s) of interest (GOI) caused by copy number. This system was predicted to be independent of copy number

as long as the repression of the stabilized promoter is perfectly non-cooperative. Segall-Shapiro used transcription-activator-like-effector (TALE) proteins as their choice of repressor because they can be designed to tightly bind an operator sequence placed in the promoter to achieve at least 100-fold repression in *E. coli* with a non-cooperative transfer function (Copeland, Politz, Johnson, Markley, & Pfleger, 2016; Rogers et al., 2015).

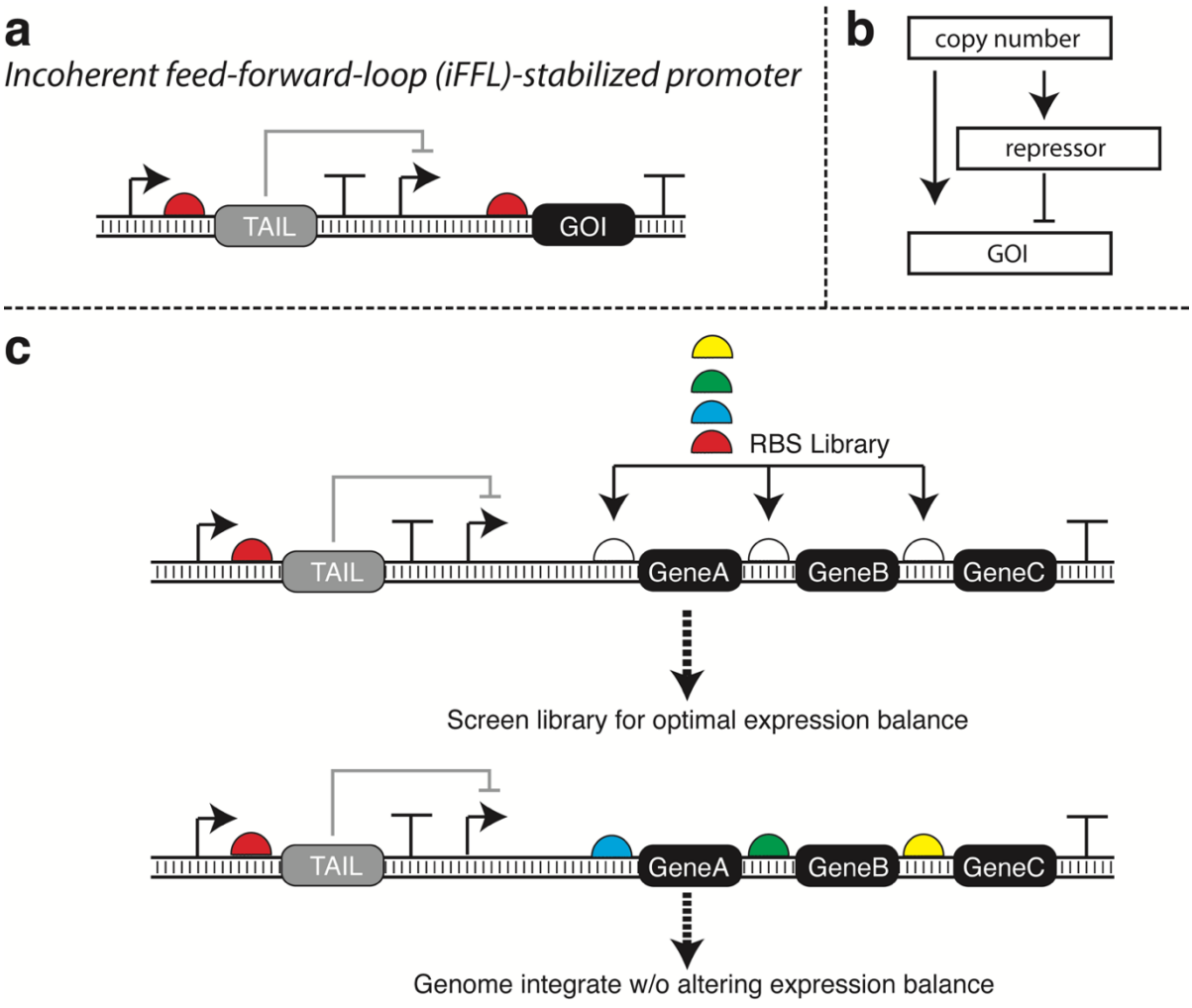


Figure 5-1. Stabilized promoters enable constant gene expression at any copy number in bacteria.

(**A**) Design of an iFFL stabilized promoter using a TALE protein as a repressor. (**B**) Schematic of an iFFL using copy number as an input and a repressor as an intermediate repressor. (**C**) Demonstration of the utility of stabilized promters for integrating a metabolic pathway into the genome without disrupting the carefully engineered balance of enzyme expression.

Functional tests of the engineered stabilized promoters strikingly resulted in nearly perfect adaption to changes in copy number across several orders of magnitude, from genome-based expression to plasmids with ~150 copies (Segall-Shapiro et al., 2018). Stabilized promoters were even shown to be effective against transient changes in plasmid copy numbers, media conditions, and growth rates. To demonstrate the utility of the stabilized promoter system, Segall-Shapiro *et al*., showed that a balanced deoxychromoviridans pathway could be transferred from a high-copy plasmid into the genome without disrupting the engineered balance of pathway expression (**Fig. 5-1C**). This demonstration was compelling because it is much easier to balance a pathway encoded on a plasmid but, for reasons of genetic stability, it is often desired to have the final pathway expressed from the genome in a biomanufacturing process. Thus, the ability to balance pathway expression on a plasmid and later integrate into the genome without disrupting the carefully engineered balance is highly desirable.

Much like stress-response promoters in Chapter 4, stabilized promoters lack the ability to control gene expression timing, which is critical for creating a separation of growth and production phase in biomanufacturing processes (Brockman & Prather, 2015). We applied the concept of gating promoter output with small-transcription

activating RNAs (STARs) to create novel riboregulated switchable stabilized promoters (rSSPs) that enable timing control of stabilized promoters.

5.3 Results

5.3.1. STARs introduce inducible functionality into stabilized promoters

Our initial rSSP design utilized a previously developed STAR (Chappell et al., 2017) under the well-characterized inducible system TetR/$P_{L,TetO1}$ (Lutz & Bujard, 1997b) interfaced with an iFFL-stabilized promoter (Segall-Shapiro et al., 2018). To construct and characterize these rSSPs, a STAR Target sequence was cloned immediately 3' of the stabilized promoter to regulate expression of an sfGFP reporter, and its cognate STAR was cloned in a second $P_{L,TetO1}$ plasmid (**Fig. 5-2A**). Plasmids were transformed into *E. coli* and fluorescence was measured with and without the presence of the $P_{L,TetO1}$ inducer anhydrotetracycline (aTc) at saturating levels (100 ng/mL). We found that induction of $P_{L,TetO1}$-STAR resulted in significant activation of the stabilized promoter (~30x-fold) and a STAR-regulated constitutive promoter lacking the TAIL repressor protein (~90x-fold) (**Fig. 5-2B**).

To demonstrate the effect of plasmid copy number on both a STAR-regulated constitutive promoter and the rSSP, we cloned mutants of the commonly used pSC101 plasmid backbone that exhibit a range of different copy numbers (Peterson & Phillips, 2008), between ~2 to ~30. We observed that the STAR-regulated constitutive promoter system increased sfGFP expression as the pSC101 plasmid backbone was varied to increase copy number, as expected. Strikingly, there was no significant change in sfGFP expression when rSSPs were expressed from the various pSC101 mutant backbones.

Overall, these results show that rSSPs introduce inducible functionality to stabilized promoters while maintaining insulation from plasmid copy number. One caveat of the system is that the inducible STAR must be expressed from a second plasmid that is not insulated from changes in copy number, somewhat limiting the utility of the rSSP system.

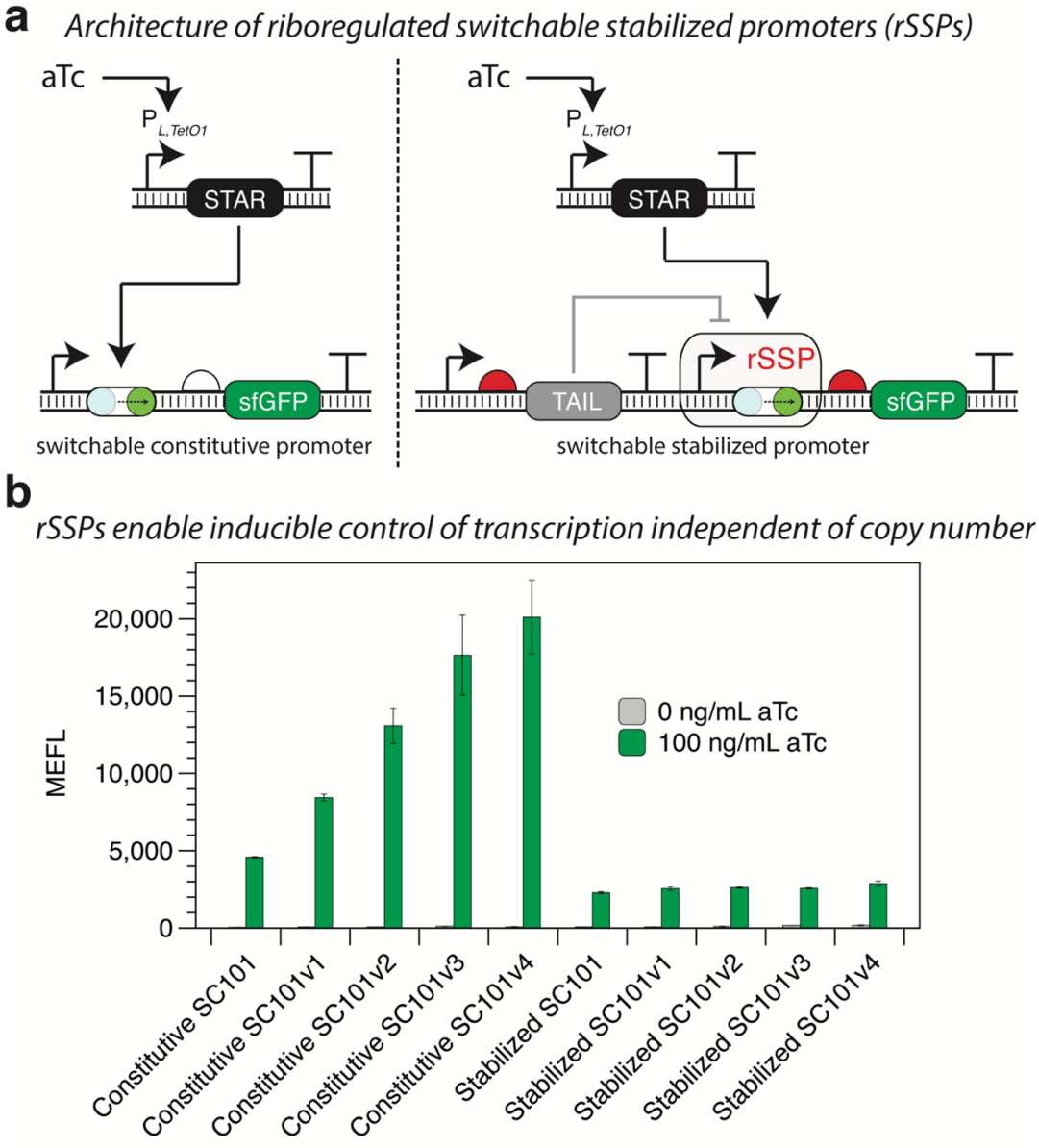


Figure 5-2. Riboregulated switchable stabilized promoters (rSSPs) introduce inducible functionality into stabilized promoters.

(**A**) Architecture of a STAR-regulated constitutive promoter (left) and rSSP (right). STAR expression is mediated by the commonly used inducible system $P_{L,TetO1}$ in both cases. (**B**) Induction of STAR expression with aTc significantly activates both the STAR-regulated constitutive promoter and rSSP. Changes in plasmid copy number through mutations in the pSC101 plasmid backbone lead to increased gene expression with the STAR-regulated constitutive promoter, but not with the rSSP, as designed.

5.4 Discussion

rSSPs enable timing control and expression level titration of stabilized promoters that may have utility in metabolic pathway engineering. In addition, it should be possible to flexibly control rSSPs with a number of inducible systems, and, with further engineering, it should be possible to create orthogonal rSSPs using STARs with minimal cross-talk to non-cognate Target sequences (Chappell et al., 2017). This would enable independent control of multiple rSSPs in a single strain with orthogonal inducible promoter systems.

Recently, Meyer *et al.* engineered strains of *E. coli,* called Marionette strains, containing 12 highly optimized and orthogonal inducible promoter systems (Meyer et al., 2019). These strains were created using a directed evolution scheme to increase the dynamic range of inducible promoters and their cognate transcription factors in response to ligand induction while simultaneously selecting against cross-talk between non-cognate transcription factor-promoter pairs. The final Marionette platform consisted of all 12 transcription factors integrated into the genome of three commonly used *E. coli* strains: MG1655 (Marionette-Wild), BL21 (Marionette-Pro), and DH10B (Marrionette-Clo). In

order to increase the flexibility of the rSSP system, we expect that Marionette promtoers could be used to control STAR expression and activate rSSP expression.

To demonstrate the potential utility of rSSPs in metabolic pathway engineering, we provide two examples for balancing a metabolic pathway for deoxychromoviridans production. First, we show how rSSPs could be used to integrate timing control into the pathway that can be balanced with a plasmid-based library before integration into the genome (**Fig. 5-4**). Implementation of rSSPs in this manner would enable a separation of growth phase and production phase in biomanufacturing processes, which is frequently desirable to avoid accumulation of genetic mutations during exponential growth of seed trains (Wehrs et al., 2019).

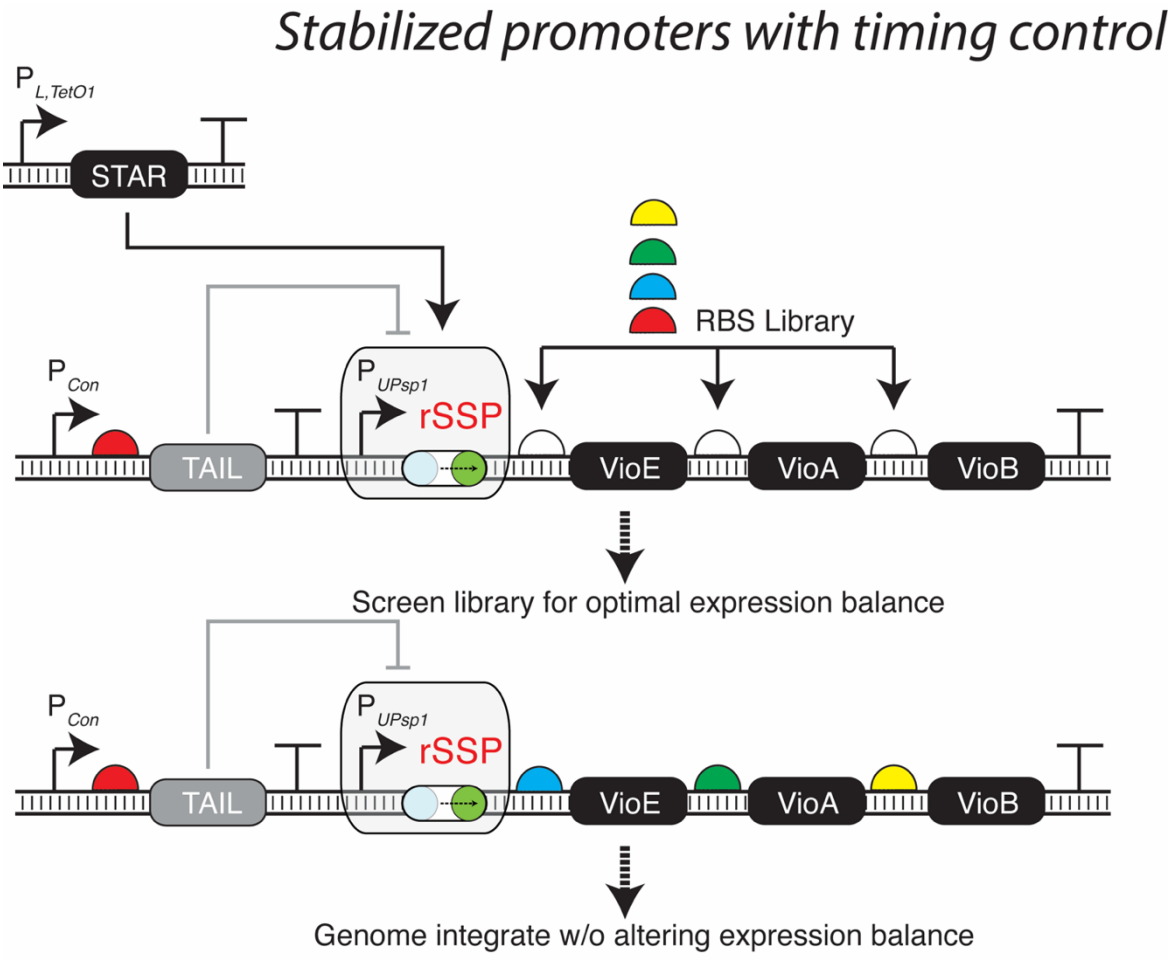


Figure 5-3. rSSPs enable timing control of balanced pathways that can be integrated into the genome without disruption of expression levels.

An example pathway for deoxychromoviridans production is shown. Pathway expression can be balanced using a library of RBS sequences controlling each pathway enzyme. Upon determination of optimal RBS sequences for each enzyme, the entire operon can be integrated into the genome without disruption of expression level. These genome-integrated constructs can be controlled with the inducible promoter of choice.

Next, we provide an example of how rSSPs could be used to balance pathway expression of the deoxychromoviridans pathway without costly and laborious library construction (**Fig. 5-5**). This would require rSSPs with orthogonal STARs that can independently control individual pathway enzymes with Marionette promoters. In this way, combinatorial titration of each rSSP inducer could be used to determine the optimal expression level of each enzyme. The rSSP Target sequence could then be replaced with an RBS sequence that matches the determined optimal induction level of each pathway enzyme to create a new construct that does not require inducers for pathway expression. This new could construct could then be integrated into the genome without disrupting its carefully engineered balance.

Overall, we shown that stabilized promoters can be controlled with STARs to create stabilized promoters to create rSSPs with expression independent of copy number. Further development of the rSSP system could enable independent control of multiple rSSPs in a single strain using Marionette promoters and STARs that do not exhibit cross-talk with non-cognate Target sequences. In the future, we expect that rSSPs could have

utility in metabolic pathway engineering to implement a separation of growth and production phase and to balance pathway expression without costly and laborious library construction, all while using stabilized promoters that are independent of copy number.

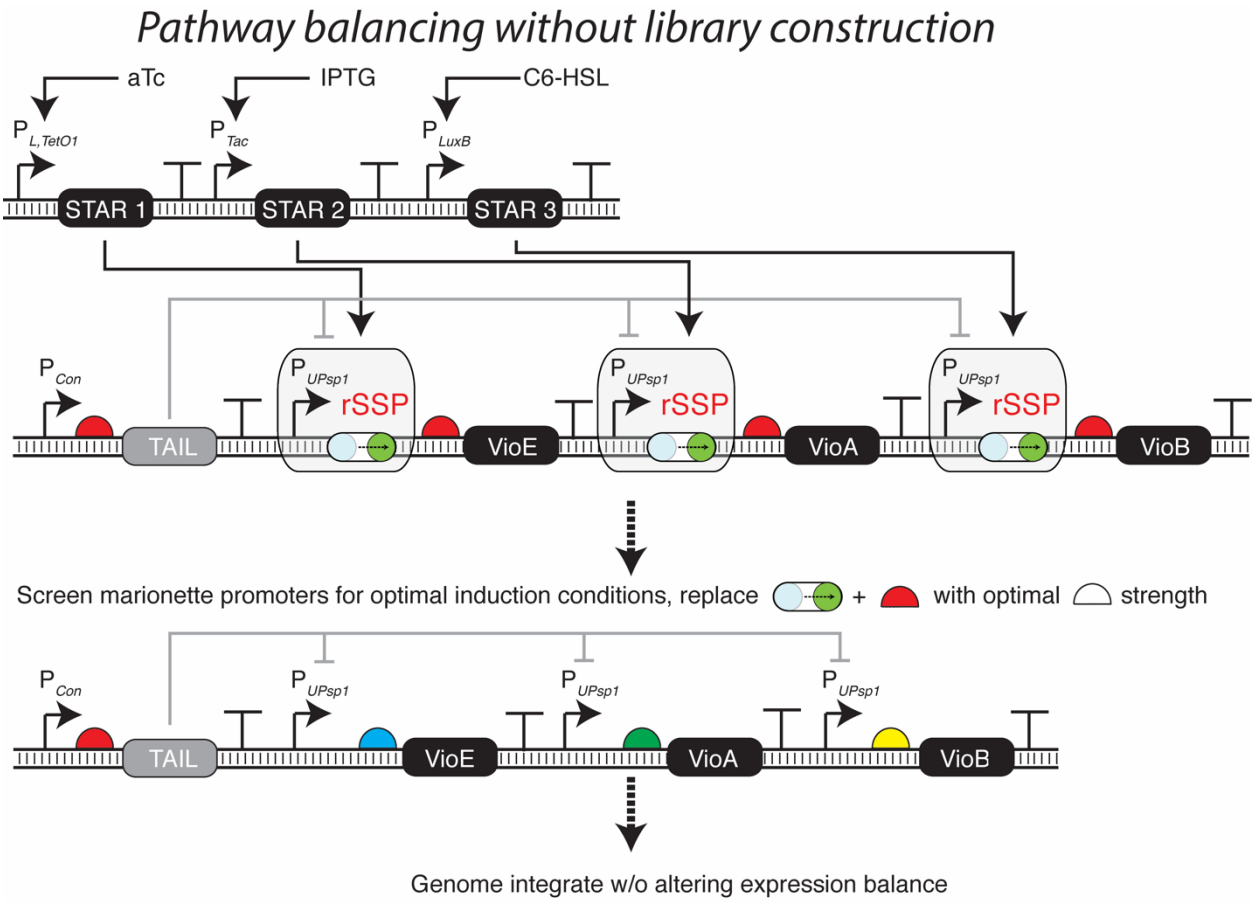


Figure 5-4. rSSPs enable balancing of pathway expression without library construction.

An example pathway for deoxychromoviridans production is shown. Each pathway enzyme can be individually controlled by an rSSP regulated with a unique orthogonal STAR that does not have cross-talk with other STARs in the system. Combinatorial titration of orthogonal STAR expression with Marionette promoters then allows balancing of pathway enzyme expression levels without costly and laborious library construction.

The STAR Target sequence regulating each enzyme can then be replaced with a new RBS sequence that matches the determined optimal induction level to create a new constitutively expressed construct. This final construct can be integrated into the genome without altering expression balance.

5.5 Material and Methods

5.5.1 Plasmid and strain construction

All plasmids were constructed using Gibson assembly or inverse PCR. The chloramphenicol resistance marker in *E. coli* Marionette-Wild was removed using the pE-FLP plasmid from a previously published protocol (St-Pierre et al., 2013).

5.5.2 Flow cytometry data collection and analysis for sfGFP fluorescence analysis

All flow cytometry experiments were performed in *E. coli* strain TG1 (*F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB)*) or *E. coli* Marionette-Wild (Meyer et al., 2019) with the chloramphenicol resistance marker removed. Cells were transformed with the relevant combination of reporter and STAR plasmids or target plasmid with pJBL002 (a blank plasmid). An autofluorescence control was included by transforming blank target and trigger plasmids (pJBL001 and pJBL002, respectively) which were both taken from Lucks *et al*.(Lucks, Qi, Mutalik, Wang, & Arkin, 2011). Plasmid combinations were transformed into chemically competent *E. coli* cells, plated on Difco LB+Agar plates containing 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol, and grown overnight at 37 °C. Following overnight incubation, plates

were left at room temperature for approximately 9 h. Individual colonies were grown overnight in LB, then diluted 1:50 into M9 minimal media. After 6 h, cells were diluted 1:100 in 1x Phosphate Buffered Saline (PBS) containing 2 mg/mL kanamycin. A BD Accuri C6 Plus flow cytometer fitted with a high-throughput sampler was then used to measure sfGFP fluorescence. Measurements were taken for at least 3 biological replicates.

Flow cytometry data analysis was performed using FlowJo (v10.4.1). Cells were gated by FSC-A and SSC-A, and the same gate was used for all samples. Samples exhibiting non-unimodal fluorescence distributions were excluded from further analysis, and the geometric mean fluorescence was calculated for each sample. All fluorescence measurements were converted to Molecules of Equivalent Fluorescein (MEFL) using CS&T RUO Beads (BD cat#661414), which were run on each day of data collection. The average fluorescence (MEFL) over replicates of cells expressing empty plasmids (pJBL001 and pJBL002)[37] was then subtracted from each measured fluorescence value.

5.6 Acknowledgements

Stabilized promoter plasmids and the *E. coli* Marionette-Wild strain were a gift from Christopher Voigt.

5.7 References

Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., & Benenson, Y. (2011). Synthetic

incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*. https://doi.org/10.1038/msb.2011.49

Block, D. H. S., Hussein, R., Liang, L. W., & Lim, H. N. (2012). Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks694

Brockman, I. M., & Prather, K. L. J. (2015). Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400422

Cardinale, S., & Arkin, A. P. (2012). Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal*. https://doi.org/10.1002/biot.201200085

Chandler, M. G., & Pritchard, R. H. (1975). The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. *MGG Molecular & General Genetics*. https://doi.org/10.1007/BF02428117

Chappell, J., Takahashi, M. K., & Lucks, J. B. (2015). Creating small transcription activating RNAs. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1737

Chappell, J., Westbrook, A., Verosloff, M., & Lucks, J. B. (2017). Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. *Nature Communications*. https://doi.org/10.1038/s41467-017-01082-6

Cheah, U. E., Weigand, W. A., & Stark, B. C. (1987). Effects of recombinant plasmid size on cellular processes in Escherichia coli. *Plasmid*. https://doi.org/10.1016/0147-619X(87)90040-0

Copeland, M. F., Politz, M. C., Johnson, C. B., Markley, A. L., & Pfleger, B. F. (2016). A

transcription activator-like effector (TALE) induction system mediated by proteolysis. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.2021

Corchero, J. L., & Villaverde, A. (1998). Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnology and Bioengineering*. https://doi.org/10.1002/(SICI)1097-0290(19980620)58:6<625::AID-BIT8>3.0.CO;2-K

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: De-novo-designed regulators of gene expression. *Cell*. https://doi.org/10.1016/j.cell.2014.10.002

Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., … Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*. https://doi.org/10.1186/1754-1611-3-4

Kittleson, J. T., Wu, G. C., & Anderson, J. C. (2012). Successes and failures in modular genetic engineering. *Current Opinion in Chemical Biology*. https://doi.org/10.1016/j.cbpa.2012.06.009

Lin-Chao, S., & Bremer, H. (1986). Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli. *MGG Molecular & General Genetics*. https://doi.org/10.1007/BF00330395

Lin-Chao, S., Chen, W. -T, & Wong, T. -T. (1992). High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Molecular Microbiology*. https://doi.org/10.1111/j.1365-2958.1992.tb02206.x

Lopilato, J., Bortner, S., & Beckwith, J. (1986). Mutations in a new chromosomal gene of Escherichia coli K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives. *MGG Molecular & General Genetics*. https://doi.org/10.1007/BF00430440

Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., & Arkin, A. P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1015741108

Lutz, R., & Bujard, H. (1997a). Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.6.1203

Lutz, R., & Bujard, H. (1997b). Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.6.1203

Ma, W., Trusina, A., El-Samad, H., Lim, W. A., & Tang, C. (2009). Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell*. https://doi.org/10.1016/j.cell.2009.06.013

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2133841100

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2019). Escherichia coli “Marionette” strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-018-0168-3

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.

A., … Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*. https://doi.org/10.1038/nmeth.2404

Peterson, J., & Phillips, G. J. (2008). New pSC101-derivative cloning vectors with elevated copy numbers. *Plasmid*. https://doi.org/10.1016/j.plasmid.2008.01.004

Pfleger, B. F., Pitera, D. J., Smolke, C. D., & Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*. https://doi.org/10.1038/nbt1226

Rogers, J. M., Barrera, L. A., Reyon, D., Sander, J. D., Kellis, M., Keith Joung, J., & Bulyk, M. L. (2015). Context influences on TALE-DNA binding revealed by quantitative profiling. *Nature Communications*. https://doi.org/10.1038/ncomms8440

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1568

Segall-Shapiro, T. H., Sontag, E. D., & Voigt, C. A. (2018). Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4111

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology*. https://doi.org/10.1021/sb400021j

Stueber, D., & Bujard, H. (1982). Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *The EMBO Journal*.

Wegrzyn, G. (1999). Replication of plasmids during bacterial response to amino acid starvation. *Plasmid*. https://doi.org/10.1006/plas.1998.1377

Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T. R., & Mukhopadhyay, A. (2019). Engineering Robust Production Microbes for Large-Scale Cultivation. *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2019.01.006

Wong Ng, J., Chatenay, D., Robert, J., & Poirier, M. G. (2010). Plasmid copy number noise in monoclonal populations of bacteria. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. https://doi.org/10.1103/PhysRevE.81.011909

CHAPTER 6 – *Conclusions*

In this work, we optimized a pathway for eukaryotic protein glycosylation in *E. coli* using classical tools from metabolic engineering and a flow cytometry-based screen for glycan biosynthesis. This work resulted in hard-earned and significant improvements to glycan biosynthesis and protein glycosylation efficiency but was not without challenges. Along the way, we elucidated contemporary obstacles in the engineering of cells for biomanufacturing and sought to develop new tools for strain and pathway engineering, inspired by recent progress in synthetic biology and RNA gene expression regulators.

In Chapter 3 we made progress towards an approach for high-throughput strain engineering with RNA gene expression regulators and provide a blueprint for the application of this approach in combinatorial knockdown screening of endogenous gene expression. In Chapter 4, we developed an entirely novel approach for controlling the output of stress-response promoters with small-transcription activating RNAs, called riboregulated switchable feedback promoters (rSFPs). Stress-response promoters have been used effectively to implement dynamic pathway regulation in response to cellular stress, however, thus far there has been no simple mechanism for controlling their output. We applied rSFPs to improve a pathway for production of a Taxol precursor >2x-fold, clearly demonstrating their utility for regulating metabolic pathway expression. Finally, in a system called riboregulated switchable stabilized promoters (rSSPs), chapter 5 shows how we used the same concept to provide timing control over the output of stabilized promoters that buffer gene expression from changes in DNA copy. It is expected that future research will further develop and enhance these technologies to aid in the development of sustainable and cost-effective biomanufacturing platforms.

CHAPTER 7 - *Appendix A*

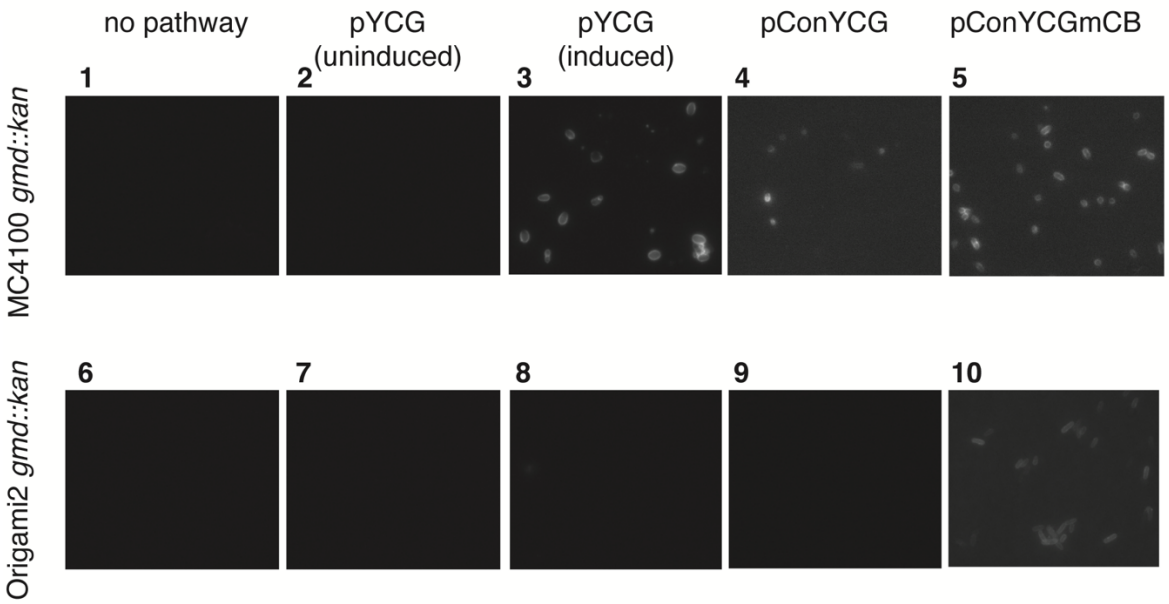
Supplementary Information: *A flow cytometric approach to engineering* Escherichia coli *for improved eukaryotic protein glycosylation.*

Supplementary Table A.1. Important DNA sequences used in this study

Name	Sequence
pConYCG (**BBa_J23109**-ALG13-ALG14-ALG1-ALG2-**rrnB**)	**TTTACAGCTAGCTCAGTCCTAGGGACTGTGCTAGC**GAATTCGAGCTCCCGGGAGGAGGAACGATGG GTATTATTGAAGAAAAGGCTCTTTTTGTTACGTGTGGGGCAACGGTGCCATTTCCAAAGCTCGTCTCAT GTGTGCTAAGCGACGAATTCTGCCAAGAATTGATTCAATATGGATTCGTACGTCTAATCATTCAGTTTG GGAGAAACTACAGTTCTGAATTTGAGCATTTAGTGCAAGAACGCGGGGGCCAAAGAGAAAGCCAAAAA ATTCCAATTGACCAGTTTGGCTGTGGCGACACCGCAAGACAGTATGTCCTGATGAACGGGAAATTAAA GGTGATCGGGTTTGACTTTTCGACCAAGATGCAAAGTATTATACGTGATTATTCAGATTTGGTCATATC ACACGCTGGAACGGGCTCTATACTAGATTCTCTACGGTTGAATAAACCGTTGATAGTTTGCGTAAACGA TTCTTTGATGGATAACCACCAGCAGCAGATAGCAGACAAGTTTGTAGAGTTGGGCTACGTATGGTCTT GTGCACCCACTGAAACAGGTTTGATAGCTGGTTTACGTGCATCTCAAACAGAGAAACTCAAACCATTCC CAGTTTCTCATAACCCGTCATTTGAGCGATTGCTAGTTGAAACTATATACAGCTAAGGGCCCCTACCAG GAGGAACGATGAAAACGGCCTACTTGGCGTCATTGGTGCTCATCGTATCGACAGCATATGTTATTAGG TTGATAGCGATTCTGCCTTTTTTCCACACTCAAGCAGGTACAGAAAAGGATACGAAAGATGGAGTTAAC CTACTGAAAATACGAAAATCGTCAAAGAAACCGCTCAAGATTTTTGTATTCTTAGGATCGGGAGGTCAT ACTGGTGAAATGATCCGTCTTCTAGAAAATTACCAGGATCTTTTACTGGGTAAGTCGATTGTGTACTTG GGTTATTCTGATGAGGCTTCCAGGCAAAGATTCGCCCACTTTATAAAAAAATTTGGTCATTGCAAAGTA AAATACTATGAATTCATGAAAGCTAGGGAAGTTAAAGCGACTCTCCTACAAAGTGTAAAGACCATCATT GGAACGTTGGTACAATCTTTTGTGCACGTGGTTAGAATCAGATTTGCTATGTGTGGTTCCCCTCATCTG TTTTTATTGAATGGGCCTGGAACATGCTGTATAATATCCTTTTGGTTGAAAATTATGGAACTTCTTTTGC CCCTGTTGGGTTCCTCCCATATAGTTTATGTAGAATCGCTGGCAAGGATTAATACTCCTAGTCTGACCG GAAAAATATTATATTGGGTAGTGGATGAATTCATTGTCCAGTGGCAAGAATTGAGGGACAATTATTTAC CAAGATCCAAGTGGTTCGGCATCCTTGTTTAAGGTACCCTCGAGAGGAGGAACGATGTTTTTGGAAAT TCCTCGGTGGTTACTTGCCTTAATAATATTATACCTTTCCATACCGTTAGTGGTTTATTATGTTATACCCT ACTTGTTTTATGGCAACAAGTCGACCAAAAAAAGGATCATCATATTTGTGCTGGGTGATGTAGGACACT CTCCAAGGATATGCTATCACGCTATAAGTTTCAGTAAGTTAGGTTGGCAAGTCGAGCTATGCGGTTATG TGGAGGACACTCTACCCAAAATTATTTCCAGTGATCCAAATATCACCGTCCATCATATGTCAAACTTGA AAAGAAAGGGAGGCGGAACATCAGTTATATTTATGGTAAAGAAGGTGCTTTTTCAAGTTTTAAGTATTTT CAAATTACTTTGGGAATTGAGAGGAAGCGATTACATACTAGTTCAAAATCCACCGAGCATACCCATTCT TCCGATTGCTGTGCTATACAAGTTGACCGGTTGTAAACTAATTATTGATTGGCACAATCTAGCATATTC GATATTGCAACTAAAATTTAAAGGAAACTTTTACCATCCTTTAGTGTTGATATCTTACATGGTAGAGATG ATATTCAGCAAATTTGCTGATTATAACTTGACTGTTACTGAAGCAATGAGGAAATATTTAATTCAAAGCT TTCACTTGAATCCAAAGAGATGTGCTGTTCTCTACGACCGCCCGGCTTCCCAATTTCAACCTTTGGCAG GTGACATTTCTCGTCAAAAAGCCCTAACTACCAAAGCCTTTATAAAGAATTATATTCGCGATGATTTTGA TACAGAAAAAGGCGATAAAATTATTGTGACTTCAACATCATTCACCCCTGATGAAGATATTGGTATTTTA TTAGGTGCCCTAAAGATTTACGAAAACTCTTATGTCAAATTTGATTCAAGTTTGCCTAAGATCTTGTGTT TTATAACGGGTAAAGGACCACTAAAGGAGAAATATATGAAGCAAGTAGAAGAATATGACTGGAAGCGC TGTCAAATCGAATTTGTGTGGTTGTCAGCAGAGGATTACCCAAAGTTATTACAATTATGCGATTACGGA GTTTCCCTGCATACTTCAAGTTCAGGGTTGGACCTGCCAATGAAAATTTTAGATATGTTTGGCTCAGGT CTTCCTGTTATTGCAATGAACTATCCAGTGCTTGACGAATTAGTACAACACAATGTAAATGGGTTAAAAT TTGTTGATAGAAGGGAGCTTCATGAATCTCTGATTTTTGCTATGAAAGATGCTGATTTATACCAAAAATT GAAGAAAAATGTAACGCAGGAAGCTGAGAACAGATGGCAATCAAATTGGGAACGAACAATGAGAGATT TGAAGCTAATTCATTGACTCGAGCATGCAGGAGGAACGATGATTGAAAAGGATAAAAGAACGATTGCT TTTATTCATCCAGACCTAGGTATTGGGGGCGCTGAAAGGTTAGTCGTCGATGCAGCATTAGGTCTACA GCAACAAGGACATAGTGTAATCATCTATACTAGTCACTGTGATAAATCACATTGTTTCGAAGAAGTTAAA AACGGCCAATTAAAAGTCGAAGTTTATGGTGATTTTTTACCGACAAACTTTTTGGGTCGTTTTTTTATTG TTTTCGCAACAATTAGACAGCTTTATTTAGTTATTCAATTGATCCTACAGAAAAAAGTGAATGCGTACCA ATTAATTATCATTGATCAACTGTCTACATGTATTCCGCTTCTGCATATCTTTAGTTCTGCCACTTTGATGT TTTATTGTCATTTCCCCGACCAATTATTGGCTCAAAGAGCTGGGCTATTGAAGAAAATATACAGACTAC CATTTGACTTAATAGAACAGTTTTCCGTGAGTGCTGCCGATACTGTTGTGGTAAATTCAAATTTCACTAA GAATACGTTCCACCAAACGTTCAAGTATTTATCCAATGATCCAGACGTCATTTATCCATGCGTGGATTTA TCAACAATCGAAATTGAAGATATTGACAAGAAATTTTTCAAAACAGTGTTTAACGAAGGCGATAGATTTT ACCTAAGTATAAATCGTTTTGAGAAAAAAAAAGGATGTTGCGCTGGCTATAAAGGCTTTTGCGTTATCTG AAGATCAAATCAATGACAACGTTAAGTTAGTTATTTGCGGTGGTTATGACGAGAGGGTTGCAGAAAATG TGGAGTACTTGAAGGAACTACAGTCTCTGGCCGATGAATACGAATTATCCCATACAACCATATACTACC

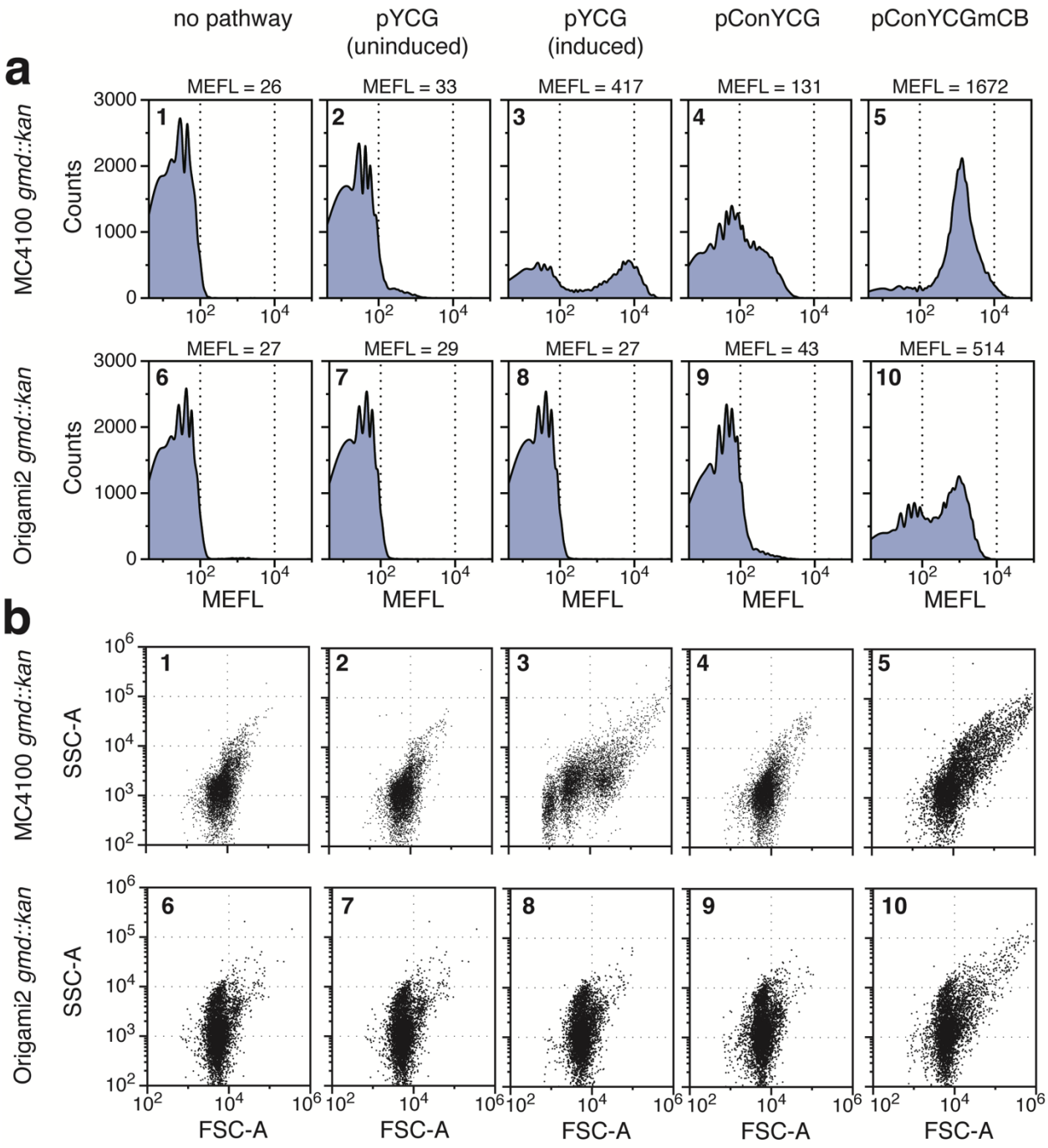
	AAGAAATAAAGCGCGTCTCCGATTTAGAGTCATTCAAAACCAATAATAGTAAAATTATATTTTTAACTTC CATTTCATCATCTCTGAAAGAATTACTGCTCGAAAGAACCGAAATGTTATTGTATACACCAGCATATGAG CACTTTGGTATTGTTCCTTTAGAAGCCATGAAATTAGGTAAGCCTGTACTAGCAGTAAACAATGGAGGT CCTTTGGAGACTATCAAATCTTACGTTGCTGGTGAAAATGAAAGTTCTGCCACTGGGTGGCTAAAACCT GCCGTCCCTATTCAATGGGCTACTGCAATTGATGAAAGCAGAAAGATCTTGCAGAACGGTTCTGTGAA CTTTGAGAGGAATGGCCCGCTAAGAGTCAAGAAATACTTTTCTAGGGAAGCAATGACTCAGTCATTTGA AGAAAACGTCGAGAAAGTCATATGGAAAGAAAAAAAGTATTATCCTTGGGAAATATTCGGTATTTCATT CTCTAATTTTATTTTGCATATGGCATTTATAAAAATTCTACCCAATAATCCATGGCCCTTCCTATTTATGG CCACTTTTATGGTATTATATTTTAAGAACTACTTATGGGGAATTTACTGGGCATTTGTATTCGCTCTCTC CTACCCTTATGAAGAAATATAAGCGGCCGCTTAATTAATCTAGAGTCGACCTGCAGGCATGCAAGCTT GGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTA**AATCAGAACGCAGAAGCGGTC** **TGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAG** **AAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAG** **GCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG** **AACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGG** **AGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCAT**
manC-manB	GCGGCCGCAGGAGGAACGATGGCGCAGTCGAAACTCTATCCAGTTGTGATGGCAGGTGGCTCCGGT AGCCGCTTATGGCCGCTTTCCCGCGTACTTTATCCCAAGCAGTTTTTATGCCTGAAAGGCGATCTCAC CATGCTGCAAACCACCATCTGCCGCCTGAACGGCGTGGAGTGCGAAAGCCCGGTGGTGATTTGCAAT GAGCAGCACCGCTTTATTGTCGCGGAACAGCTGCGTCAACTGAACAAACTTACCGAGAACATTATTCT CGAACCGGCAGGGCGAAACACGGCACCTGCCATTGCGCTGGCGGCGCTGGCGGCAAAACGTCATAG CCCGGAGAGCGACCCGTTAATGCTGGTATTGGCGGCGGATCATGTGATTGCCGATGAAGACGCGTTC CGTGCCGCCGTGCGTAATGCCATGCCATATGCCGAAGCGGGCAAGCTGGTGACCTTCGGCATTGTGC CGGATCTACCAGAAACCGGTTATGGCTATATTCGTCGCGGTGAAGTGTCTGCGGGTGAGCAGGATAT GGTGGCCTTTGAAGTGGCGCAGTTTGTCGAAAAACCGAATCTGGAAACCGCTCAGGCCTATGTGGCA AGCGGCGAATATTACTGGAACAGCGGTATGTTCCTGTTCCGCGCCGGACGCTATCTCGAAGAACTGAA AAAATATCGCCCGGATATCCTCGATGCCTGTGAAAAAGCGATGAGCGCCGTCGATCCGGATCTCAATT TTATTCGCGTGGATGAAGAAGCGTTTCTCGCCTGCCCGGAAGAGTCGGTGGATTACGCGGTCATGGA ACGTACGGCAGATGCTGTTGTGGTGCCGATGGATGCGGGCTGGAGCGATGTTGGCTCCTGGTCTTCA TTATGGGAGATCAGCGCCCACACCGCCGAGGGCAACGTTTGCCACGGCGATGTGATTAATCACAAAA CTGAAAACAGCTATGTGTATGCTGAATCTGGCCTGGTCACCACCGTCGGGGTGAAAGATCTGGTAGTG GTGCAGACCAAAGATGCGGTGCTGATTGCCGACCGTAACGCGGTACAGGATGTGAAAAAAGTGGTCG AGCAGATCAAAGCCGATGGTCGCCATGAGCATCGGGTGCATCGCGAAGTGTATCGTCCGTGGGGCAA ATATGACTCTATCGACGCGGGCGACCGCTACCAGGTGAAACGCATCACCGTGAAACCGGGCGAGGG CTTGTCGGTACAGATGCACCATCACCGCGCGGAACACTGGGTGGTTGTCGCGGGAACGGCAAAAGTC ACCATTGATGGTGATATCAAACTGCTTGGTGAAAACGAGTCCATTTATATTCCGCTGGGGGCGACGCA TTGCCTGGAAAACCCGGGGAAAATTCCGCTCGATTTAATTGAAGTGCGCTCCGGCTCTTATCTCGAAG AGGATGATGTGGTGCGTTTCGCGGATCGCTACGGACGGGTGTAAACGTCGCATCAGGCAATGAATGC GAAACCGCGGTGTAAATAACGACAAAAATAAAATTGGCCGCTTCGGTCAGGGCCAACTATTGCCTGAA AAAGGGTAACGATATGAAAAAATTAACCTGCTTTAAAGCCTATGATATTCGCGGGAAATTAGGCGAAGA ACTGAATGAAGATATCGCCTGGCGCATTGGTCGCGCCTATGGCGAATTTCTCAAACCGAAAACCATTG TGTTAGGCGGTGATGTCCGCCTCACCAGCGAAACCTTAAAACTGGCGCTGGCGAAAGGTTTACAGGA TGCGGGCGTTGACGTGCTGGATATTGGTATGTCCGGCACCGAAGAGATCTATTTCGCCACGTTCCATC TCGGCGTGGATGGCGGCATTGAAGTTACCGCCAGCCATAATCCGATGGATTATAACGGCATGAAGCT GGTTCGCGAGGGGGCTCGCCCGATCAGCGGAGATACCGGACTGCGCGACGTCCAGCGTCTGGCTGA AGCCAACGACTTTCCTCCCGTCGATGAAACCAAACGCGGTCGCTATCAGCAAATCAACCTGCGTGACG CTTACGTTGATCACCTGTTCGGTTATATCAATGTCAAAAACCTCACGCCGCTCAAGCTGGTGATCAACT CCGGGAACGGCGCAGCGGGTCCGGTGGTGGACGCCATTGAAGCCCGCTTTAAAGCCCTCGGCGCGC CCGTGGAATTAATCAAAGTGCACAACACGCCGGACGGCAATTTCCCCAACGGTATTCCTAACCCACTA CTGCCGGAATGCCGCGACGACACCCGCAATGCGGTCATCAAACACGGCGCGGATATGGGCATTGCTT TTGATGGCGATTTTGACCGCTGTTTCCTGTTTGACGAAAAAGGGCAGTTTATTGAGGGCTACTACATTG TCGGCCTGTTGGCAGAAGCATTCCTCGAAAAAAATCCCGGCGCGAAGATCATCCACGATCCACGTCTC TCCTGGAACACCGTTGATGTGGTGACTGCCGCAGGTGGCACGCCGGTAATGTCGAAAACCGGACACG CCTTTATTAAAGAACGTATGCGCAAGGAAGACGCCATCTATGGTGGCGAAATGAGCGCCCACCATTAC TTCCGTGATTTCGCTTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTCGCCGAACTGGTGTGCC TGAAAGATAAAACGCTGGGCGAACTGGTACGCGACCGGATGGCGGCGTTTCCGGCAAGCGGTGAGA TCAACAGCAAACTGGCGCAACCCGTTGAGGCGATTAACCGCGTGGAACAGCATTTTAGCCGTGAGGC GCTGGCGGTGGATCGCACCGATGGCATCAGCATGACCTTTGCCGACTGGCGCTTTAACCTGCGCACC TCCAATACCGAACCGGTGGTGCGCCTGAATGTGGAATCGCGCGGTGATGTGCCGCTGATGGAAGCGC GAACGCGAACTCTGCTGACGTTGCTGAACGAGTAACTTAATTAATCTAGAGTCGACCTGCAGGCATGC AAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTA

*Underlined text denotes start and stop codons for each cloned gene.



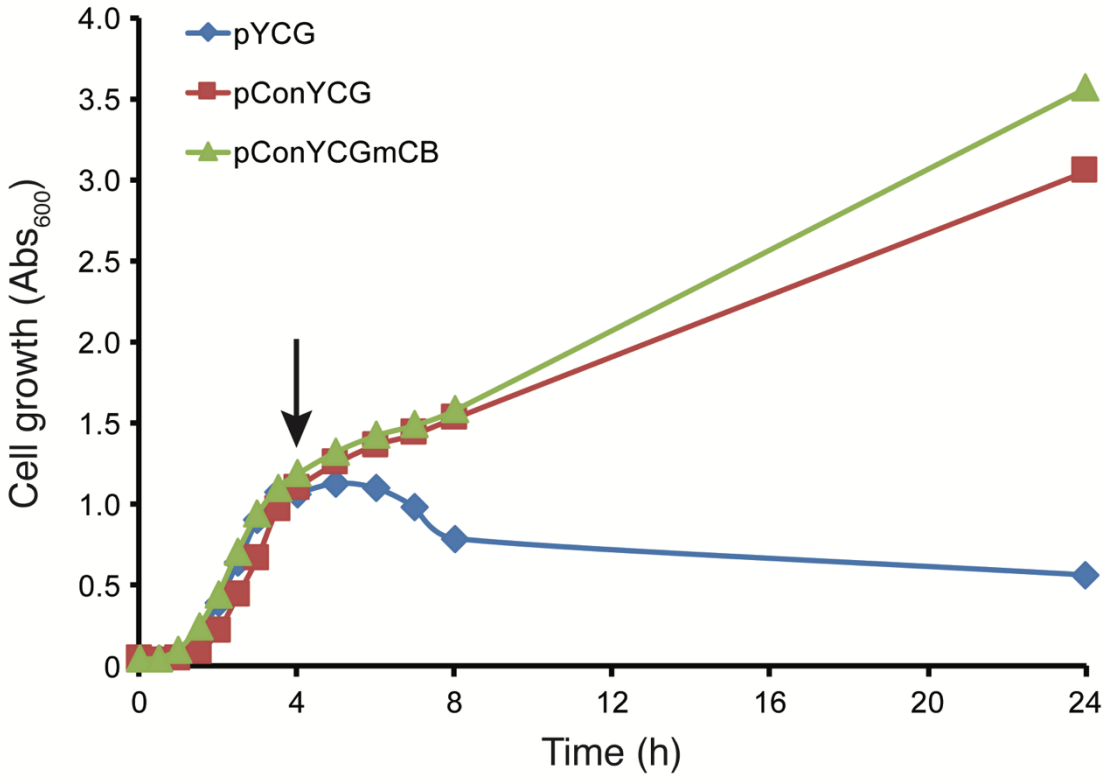
SI Figure A.6-1. Fluorescence microscopy of ConA-labeled bacteria.

MC4100 *gmd::kan* strains (top row; 1- 5) or Origami2 *gmd::kan* strains (bottom row; 6-10) carrying the plasmids indicated were labeled with ConA-AlexaFluor and analyzed by fluorescence microscopy.



SI Figure A.6-2. Flow cytometric analysis of glycoengineered *E. coli* strains. Flow cytometry data including representative (a) MEFL fluorescence histograms and (b) dot plots of forward scatter (FSC-A) versus side scatter (SSC- A) was collected for MC4100 *gmd::kan* strains (1-5) or Origami2 *gmd::kan* strains (6-10)

with different plasmids as indicated. Cells were labeled with ConA-AlexaFluor prior to flow cytometry. Molecules of equivalent fluorescein (MEFL) values are given for each histogram. Generation of histograms was performed using FlowCal software and FSC-A versus SSC-A dot plots were generated using FlowJo 10.4.1 software.



SI Figure A.6-3. Growth analysis of glycoengineered *E. coli* strains.

MC4100 *gmd::kan* cells carrying pYCG, pConYCG, or pConYCGmCB were grown at 37oC until Abs_{600} reached ≈1.5. Culture temperature was reduced to 30oC and cells were induced (pYCG cells only) with 0.2% (w/v) L-arabinose. Measurements of Abs_{600} were made at the time points indicated. Data are the average of biological triplicates and the error was less than 5% in all cases.

CHAPTER 8 - *Appendix B*

Supplementary Information: *Dynamic control of pathway expression with riboregulated switchable feedback promoters.*

Table of Contents:

Supplementary Results

Supplementary Tables

Supplementary Table B1. Plasmids used in this study. apFAB parts were obtained from a previously published library of genetic parts[1]. Abbreviations are as follows: RBS 1 = ribosome binding site variant (see Supplementary Table 3), P_R = tetR promoter[2], $P_{LTet,O1}$ = TetR repressible promoter[3], P_{Lux} (BBa_R0062)* = LuxR inducible promoter[4], mCherry = red fluorescent protein, LuxR (BBa_C0062)* = AHL inducible transcription factor[4], tetR = tet repressor protein[3], TrrnB = rrnB terminator, BBa_B0015* = B0015 terminator, T500 = T500 terminator, dblTerm = dblTerm terminator, PJ2315 = BBa_J23115 promoter from

the iGEM Registry of Standard Biological Parts (parts.igem.org), CmR = chloramphenicol resistance cassette, AmpR = ampicillin resistance cassette, SpcR = spectinomycin resistance cassette, p15A = p15A origin of replication, ColE1 = ColE1 origin of replication and CDF = CDF origin of replication.

Plasmid #	Plasmid architecture	Name	Figure
pJBL002	AmpR – ColE1 origin (Empty vector)	pJBL002	1d-e, 4b, SI Fig. 1, SI Fig. 3
pJBL644	SpcR – pCDF origin (Empty vecor)	pJBL644	1d-e, 4b, SI Fig. 1, SI Fig. 3
pJBL6654	P_R – TetR – dblTerm – PL,TetO1 – STAR 8 – T500 – AmpR – ColE1 origin	$P_{L,TetO1}$-STAR	1d-e, 2d, 3c-f, 4c-d, SI Fig. 1, SI Fig. 2, SI Fig. 3, SI Fig. 4, SI Fig 5
pJBL6655	PLux – STAR 8 – T500 – AmpR – ColE1 origin	P_{Lux}-STAR	4b-c
pJBL6656	PgntK – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PgntK-mCherry	1d, SI Fig. 1
pJBL6657	PompF – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PompF-mCherry	1d-e, 4b, , SI Fig. 1, SI Fig. 3
pJBL6658	PyeeF – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PyeeF-mCherry	1d-e, SI Fig. 1
pJBL6659	PompT – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PompT-mCherry	1d-e, SI Fig. 1
pJBL6660	PmetN – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PmetN-mCherry	1d-e, 4b, SI Fig. 1, SI Fig. 3
pJBL6661	Pb1762 – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	Pb1762-mCherry	1d, SI Fig. 1
pJBL6662	PcarA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PcarA-mCherry	1d, SI Fig. 1
pJBL6663	PfadL – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PfadL-mCherry	1d, SI Fig. 1
pJBL6664	PfecA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PfecA-mCherry	1d, SI Fig. 1
pJBL6665	PuraA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PuraA-mCherry	1d, SI Fig. 1
pJBL6666	PgrxA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PgrxA-mCherry	1d, SI Fig. 1
pJBL6667	PmtgA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PmtgA-mCherry	1d, SI Fig. 1
pJBL6668	PybcU – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PycbU-mCherry	1d, SI Fig. 1
pJBL6669	PycbS – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PycbS-mCherry	1d, SI Fig. 1

pJBL6670	PyhjX – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PyhjX-mCherry	1d, SI Fig. 1
pJBL6671	PatoA – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PatoA-mCherry	1d, SI Fig. 1
pJBL6672	Pb2970 – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	Pb2970-mCherry	1d, SI Fig. 1
pJBL6673	PecpD – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PecpD-mCherry	1d, SI Fig. 1
pJBL6674	PgntK – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – p15a origin	PgntK-P450	2d, SI Fig 2
pJBL6675	PompF – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PompF-P450	2d, 3d, 3f, 4d, SI Fig. 2, SI Fig. 5
pJBL6676	PyeeF – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PyeeF-P450	2d, SI Fig 2
pJBL6677	PompT – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PompT-P450	2d, SI Fig 2
pJBL6678	PmetN – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PmetN-P450	2d, 3c, 3e, 4c, SI Fig. 2, SI Fig. 4
pJBL6679	Pb1762 – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	Pb1762-P450	2d, SI Fig 2
pJBL6680	PcarA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PcarA-P450	2d, SI Fig 2
pJBL6681	PfadL – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PfadL-P450	SI Fig 2
pJBL6682	PfecA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PfecA-P450	SI Fig 2
pJBL6683	PuraA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PuraA-P450	SI Fig 2
pJBL6684	PgrxA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PgrxA-P450	SI Fig 2
pJBL6685	PmtgA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PmtgA-P450	SI Fig 2
pJBL6686	PybcU – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PybcU-P450	SI Fig 2
pJBL6687	PycbS – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PycbS-P450	SI Fig 2
pJBL6688	PyhjX – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PyhjX-P450	SI Fig. 2
pJBL6689	PatoA – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PatoA-P450	SI Fig. 2
pJBL6690	PecpD – Target 8 – RBS 1 – CYP725A4-tcCPR – dblTerm – CmR – P15a origin	PecpD-P450	SI Fig. 2
N/A	PTrc – CYP725A4-tcCPR – rrnB – SpcR – SC101	P5Trc	2c
N/A	PTrc – CYP725A4-tcCPR – rrnB – CmR – p15a	P10Trc	2c
pJBL6691	apFAB346 – apFAB682 – EsaI -- LuxR — dblTerm – SpcR – SC101*	pQS	N/A

pJBL6692	PJ23115 – Target 8 – RBS 1 – mCherry – dblTerm – SpcR – pCDF origin	PJ23115-mCherry	SI. Fig 3

Supplementary Table B2. Examples of DNA plasmid sequences. Abbreviations as described in Supplementary Table 1.

Name	Sequence
PL,TetO1-STAR (P_R-TetR-dblTerm-PL,TetO1-STAR8-T500)	GAATTCTAAAGATCTTTTTCTCTATCACTGATAGGGAGTGGTAAAATAACTCTATCAACGAT AGAGTGTCAACAAAAATTAGGAATTAATGATGTCGAGATTAGATAAAAGTAAAGTGATTAAC AGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCC CAGAAGCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTG CTCGACGCCTTAGCCATTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAG GGGAAAGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGT CATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACAGTATGAAACT CTCGAAAATCAATTAGCCTTTTTATGCCAACAAGGTTTTTCACTAGAGAATGCATTATATGC ACTCAGCGCTGTGGGGCATTTTACTTTAGGTTGCGTATTGGAAGATCAAGAGCATCAAGTC GCTAAAGAAGAAAGGGAAACACCTACTACTGATAGTATGCCGCCATTATTACGACAAGCTA TCGAATTATTTGATCACCAAGGTGCAGAGCCAGCCTTCTTATTCGGCCTTGAATTGATCAT ATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAATAACACTGATAGTGCTAG TGTAGATCACTACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGG CCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACC TTCGGGTGGGCCTTTCTGCGTTTATATACTAGAGTCCCTATCAGTGATAGAGATTGACATC CCTATCAGTGATAGAGATACTGAGCACTGAACTGTATACATTCCCCGCAGGATAAGAGTAA GTGAGAGTAGGTAGAGATTGAGGATGGGGATCTCAAAGCCCGCCGAAAGGCGGGCTTTT TTTTGGATCCTTACTCGAGTCTAGACTGCAGGCTTCCTC
PLux-STAR (PLux-STAR8-T500)	CTAAAGATCTATATACTAGAGACCTGTAGGATCGTACAGGTTTACGCAAGAAAATGGTTTG TTATAGTCGAATAAATGAACTGTATACATTCCCCGCAGGATAAGAGTAAGTGAGAGTAGGT AGAGATTGAGGATGGGGATCTCAAAGCCCGCCGAAAGGCGGGCTTTTTTTTGGATCCTTA CTCGAGTCTAGACTGCAGGCTTCCTC
Example rSFP Plasmid (PompF-TARGET 8-RBS 1-mCherry-dblTerm)	CGATCATCCTGTTACGGAATATTACATTGCAACATTTACGCGCAAAAACTAATCCGCATTCT TATTGCGGATTAGTTTTTTCTTAGCTAATAGCACAATTTTCATACTATTTTTTGGCATTCTGG ATGTCTGAAAGAAGATTTTGTGCCAGGTCGATAAAGTTTCCATCAGAAACAAAATTTCCGTT TAGTTAATTTAAATATAAGGAAATCATATAAATAGATTAAAATTGCTGTAAATATCATCACGT CTCTATGGAAATATGACGGTGTTCACAAAGTTCCTTAAATTTTACTTTTGGTTACATATTTTT TCTTTTTGAAACCAAATCTTTATCTTTGTAGCACTTTCACGGTAGCGAAACGTTAGTTTGAA TGGAAAGATGCCTGCAGACACATAAAGACACCAAACTCTCATCAATAGTTCCGTAAATTTT TATTGACAGAACTTATTGACGGCAGTGGCAGGTGTCATAAAAAAAACCATGAGGGTAATAA ATACCATCCTCAATCTCTACCTACTCTCACTTACTCTTATCCTGCGGGGAATGTATACAGTT CATGTATATATTCCCCGCTTTTTTTTTGGATCTAGGAGGAAGGATCTATGGCGAGTAGCGA AGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCAC GAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAA ACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTT CCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACT GTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGT TACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGT GGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCT TCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTG AAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAA CCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAAC GAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCT TAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGA AAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACAC TGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA
CYP725A4/tc CPR fusion	ATGGCTCTGTTATTAGCAGTTTTTTTTAGCATCGCTTTGAGTGCAATTGCCGGGATCTTGCT GTTGCTCCTGCTGTTTCGCTCGAAACGTCATAGTAGCCTGAAATTACCTCCGGGCAAACT GGGCATTCCGTTTATCGGTGAGTCCTTTATTTTTTTGCGCGCGCTGAGGAGCAATTCTCTG GAACAGTTCTTTGATGAACGTGTGAAGAAGTTCGGCCTGGTATTTAAAACGTCCCTTATCG GTCACCCGACGGTTGTCCTGTGCGGGCCCGCAGGTAATCGCCTCATCCTGAGCAACGAA GAAAAGCTGGTACAGATGTCCTGGCCGGCGCAGTTTATGAAGCTGATGGGAGAGAACTCA GTTGCGACCCGCCGTGGTGAAGATCACATTGTTATGCGCTCCGCGTTGGCAGGCTTTTTC GGCCCGGGAGCTCTGCAATCCTATATCGGCAAGATGAACACGGAAATCCAAAGCCATATT AATGAAAAGTGGAAAGGGAAGGACGAGGTTAATGTCTTACCCCTGGTGCGGGAACTGGTT TTTAACATCAGCGCTATTCTGTTCTTTAACATTTACGATAAGCAGGAACAAGACCGTCTGCA

	CAAGTTGTTAGAAACCATTCTGGTAGGCTCGTTTGCCTTACCAATTGATTTACCGGGTTTC GGGTTTCACCGCGCTTTACAAGGTCGTGCAAAACTCAATAAAATCATGTTGTCGCTTATTA AAAAACGTAAAGAGGACTTACAGTCGGGATCGGCCACCGCGACGCAGGACCTGTTGTCT GTGCTTCTGACTTTCCGTGATGATAAGGGCACCCCGTTAACCAATGACGAAATCCTGGAC AACTTTAGCTCACTGCTTCACGCCTCTTACGACACCACGACTAGTCCAATGGCTCTGATTT TCAAATTACTGTCAAGTAACCCTGAATGCTATCAGAAAGTCGTGCAAGAGCAACTCGAGAT TCTGAGCAATAAGGAAGAAGGTGAAGAAATTACCTGGAAAGATCTTAAGGCCATGAAATAC ACGTGGCAGGTTGCGCAGGAGACACTTCGCATGTTTCCACCGGTGTTCGGGACCTTCCG CAAAGCGATCACGGATATTCAGTATGACGGATACACAATCCCGAAAGGTTGGAAACTGTT GTGGACTACCTATAGCACTCATCCTAAGGACCTTTACTTCAACGAACCGGAGAAATTTATG CCTAGTCGTTTCGATCAGGAAGGCAAACATGTTGCGCCCTATACCTTCCTGCCCTTTGGA GGCGGTCAGCGGAGTTGTGTGGGTTGGGAGTTCTCTAAGATGGAGATTCTCCTCTTCGTG CATCATTTCGTGAAAACATTTTCGAGCTATACCCCGGTCGATCCCGATGAAAAAATTTCCG GCGATCCACTGCCGCCGTTACCGAGCAAAGGGTTTTCAATCAAACTGTTCCCTCGTCCGg gcagcaccggatccCGCCGTGGTGGAAGTGATACACAGAAGCCCGCCGTACGTCCCACACCTC TTGTTAAAGAAGAGGACGAAGAAGAAGAAGATGATAGCGCCAAGAAAAAGGTCACAATAT TTTTTGGCACCCAGACCGGCACCGCCGAAGGTTTCGCAAAGGCCTTAGCTGAGGAAGCA AAGGCACGTTATGAAAAGGCGGTATTTAAAGTCGTGGATTTGGATAACTATGCAGCGGAT GACGAACAGTACGAAGAGAAGTTGAAAAAGGAAAAGCTAGCGTTCTTCATGCTCGCCACC TACGGTGACGGCGAACCGACTGATAATGCCGCTCGCTTTTATAAATGGTTTCTCGAGGGT AAAGAGCGCGAGCCATGGTTGTCAGATCTGACTTATGGCGTGTTTGGCTTAGGTAACCGT CAGTATGAACACTTTAACAAGGTCGCGAAAGCGGTGGACGAAGTGCTCATTGAACAAGGC GCCAAACGTCTGGTACCGGTAGGGCTTGGTGATGATGATCAGTGCATTGAGGACGACTTC ACTGCCTGGAGAGAACAAGTGTGGCCTGAGCTGGATCAGCTCTTACGTGATGAAGATGAC GAGCCGACGTCTGCGACCCCGTACACGGCGGCTATTCCAGAATACCGGGTGGAAATCTA CGACTCAGTAGTGTCGGTCTATGAGGAAACCCATGCGCTGAAACAAAATGGACAAGCCGT ATACGATATCCACCACCCGTGTCGCAGCAACGTGGCAGTACGTCGTGAGCTGCATACCCC GCTGTCGGATCGTAGTTGTATTCATCTGGAATTCGATATTAGTGATACTGGGTTAATCTAT GAGACGGGCGACCACGTTGGAGTTCATACCGAGAATTCAATTGAAACCGTGGAAGAAGCA GCTAAACTGTTAGGTTACCAACTGGATACAATCTTCAGCGTGCATGGGGACAAGGAAGAT GGAACACCATTGGGCGGGAGTAGCCTGCCACCGCCGTTTCCGGGGCCCTGCACGCTGC GGACGGCGCTGGCACGTTACGCGGACCTGCTGAACCCTCCGCGCAAAGCCGCCTTCCTG GCACTGGCCGCACACGCGTCAGATCCGGCTGAAGCTGAACGCCTTAAATTTCTCAGTTCT CCAGCCGGAAAAGACGAATACTCACAGTGGGTCACTGCGTCCCAACGCAGCCTCCTCGA GATTATGGCCGAATTCCCCAGCGCGAAACCGCCGCTGGGAGTGTTTTTCGCCGCAATAGC GCCGCGCTTGCAACCTAGGTATTATAGCATCTCCTCCTCCCCGCGTTTCGCGCCGTCTCG TATCCATGTAACGTGCGCGCTGGTCTATGGTCCTAGCCCTACGGGGCGTATTCATAAAGG TGTGTGCAGCAACTGGATGAAGAATTCTTTGCCCTCCGAAGAAACCCACGATTGCAGCTG GGCACCGGTCTTTGTGCGCCAGTCAAACTTTAAACTGCCCGCCGATTCGACGACGCCAAT CGTGATGGTTGGACCTGGAACCGGCTTCGCTCCATTTCGCGGCTTCCTTCAGGAACGCG CAAAACTGCAGGAAGCGGGCGAAAAATTGGGCCCGGCAGTGCTGTTTTTTGGGTGCCGC AACCGCCAGATGGATTACATCTATGAAGATGAGCTTAAGGGTTACGTTGAAAAAGGTATTC TGACGAATCTGATCGTTGCATTTTCACGAGAAGGCGCCACCAAAGAGTATGTTCAGCACA AGATGTTAGAGAAAGCCTCCGACACGTGGTCTTTAATCGCCCAGGGTGGTTATCTGTATG TTTGCGGTGATGCGAAGGGTATGGCCAGAGACGTACATCGCACCCTGCATACAATCGTTC AGGAACAAGAATCCGTAGACTCGTCAAAAGCGGAGTTTTTAGTCAAAAAGCTGCAAATGG ATGGACGCTACTTACGGGATATTTGGTAA
QS operon (apFAB346- apFAB682- EsaI-B0034- LuxR- dblTerm)	TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGAGGGCCCAAGTTCACTTAAAAAGGAG ATCAACAATGAAAGCAATTTTCGTACTGAAACATCTTAATCATGCTAAGGAGGTTTTCTAAT GATGCTTGAACTGTTTGACGTCAGTTACGAAGAACTGCAAACCACCCGTTCAGAAGAACTT TATAAACTTCGCAAGAAAACATTTAGCGATCGTCTGGGATGGGAAGTCATTTGCAGTCAGG GAATGGAGTCCGATGAATTTGATGGGCCCGGTACACGTTATATTCTGGGAATCTGCGAAG GACAATTAGTGTGCAGCGTACGTTTTACCAGCCTCGATCGTCCCAACATGATCACGCACA CTTTTCAGCACTGCTTCAGTGATGTCACCCTGCCCGCCTATGGTACCGAATCCAGCCGTTT TTTTGTCGACAAAGCCCGCGCACGTGCGCTGTTAGGTGAGCACTACCCTATCAGCCAGGT CCTGTTTTTAGCGATGGTGAACTGGGCGCAAAATAATGCCTACGGCAATATCTATACGATT GTCAGCCGCGCGATGTTGAAAATTCTCACTCGCTCTGGCTGGCAAATCAAAGTCATTAAA GAGGCTTTCCTGACCGAAAAGGAACGTATCTATTTGCTGACGCTGCCAGCAGGTCAGGAT GACAAGCAGCAACTCGGTGGTGATGTGGTGTCACGTACGGGCTGTCCGCCCGTCGCAGT CACTACCTGGCCGCTGACGCTGCCGGTCTGATACTAGAGAAAGAGGAGAAATACTAGATG AAAAACATAAATGCCGACGACACATACAGAATAATTAATAAAATTAAAGCTTGTAGAAGCAA TAATGATATTAATCAATGCTTATCTGATATGACTAAAATGGTACATTGTGAATATTATTTACT CGCGATCATTTATCCTCATTCTATGGTTAAATCTGATATTTCAATCCTAGATAATTACCCTAA AAAATGGAGGCAATATTATGATGACGCTAATTTAATAAAATATGATCCTATAGTAGATTATT CTAACTCCAATCATTCACCAATTAATTGGAATATATTTGAAAACAATGCTGTAAATAAAAAAT CTCCAAATGTAATTAAAGAAGCGAAAACATCAGGTCTTATCACTGGGTTTAGTTTCCCTATT CATACGGCTAACAATGGCTTCGGAATGCTTAGTTTTGCACATTCAGAAAAAGACAACTATA TAGATAGTTTATTTTTACATGCGTGTATGAACATACCATTAATTGTTCCTTCTCTAGTTGATA ATTATCGAAAAATAAATATAGCAAATAATAAATCAAACAACGATTTAACCAAAAGAGAAAAA GAATGTTTAGCGTGGGCATGCGAAGGAAAAAGCTCTTGGGATATTTCAAAAATATTAGGTT

	GCAGTGAGCGTACTGTCACTTTCCATTTAACCAATGCGCAAATGAAACTCAATACAACAAA CCGCTGCCAAAGTATTTCTAAAGCAATTTTAACAGGAGCAATTGATTGCCCATACTTTAAAA ATTAATCTAGAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCT CAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGA GTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATACCTAGG

Supplementary Table B3. Sequence of Promoter and RBS variants. Pstress promoters were PCR amplified from the *E. coli* K-12 MG1655 genome.

Name	Sequence
RBS 1	AGGAGGAA
B0034 RBS	AAGAGGAGAAA
$P_{L,TetO1}$	TCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAG CAC
P_{Lux}	ACCTGTAGGATCGTACAGGTTTACGCAAGAAAATGGTTTGTTATAGTCGAA TAAA
PJ23115	TTTATAGCTAGCTCAGCCCTTGGTACAATGCTAGC
PgntK	AATCTGTGACACCGAAAATGTTAGATTTAGGTTTCACCTTGTCACCGGGCG GATCTATTTAAGCCCACAAATTTGAAGTAGCTCACACTTATACACTTAAGG CATGGATGGATATTGCTTCTGATATTGTCCGGCTGGACAATGTTACCGATA ACAGTTACCCGTAACATTTTTAATTCTTGTATTGTGGGGGCACCACT
PompF	CGATCATCCTGTTACGGAATATTACATTGCAACATTTACGCGCAAAAACTA ATCCGCATTCTTATTGCGGATTAGTTTTTTCTTAGCTAATAGCACAATTTTC ATACTATTTTTTGGCATTCTGGATGTCTGAAAGAAGATTTTGTGCCAGGTC GATAAAGTTTCCATCAGAAACAAAATTTCCGTTTAGTTAATTTAAATATAAG GAAATCATATAAATAGATTAAAATTGCTGTAAATATCATCACGTCTCTATGG AAATATGACGGTGTTCACAAAGTTCCTTAAATTTTACTTTTGGTTACATATTT TTTCTTTTTGAAACCAAATCTTTATCTTTGTAGCACTTTCACGGTAGCGAAA CGTTAGTTTGAATGGAAAGATGCCTGCAGACACATAAAGACACCAAACTCT CATCAATAGTTCCGTAAATTTTTATTGACAGAACTTATTGACGGCAGTGGC AGGTGTCATAAAAAAAACCATGAGGGTAATAAATA
PyeeF	ATTAGCGGCCTCGGCTGCGGCTATTTACCCCGTTATCTGGCGCAACGTTT TCTCGATAGTGGCGCGTTAATCGAGAAGAAAGTGGTCGCCCAAACTCTCT TTGAACCCGTCTGGATTGGCTGGAACGAACAGACCGCAGGACTTGCCAGT GGCTGGTGGCGGGATGAAATTTTAGCAAATAGTGCGATCGCCGGTGTTTA TGCAAAATCTGATGACGGAAAATCAGCCATTTAAAGAAAAATTATTCTGAC AAGCCTCTCATTCTCTTGTCATTTCCCCCCCATTTAGGCACAATGCGCCGC TGTCAAAAAATGACTAAAAACCGACGTTTCATCAGCGTCGGTTATTTTTTG CTTCAAACCAATCATTCATACCAAGAGGCCGGGCTTCGTACCGGATAGAT ATTTACTAAAAATCGACAGTTGTTGTCGCTGAGGAATCCAAAAAAATGGGG CAATTTTTTGCTTACGCGACGGTTATCACCGTAAAGGAGAATGACC
PompT	AACGGATAAGACGGGCATAAATGAGGAAGAAATGGCGCGCCCTGCAGGA TTCGAACCTGCGGCCCACGACTTAGAAGTTCCTAGAACGACATTTTAAGTC AACAACTTACCGCGCCATCTCTGCGCTCACACGTCCCACTACCTCAAAAC ATGTAAAGCCTTGCAAGCCATTGCGAGGCCTTATGTGTCTCAGTTTTGTCC CTCTTTTTTGTACTAAAAAACATAGTAATTGAGGATAAACCTCATGCTATTT TCGCTTATATGCCTCTAAAGGCATGGCACTTAAATAGATAAAAGCACCACA AAAGCATAAAAAAACCACACAGTAAAACCGAAATATGAAACAATAACAGAT AATTAAACCAAAAACAGATAGCGCATTGTGATAATCATTCAATACTAAACAA AATATAAACAGTGGAGCAATATGTAATTGACTCATTAAGTTAGATATAAAAA ATACATATTCAATCATTAAAACGATTGAATGGAGAACTTTT
PmetN	GGCGAAACTCTTCAACACTACCCTGCGGATGATGCGGGCAATAATAGATA CCATCCAGATCGACATCTCGGTCCGCCAGCGACCAGTCCATCCACTCGGT CAGCGTTTCAAACTGTGCTTCGGTAAATTTACCGCGAGCAATGCCAGACT GGTTGGTTACTACCACCAGCGCAAAGCCCATTTTTTTTAGCTCGCGCATG GCGTCAATAACACCGTCGATAAATTCAAAGTTGTCGATCTCATGGACATAG CCGTGATCGACATTAATGGTGCCATCACGGTCAAGAAAAATTGCGGGTAC GCTCTTCGCCACCTTTTATAGCTCCTTAATAAGGCATGTGACGCTAGTATC GCATGTTTCGACCTGCAAGAAAGTGCTCTTCGCATAAACCTGATTGATTTA GACGTCTGGATGCCTTAACATCCATTTCATTGACGGCGTTGCCCGTTTCAG GCATTCGAGATGCCACGACTAACTTAATGACGATAATAAATAATCA

Pb1762	GATTATTGAACTGTTGTTCAAGCGTGGTTTCCTGACCAAAAAAGGGCGCTA TATCCACTCCACCGACGCCGGAAAAGCGCTATTCCATTCGCTGCCGGAGA TGGCGACGCGACCGGACATGACCGCGCACTGGGAATCGGTGCTGACGCA AATCAGCGAAAAGCAGTGTCGCTATCAGGACTTTATGCAGCCGCTGGTGG GGACGCTATATCAGCTTATTGATCAAGCCAAACGTACGCCGGTGCGGCAG TTTCGCGGCATTGTGGCTCCGGGCAGTGGTGGCAGTGCTGATAAGAAAAA GGCTGCACCGCGTAAACGTAGTGCGAAAAAAAGTCCGCCAGCAGATGAA GTCGGAAGCGGGGCGATAGCGTAAGCGAGTGAATCTTTCGTGCTATTCGA GTCATATTCTGAAATATCCAGCGGATCAAGAAAATTCGTTGGATATTTTTTT TGCATGGATAAAATTATCGCCTCTAAAGTATGTAATAACAGGGAATGTG
PcarA	GGTCTTTTTGATATGCGAGATGTACTTGATCTCAATAATTTGTAACCACAAA ATATTTGTTATGGTGCAAAAATAACACATTTAATTTATTGATTATAAAGGGC TTTAATTTTTGGCCCTTTTATTTTTGGTGTTATGTTTTTAAATTGTCTATAAG TGCCAAAAATTACATGTTTTGTCTTCTGTTTTTGTTGTTTTAATGTAAATTTT GACCATTTGGTCCACTTTTTTCTGCTCGTTTTTATTTCATGCAATCTTCTTG CTGCGCAAGCGTTTTCCAGAACAGGTTAGATGATCTTTTTGTCGCTTAATG CCTGTAAAACATGCATGAGCCACAAAATAATATAAAAAATCCCGCCATTAA GTTGACTTTTAGCGCCCATATCTCCAGAATGCCGCCGTTTGCCAGAAATTC GTCGGTAAGCAGATTTGCATTGATTTACGTCATCATTGTGAATTAATATGCA AATAAAGTGAGTGAATATTCTCTGGAGGGTGTT
PfadL	CGTTTGCTCTGCTTCTGCGCGGTTTGCAAACACGCGGCTGTAAGACGCGG TGCAGTCGGAGTTGTCCATAATGGTGCCAACATCCATACAGCAGCAAACC GGGGTTTCATCAGCACTACATTTACTCATCGTTGATTTCCTCTGTATGTGC ACCCAAGGTGCCAGATAAACGTTGTGGATATTTTACGCTTCCGGAAAGTG CTGCTCCAGTTGTTAATTCTGCAAAATCGGATAAGTGACCGAAATCACACT TAAAAATGATCTAAAACAAAATTCACCCGAATCCATGAGTGCGCCACCTCC AAATTTTGCCAGCTGGATCGCGTTTCTTAGATCATATTTGAAAAAAGATAGA AACATACTTGCAACATTCCAGCTGGTCCGACCTATACTCTCGCCACTGGTC TGATTTCTAAGATGTACCTCAGACCCTACACTTCGCGCTCCTGTTACAGCA CGTAACATAGTTTGTATAAAAATAAATCATTGAGGTTATGGTC
PuraA	TACCGTTGTGCCAATTCTGCGTGCGGGTCTTGGTATGATGGACGGTGTGC TGGAAAACGTTCCGAGCGCGCGCATCAGCGTTGTCGGTATGTACCGTAAT GAAGAAACGCTGGAGCCGGTACCGTACTTCCAGAAACTGGTTTCTAACAT CGATGAGCGTATGGCGCTGATCGTTGACCCAATGCTGGCAACCGGTGGTT CCGTTATCGCGACCATCGACCTGCTGAAAAAAGCGGGCTGCAGCAGCATC AAAGTTCTGGTGCTGGTAGCTGCGCCAGAAGGTATCGCTGCGCTGGAAAA AGCGCACCCGGACGTCGAACTGTATACCGCATCGATTGATCAGGGACTGA ACGAGCACGGATACATTATTCCGGGCCTCGGCGATGCCGGTGACAAAATC TTTGGTACGAAATAAAGAATAAAAATAATTAAAGCCGACTTTAAGAGTCGG CTTTTTTTTGAGTAAAGCGCCTATAACACATAATACAGAGGATAATACT
PgrxA	CGGAAATGGGTTCATCAGTGAAATGGCGAATGGAGCGATGGCCACAAATA AGTTCAATGGTTGGCGTCATTATCTTTTTCTCTTTCTGAACGTGAATATTGC GGTGGACGGTTCATCAGCTGTGGGGCAAGACGTTTTGCCACCTGAAGAAT AACCACCACCGCAGCGGGAAGCATGAGCAAAACACCGAGAAAAATCATCA GAATCTGCACTTCTGGCCGAGAAAATGGCTCAGGCAGCGACAGGGAGTC GCTTACCGACAGCAGCGCCACCGCCAGTAGCATCATTCCGATAAATTCCA GTATCAACACGCCTTTAGGCAATTTACCGATCGCGCGCATACGCTTCCCT CTGCAAAGTGAGCCTTCAGTCTAAAACTTTTCACTGTATTGTGTTTAACAGT TATAGCTTTTAGCAATTAATGCAACAGGTTAAACCTACTTTCAGCGAATACA TTTTAGCGTGATCATTACAGGCATAAATCTATGAGGAGAGAAATA
PmtgA	CATTTAACTGGCGAAGCGATGACGGAAACGCGCAATGTGCTGATTGAAGC GGCACGAATAACGCGCGGTGAAATCCGTCCTCTGGCCCAGGCCGATGCC GCTGAACTGGATGCGTTGATTGTGCCGGGGGGGGTTTGGCGCGGCGAAGA ATTTAAGCAATTTTGCCAGTCTTGGTAGCGAATGCACCGTTGACCGTGAAT TAAAGGCGCTGGCACAAGCGATGCATCAGGCCGGAAAACCGCTTGGTTTT ATGTGTATTGCCCCGGCGATGCTGCCGAAAATTTTCGATTTCCCGCTGCG TTTGACCATCGGTACTGATATCGATACCGCAGAAGTGCTGGAAGAGATGG GCGCGGAGCATGTGCCGTGTCCTGTCGATGATATCGTGGTTGATGAAGAC AATAAGATTGTCACCACCCCAGCATATATGCTGGCGCAGAACATTGCAGA AGCGGCGAGCGGCATTGATAAGCTGGTTTCCCGCGTGCTGGTTCTGGCT GA
PybcU	TCAAGAAATACGCATCTTATAGAAACGTCCTATGATAGGTTGAAATCAAGA GAAATCACATTTCAGCAATACAGGGAAAATCTTGCTAAAGCAGGAGTTTTC CGATGGATTACAAATATCCACGAACATAAAAGATATTACTATACCTTTGATA ATTCATTACTATTTACTGAGAGCATTCAGAACACTACACAAATCTTTCCACG CTAAATCATAACGTCCGGTTTCTTCCGTGTCAGCACCGGGGTGTTGGCAT AATACAATACATGTACGCGCTAAACCCTGTGTGCATCGTTTTTAATTATTCC CGGACACTCCCGCAGAGAAGTTCCCCGTCAGGGCTGTGGACATAGTTAAT CCGGGAATACAATGACGATTCATCGCACCTGGCATACATTAATAAATATTA

	ACAATATGAAATTTCAACTCATTGTTTAGGGTTTGTTTAATTTTCTACACATA CGATTCTGCGAACTTCAAAAAGCATCGGGAATAACACC
PycbS	CAGACTTTATTATTACACCACCGCTATTTGTGCTGAATCCGGCAAATGAGA ATCTGTTACGCATTATGTACATTGGAGCGCCGTTGGCGAAAGACAGAGAA ACCCTTTTCTTCACTAGCGTACGGGCAGTCCCTTCAACAACGAAGCGGAA AGAGGGAAATACCCTGAAGATTGCCACACAAAGCGTCATCAAACTTTTCTG GCGACCAAAAGGTTTAGCGTATCCCTTAGGCGAGGCTCCGGCGAAACTG CGTTGCACTTCGTCAGCTGACATGGTTACGGTCAGTAACCCAACACCTTAT TTCATTACCCTGACAGACCTGAAAATAGGTGGAAAAGTAGTTAAAAATCAA ATGATTTCCCCCTTTGATAAATACCAATTTTCTCTGCCAAAGGGGGCCAAA AATAGCAGCGTAACGTATCGAACCATCAATGACTACGGGGCGGAAACGCC GCAACTCAACTGTAAATCGTAAGCCGTCTTCAGTTAAGAGAGCGAG
PyhjX	TAGGCTGGCGTGTTGACTCCCGGCTTGGCGATCTCCGACCCTGGGCGCA AATCAGCTATAACCAGCAATTTGGCGAGAATATCTGGAAGGCGCAATCAG GCCTGAGCCGGATGACGGCGACAAACCAGAACGGCAACTGGCTGGATGT CACCGTAGGCGCTGATATGTTGCTCAATCAAAATATTGCCGCCTATGCCG CGCTAACTCAGGCAGAAAATACCACTAATAATAGCGACTATCTGTATACGA TGGGGGTTAGCGCCAGATTTTAACGTAACAGTCACAATTGAAACCATTAAA TAACAATAGTTGTGGCGATAGTGGGTGCTAACTTACCAAATAATAAATTTG GTGAATAATTGTCGCGTCATTCATTCCTGAACTAAGGCATTTCATTCCGTT CTGATGGCATTTCATGCCGTTTTTCCCCAGGCATAAAGTGCACTTCGTTAT GGTTGTCGGCAGAGATTTTTCCTTTTTATTACTGCAGGAATACTGCC
PatoA	GCGTTTGTTGATACCGGCATCGGTCCGCTCATCGTCAATGGTCGAGTCCG CAAAGTGATTGCTTCACATATCGGCACCAACCCGGAAACAGGTCGGCGCA TGATATCTGGTGAGATGGACGTCGTTCTGGTGCCGCAAGGTACGCTAATC GAGCAAATTCGCTGTGGTGGAGCTGGACTTGGTGGTTTTCTCACCCCAAC GGGTGTCGGCACCGTCGTAGAGGAAGGCAAACAGACACTGACACTCGAC GGTAAAACCTGGCTGCTCGAACGCCCACTGCGCGCCGACCTGGCGCTAA TTCGCGCTCATCGTTGCGACACACTTGGCAACCTGACCTATCAACTTAGC GCCCGCAACTTTAACCCCCTGATAGCCCTTGCGGCTGATATCACGCTGGT AGAGCCAGATGAACTGGTCGAAACCGGCGAGCTGCAACCTGACCATATTG TCACCCCTGGTGCCGTTATCGACCACATCATCGTTTCACAGGAGAGCAAA TA
Pb2970	GGATTATTAAGTGGCTGTGCCAGCCATAATGAAAATGCCAGTTTACTGGC GAAAAAACAGGCGCAAAATATCAGCCAAAACCTGCCGATTAAATCTGCGG GATATACCTTAGTGCTGGCGCAAAGTAGCGGCACAACGGTAAAAATGACC ATTATCAGCGAAGCGGGTACACAAACCACGCAGACGCCTGACGCCTTTTT AACCAGCTATCAACGACAAATGTGCGCTGACCCGACGGTGAAATTAATGA TCACTGAGGGAATTAATTACAGCATAACGATTAATGATACACGTACAGGTA ACCAGTATCAGCGGAAACTGGATCGTACCACCTGTGGAATAGTCAAAGCA TAACGTCGGGTAGATATAAATTGGCGCGGGTTGTTTTTCGTGACGCACGA ATTTATCTCATTCAATGGCTGACAAAAATTCGTCACACTCTTAACCAGAGAC AATCTCTTAATACAGACAAAGAGCATCTGCGAAAAATTGCACGCGGG
PecpD	CGGGCTGGAGGACGACGGTCAGATCAGCGCCAAAATCAACGGGCGGATT TTCCCGCTTAACGGCAAGCGTAACTATCTCCCGCTCTCTCCCTATGGAAG ATATGAGGTGGAGTTACAGAACAGCAAAAACTCACTCGACAGTTACGATAT CGTCAGCGGCCGCAAAAGTCGTCTGACTCTCTATCCAGGCAATGTCGCTG TCATTGAGCCAGAGGTGAAGCAGATGGTTACCGTCTCCGGTCGTATCCGT GCGGAAGACGGCACACTGCTGGCTAACGCACGGATTAACAACCATATCG GCCGAACCCGAACCGATGAAAACGGCGAGTTTGTCATGGACGTGGATAA GAAATACCCCACTATCGATTTTCGCTACAGTGGCAATAAAACCTGCGAAGT GGCTCTGGAACTCAACCAGGCGCGCGGTGCCGTCTGGGTCGGTGATGTG GTCTGCAGCGGCCTCTCATCGTGGGCGGCGGTGACGCAGACAGGAGAAG AGA
PfecA	GGGACAGAATTTACCGTCCGCCAGCAGGATAATTTCACGCAGCTTGACGT GCAGCAGCACGCTGTGGAAGTGCTTCTCGCCAGTGCCCCCGCGCAAAAA CGCATCGTGAACGCTGGTGAAAGCCTGCAGTTCAGCGCCTCTGAGTTTGG CGCAGTGAAACCGCTGGATGACGAGAGTACAAGCTGGACGAAGGACATC CTGAGCTTCAGCGATAAACCGCTGGGTGAGGTGATAGCCACGCTAACCC GTTACCGCAACGGCGTGCTGCGCTGCGATCCCGCCGTTGCCGGGCTGCG CCTGAGCGGGACGTTCCCGCTGAAAAATACCGATGCGATCCTGAACGTTA TCGCGCAAACGCTTCCCGTTAAAATTCAGTCTATTACGCGGTACTGGATAA ACATTTCACCACTGTAAGGAAAATAATTCTTATTTCGATTGTCCTTTTTACC CTTCTCGTTCGACTCATAGCTGAACACAACAAAAATGATGATGGGGAAGGT

Supplementary Table B4. Strains used in this study. Strains containing genomic insertions were created using the clonetegration platform to integrate the inserts using

the HK022 plasmid into the *attB* site of the *E. coli* genome[5]. Successful integrations were identified by antibiotic selection and colony PCR according to the published protocol.[5]

Strain	Strain Information	Genomic Insertion
E. coli Tax1	E. coli containing genome integrated pathway enzymes for taxadiene biosynthesis (gift from Manus Bio)	N/A
E. coli Tax1-QS	Derived from E. coli Tax1	*attB::EsaI-LuxR(*apFAB346-apFAB382-EsaI-LuxR-dblTerm – KmR)

Supplementary Table B5. Basal R-media recipe, per liter. Adapted from Biggs et al., 2016.[6]

Component	Final media concentration (g/l)
KH_2PO_4	13.3
$(NH_4)_2HPO_4$	4
Citric Acid Monohydrate	1.7
Yeast Extract	5
HEPES	23.83

Supplementary Table B6. 1000x Trace Element (TE) solution, per liter. Adapted from Biggs et al., 2016.

Component	Final media concentration (mg/l)
EDTA	8.4
H_3BO_3	3.0
$Zn(CH_3COO)_2$	8.0
$CoCl_2•6H_2O$	4.6
$CuCl2•2H_2O$	1.9
$MnCl_2•4H_2O$	24.0
$Na_2MoO_4•2H_2O$	2.9

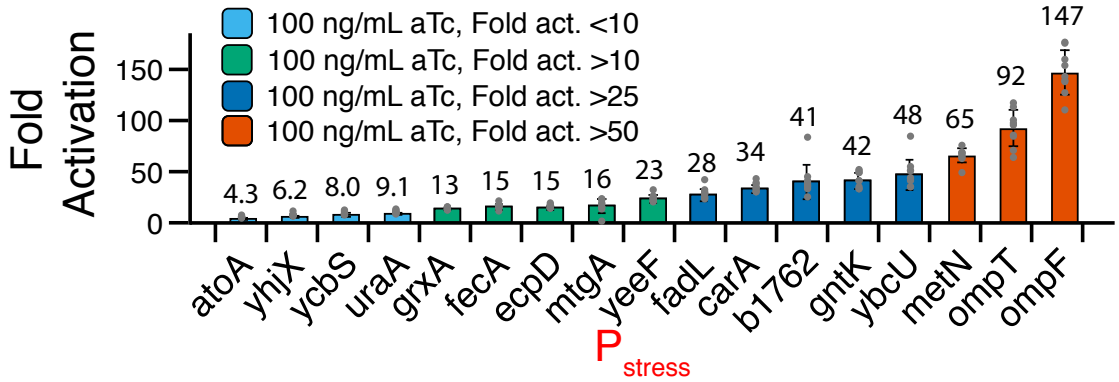
Supplementary Table B7. Complete R-media compositions utilized for hungate tube fermentations. Adapted from Biggs et al., 2016.

Component	Amount (mL)
Basal R-media	35

32% v/v Glycerol	1.3
1 M $MgSO_4$	0.171
0.1 M Ferric Citrate	0.0858
1000x TE Solution	0.035
1000x Antibiotic	0.035
1 M Thiamine HCl	0.00047

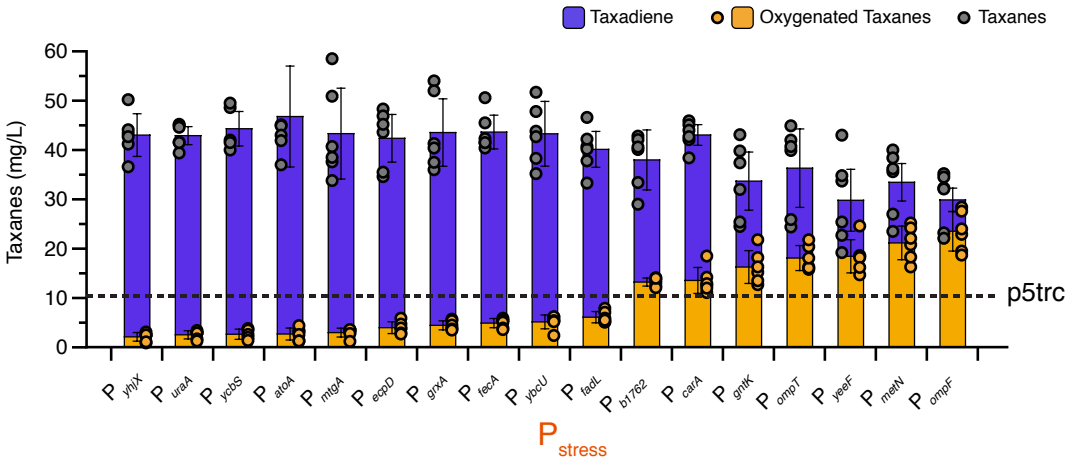
Supplementary Table B8. P values for two-tailed paired-sample t-tests in Fig. 1D.

Condition	P value
PatoA	2.73E-08
PyhjX	6.17E-07
PycbS	2.13E-06
PuraA	3.52E-09
PgrxA	8.06E-10
PfecA	6.01E-08
PecpD	2.29E-11
PmtgA	4.27E-08
PyeeF	1.15E-10
PfadL	4.54E-10
PcarA	1.86E-08
Pb1762	4.14E-09
PgntK	1.49E-08
PybcU	1.24E-07
PmetN	9.29E-08
PompT	1.36E-05
PompF	2.02E-07



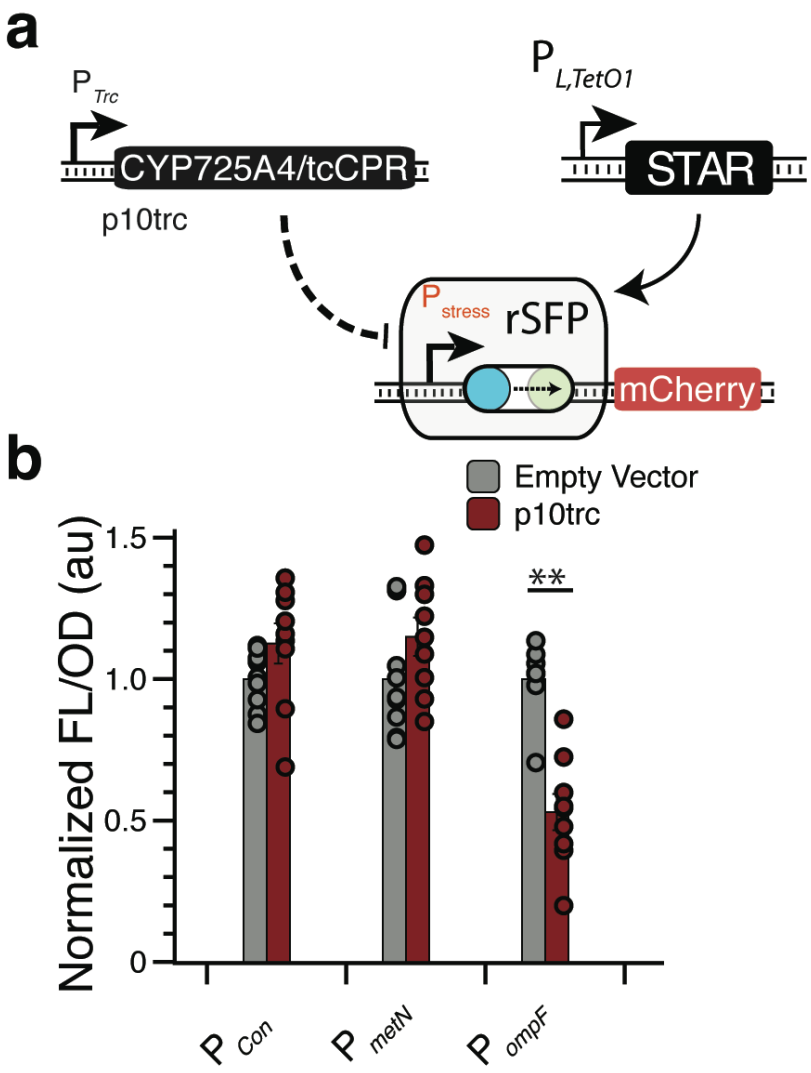
SI Figure B.8-1. Fold activation (ON/OFF) of rSFP variants containing unique envelope stress-response promoters.

Fluorescence characterization was performed on *E. coli* transformed with plasmids encoding each rSFP controlling mCherry expression and $P_{L,TetO1}$-STAR in the absence and presence of 100 ng/mL aTc. Data represent mean values of fold activation calculated as fluorescence/optical density (FL/OD) for each colony grown with 100 ng/mL aTc over the same colony without aTc. Error bars represent s.d. of at least n = 7 biological replicates. Mean fold activation for each rSFP variant is indicated above each bar. Grey points represent individual data points.



SI Figure B.8-2. Titers of fermentations after 96 hrs with *E. coli* Tax1 containing CYP725A4/tcCPR under control of complete rSFP library and $P_{L,TetO1}$-STAR with addition of 100 ng/mL aTc at inoculation.

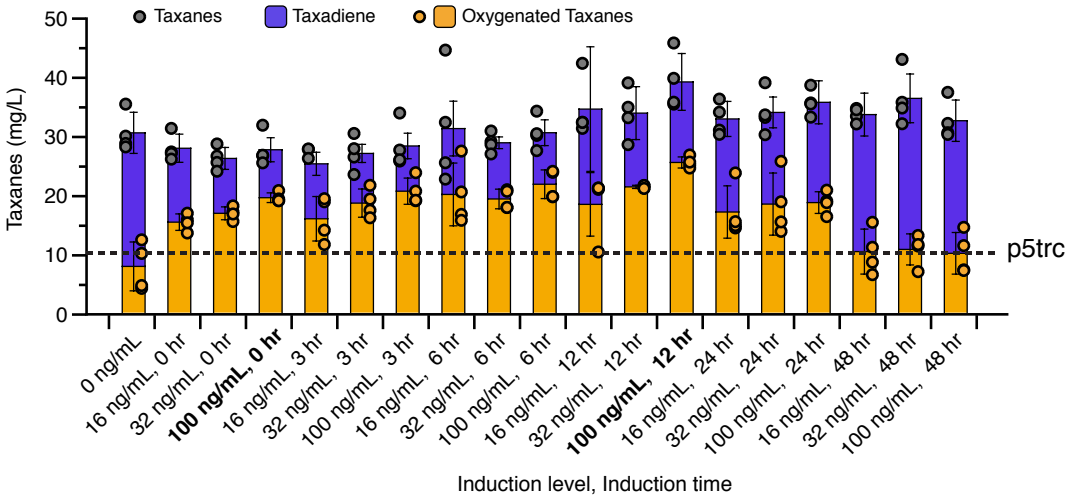
Dashed line represents production of oxygenated taxanes from p5Trc in Fig. 2C. Data represent mean values measured with GC-MS and error bars represent s.d. of at least n = 5 biological replicates. Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.



SI Figure B.8-3. Analysis of feedback-responsiveness of selected stress-response promoters to CYP725A4/tcCPR stress.

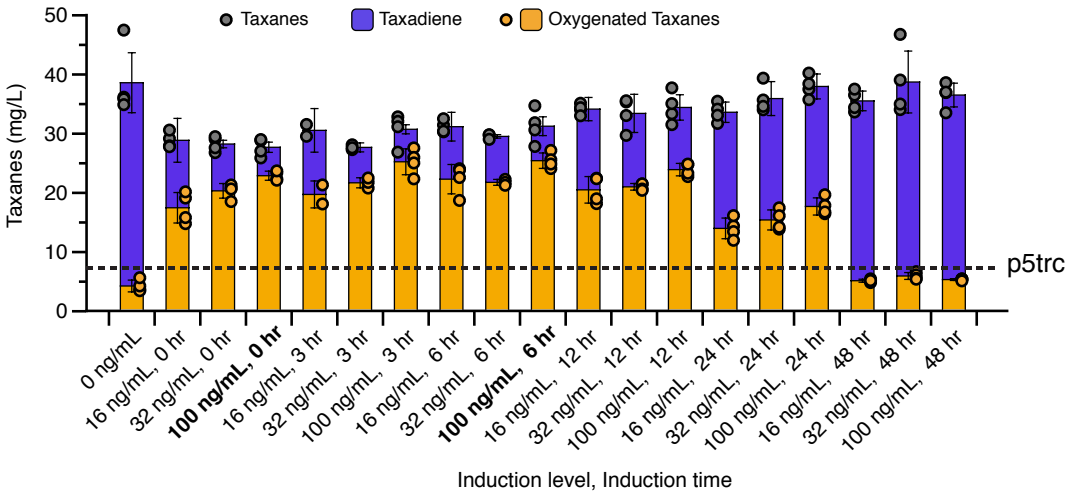
(a) Schematic of plasmids used for fluorescence characterization of rSFP stress response. $P_{L,TetO1}$-STAR was used to activate expression from select rSFP plasmids. p10Trc was used to induce CYP275A4/tcCPR stress in comparison with an empty vector. Monitoring rSFP controlled expression of mCherry then allows the response to CYP275A4/tcCPR stress to be characterized. (b) Fluorescence characterization of cells containing select $P_{L,TetO1}$-STAR activated rSFPs controlling mCherry expression with 100

ng/mL aTc, and either an empty vector or the p10Trc vector to express CYP725A4/tcCPR and induce membrane stress. Fluorescence values were normalized to the empty vector control and error bars represent standard error of the mean. Experiments were performed as in main Figure 1d except 20 µL of each overnight culture were added to 490 µL of R-media containing selective antibiotics and grown for 4 h at 22C to closely mimic hungate fermentations. 100 ng/mL aTc was added after 4 hrs growth. After another 6 hrs of growth at 22C, 100 µL were sampled for characterization by bulk fluorescence measurements. Pcon = PJ23115. Data represent mean values of arbitrary fluorescence/optical density (FL/OD) normalized to the empty vector control for each condition and error bars represent standard error of the mean (s.e.m) of at least n = 7 biological replicates. Colored points represent individual data points. * indicate a statistically significant difference in FL/OD by a two-tailed Welch's t-test (** P = 0.0000261).



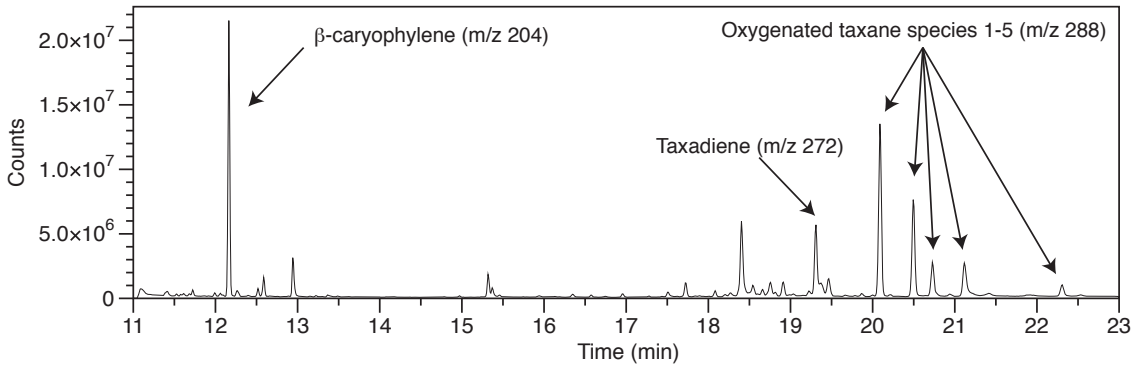
SI Figure B.8-4. Titers of fermentations after 96 hrs with *E. coli* Tax1 containing CYP725A4/tcCPR controlled by the P_{metN} rSFP and $P_{L,TetO1}$-STAR under each induction condition.

Dashed line represents production of oxygenated taxanes from p5Trc in Fig. 2C. Data represent mean values of fermentation titers and error bars represent s.d. of at least n = 3 biological replicates. Bold conditions indicate 100 ng/mL aTc induction at inoculation and the optimal inductions. Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.



SI Figure B.8-5. Titers of fermentations after 96 hrs with *E. coli* Tax1 containing CYP725A4/tcCPR controlled by the P_{ompF} rSFP and $P_{L,TetO1}$-STAR under each induction condition.

Dashed line represents production of oxygenated taxanes from p5Trc in Fig. 2C. Data represent mean values of fermentation titers and error bars represent s.d. of at least n = 3 biological replicates. Bold conditions indicate 100 ng/mL aTc induction at inoculation and the optimal inductions. Colored points represent individual data points. Grey filled points represent individual data points of overall taxanes and orange filled points represent individual data points of oxygenated taxanes.



SI Figure B.8-6. Example GC chromatogram for analysis of taxadiene and oxygenated taxane fermentations.

Taxadiene and oxygenated taxane peaks were previously described in Biggs et al.[6]

References:

1. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* (2013). doi:10.1038/nmeth.2404
2. Meier, I., Wray, L. V & Hillen, W. Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. *EMBO J.* (1988). doi:10.1002/j.1460-2075.1988.tb02846.x
3. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. *Nucleic Acids Res.* (1997). doi:10.1093/nar/25.6.1203
4. Engebrecht, J. & Silverman, M. Identification of genes and gene products necessary for bacterial bioluminescence. *Proc. Natl. Acad. Sci. U. S. A.* (1984). doi:10.1073/pnas.81.13.4154
5. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* (2013). doi:10.1021/sb400021j
6. Biggs, B. W. *et al.* Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1515826113