Title
Rational Design of Genetic Circuits and Biological Modules

Author
Tei, Mika

Publication Date
2018

Rational Design of Genetic Circuits and Biological Modules

By

Mika Tei

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Adam Paul Arkin, Chair
Professor Hana El-Samad
Professor Hernan Garcia

Fall 2018

Rational Design of Genetic Circuits and Biological Modules

Abstract

Rational Design of Genetic Circuits and Biological Modules

by

Mika Tei

Joint Doctor of Philosophy with University of California, San Francisco

in Bioengineering

University of California, Berkeley

Professor Adam Paul Arkin, Chair

Synthetic biology has exploited genetic engineering technology and systems theory to decipher the nature of interactions and apply the knowledge towards design and manufacture of new biological functions. Engineering biological systems with new functions has many promising applications in bioenergy, medicine, and natural computing. In this work, I designed and implemented various biological circuits in microbes, and evaluate functions of the systems both theoretically and experimentally through computational modeling and quantitative measurement of cellular dynamics. In the first chapter of this dissertation, I provide the comprehensive view of synthetic biology explored at three abstraction levels–at metabolite level, gene level, and community member level–and how those levels are interconnected to each other. In the second chapter, the design and implementation of a tunable delay-generating circuit for biocomputing purposes is discussed. Delay is an important building block in biological circuits to construct dynamic regulatory networks such as pulses, oscillations, and time-derivative calculations. At the end of the second chapter, the limitations of single cell engineering are raised, which motivates me to investigate on the resource constraints on the host processes and synthetic pathways in the third chapter. For bioenergy applications, it is desirable to divert energy spent on native or housekeeping gene processes to synthetic pathways to maximize product yield. In our published work, we show that the use of a sequence-dependent endoribonuclease, *mazF*, can funnel intracellular resources to a synthetic circuit and increase production of a high-valued

metabolite via the programmed degradation of non-circuit mRNA. In the fourth chapter, we analyze intercellular gene regulation in spatial configuration. To validate the theoretical framework, we engineer two strains of cells for a synergistic behavior, and elicit bistable contrasting pattern formation. Synthetic ecology, where cells work in cooperation to achieve more complex social behavior, is critical for advancing the capabilities of synthetic biology to implement an ecological function above the level of one organism. In the final chapter, I develop gene expression tools enabling the use of more diverse microbes. In this project, we create a registry of well-characterized gene regulatory elements for predictable heterologous gene expression in diverse bacterial hosts, including plant pathogens, plant-growth promoting, and bioremediation bacteria, using fluorescence-activated cell sorting and high-throughput sequencing. The large amount of sequencing data also enables statistical modeling of regulatory elements to design a pathway de novo with a set of desired gene expression levels. I believe that the findings described in this dissertation will be useful in advancing various applications of synthetic biology in agriculture, medicine, and bioenergy. For example, the resource allocation by *mazF* has further utility in bioenergy production when incorporated in polyclonal or differentiated population. One key challenge in translation of synthetic biology research to real world applications is the high mutation rates and escape mutants that typically result from engineering a large synthetic pathway in a single host. This is due to the selection pressure from high metabolic load. The findings described in this work will help inform engineering of microbial communities to carry out distributed cell behavior, so that each member produces part of the overall pathway and passage intermediates to neighboring members for complete production. This circumvents selection risk while still delivering high product yields. Another compelling application of synthetic biology is to optimize probiotic microbes for production of high-utility metabolites in plant symbiosis and human gut microbiota for applications in agriculture and healthcare. Expansion of gene expression tools and parts development in diverse microbes will accelerate exploitation of the probiotic strains to enhance their beneficial activities in the environment.

Acknowledgements

My work at the Arkin Lab has been an incredible journey of discovery and learning. This journey would have not been made possible without so much devotion from my advisor, mentors, colleagues, and mentees to progress my research.

Adam Arkin has been incredibly considerate and erudite supervisor with great generosity. Adam drives us to scrutinize our research while continuously motivating us with the big-picture goals. He has shown to us what it takes to be a great systems/synthetic biologist. Gwyneth Terry is a perspicacious administrator who is fun and dedicating and makes us able to reach Adam with great planning.

Murat Arcak, Justin Hsia, and Mindy Perkins from Arcak Lab in the Department of Electrical Engineering and Computer Sciences have been cooperative and friendly collaborators for the pattern formation project. Marcella Gomez from Arcak Lab also provided me plenty of technical advice in modeling and computational work. We had many inspiring conversations over coffee, which made the paper-writing process much more delightful.

My mentor, Ophelia Venturelli, was a young research fellow in Arkin Lab when I joined the lab as a rotation student in 2014. I was new to working with bacteria as all my previous research experience involved studying in mammalian cells. Although Ophelia herself was recently transitioning from yeast to bacteria research and had so many projects on her plate already, she dedicated her time to train me from scratch. Along with her high level of responsibility and solicitude for her lab mates, Ophelia's deep insight and rigor in research are what made her my role model in academia since the start of my Ph.D career.

To the Bioengineering cohorts–especially the 2013 classmates–and the lab mates in the Arkin Lab, I appreciate your kindness, collaboration, and support through the years. Special thanks to the people (Vivek Mutalik, Fangchao Song, Robert Egbert, Tom Ruegg, Ben Adler) whom I worked closely with on several projects. Kelly Wetmore, Hualan Liu, Jennifer Kuehl, Judy Savitskaya, Harneet Rishi, Sean Carim, and Kyle Sander have been very helpful with their advice on culturing, analyzing, and cloning different bacterial species.

My undergraduate and post-undergraduate mentees contributed significantly to the progress of our research. I greatly enjoyed working with Eugene Pang, Dylan McCormick, Lynn Dezhen Kong, and Poorvi Acharya.

Their unbiased and flexible view of the project deepened my comprehension in synthetic biology.

My family and close friends are invaluable in my graduate journey. My parents and my sister, who are professors in Materials Science, Physics, and a Ph.D. candidate in Chemistry, respectively, have supported and inspired me throughout the pursuit of Ph.D. and beyond. My roommates–Tammy Hsu, Stacey Lee, Minghui Zhang, Zhiwei Tay, Sean Nguyen, and Yanwei Lum–have been extremely helpful and encouraging during qualifying exam preparation and paper writing. Thank you for allowing me to share the joys of intellectual discovery and helped me through the most frustrating and challenging times in the Ph.D. journey.

Many thanks to Professor John Dueber to be my qualifying committee chair and giving me advice about the genetic timer.

"What I cannot create I do not understand."

Richard Feynman

Preface

Synthetic biology has grown tremendously since the first creation of biological circuits in the form of a genetic toggle switch (Gardner et al., 2000) and biological clock (Elowitz et al., 2000). The synthetic biology toolbox has been greatly expanded and complexity of circuit implementation has been advanced, to realize sequential logic computation and proportional-integral-derivative controller (Chevalier et al., 2018), .

Starting from as simple as programming single gene expression, the framework of systematic characterization and rational assembly of biological parts to perform a desired function has extended to larger scales such as genomic, intercellular, and ecological levels. Just as consumer electronics saw a huge boom after the invention and commercialization of the first integrated circuit in 1949, I believe that synthetic biology is at the doorstep of this important phase of developing a programmable and extensible platform to serve as "biological integrated circuit". I describe several projects within this dissertation that is a step towards this goal.

Progress in synthetic biology is rapid and it is a very exciting time to be in the field. We can look forward with anticipation to the developing of "integrated circuit" platforms for synthetic biology, where the high levels of reproducibility and reliability greatly ease translation to consumer applications. Just like the history of consumer electronics, biological integrated circuits will herald a new era of "consumer biology", where synthetic biology circuits will no longer be restricted to industrial bioreactors, but find their way to institutionalize the future of civilization.

Mika Tei
Berkeley, CA
Date, 2018

Contents

List of Figures

List of Tables

List of Abbreviations

DNA	**D**eoxyribo**N**ucleic **Acid**
RNA	**R**ibo **N**ucleic **A**cid
ATP	**A**denosine **T**ri **P**hosphate
RBS	**R**ibosome **B**inding **S**ite
HT	**H**igh **T**hroughput
HTS	**H**igh **T**hroughput **S**equencing
FACS	**F**luorescence **A**ctivated **C**ell **S**orting
LC	**L**iquid **C**hromatography
MS	**M**ass **S**pectrometry
FBA	**F**lux **B**alance **A**nalysis
MCA	**M**etabolic **C**ontrol **A**nalysis
EMRA	**E**nsemble **M**odeling for **R**obustness **A**nalysis
GRN	**G**ene **R**egulatory **N**etwork
ODE	**O**rdinary **D**ifferential **E**quation
PDE	**P**artial **D**ifferential **E**quation
IO	**I**nput-**O**utput
E. coli	*Escherichia **coli***
PCR	**P**olymerase **C**hain **R**eaction
CDI	**C**ontact-**D**ependent **I**nhibition
DLI	**D**iffusion-mediated **L**ateral **I**nhibition
1D	**1 D**imensional
2D	**2 D**imensional
PDMS	**P**oly**D**i**M**ethyl**S**iloxane
QS	**Q**uorum-**S**ensing
AHL	**A**cyl **H**omoserine **L**actone
HSL	**H**omo**S**erine **L**actone
AIP	**A**uto **I**nducing **P**eptide
AMP	**A**nti**M**icrobial **P**eptide
GPCR	**G** **P**rotein-**C**oupled **R**eceptor
GLV	**G**eneralized **L**otka **V**olterra
ECF	**E**xtra **C**ytoplasmic **F**unction
AS	**A**nti **S**igma

IPTG	**I**so**P**ropyl β-D-1-**T**hio**G**alactopyranoside
aTc	**a**nhydro**T**etra**C**ycline
Kan	**Kan**amycin
Amp	**Amp**icillin
Chlor	**Chlor**amphenicol
Strep	**Strep**tomycin

List of Symbols

Symbol	Name	Unit
t	time	s
δ	delay	s
N	noise in delay	s
$\|\delta\|$	range of delay	s
y_{off}	basal level of output	$\mathrm{mol\,L^{-1}}$
y_{ss}	steady-state level of output	$\mathrm{mol\,L^{-1}}$
K_D	dissociation constant	$\mathrm{mol\,L^{-1}}$
Δx	Channel length	m
l	Communication channel length	m
l_{res}	Efflux channel length	m
L	Laplacian matrix	

Chapter 1

Introduction

1.1 Introduction and Overview

Inside a cell is a small universe. Proteins, DNA, RNA, and metabolites form an intricate interacting network for survival. Synthetic biology has exploited genetic engineering technology and systems theory to decipher the nature of interactions and apply the knowledge towards design and manufacture of new biological functions.

A network consists of vertices/nodes and edges indicating interactions between nodes. An edge is the controlling agent of the attached node, and a node could be a metabolite, gene, or cell depending on the design level. The more molecular the process is, the more microscopic control we can exert. The biological or chemical study on natural systems such as functional genomics enabled identification and reverse engineering of the controller of biological components, thus contributing to the foundry of biological parts for synthetic biologists to optimize for and utilize in the designed networks. In the following sections, I discuss the progress in three design levels–chemical, genes, population – which were chosen from many possibilities specifically for their ease of representation by graph theory. I will also elaborate on how my dissertation fits in the vision to advance the field.

1.1.1 Design at Chemical Level

At the level where nodes are chemicals (metabolites), identification of an enzyme (edge) that catalyze the non-spontaneous chemical transformations is the first step to efficiently manufacture high-value chemicals. In-depth study of the case where "node = metabolite" or "edge = enzyme" is often referred to as metabolic engineering, and numerous useful discoveries have been made and translated into industrial applications. This process

is especially aided by the use of high throughput sequencing (HTS) for genomic and transcriptomic analysis to rapidly identify biosynthetic steps in an organism (J. Nielsen et al., 2016). Now, many high-value chemicals that were traditionally extracted from complex organisms, such as mammals and plants, can be now synthesized heterologously in simpler microorganisms, and further engineered by de novo pathway construction. For example, Ro et al. discovered a novel cytochrome P450 that performs a three-step oxidation of amorphadiene to arteminisinic acid in *Artemisia annua* using a comparative genomic analysis of plants in the Asteraceae family (Ro et al., 2006), and expressed the enzyme in transgenic yeast to produce arteminisinic acid by fermentation. Li et al. demonstrated de novo production of noscapine in yeast by reconstructing the biosynthetic network from diverse sources over 30 enzymes (Y. Li et al., 2018). In fact, the characterization of natural enzymes has been done so thoroughly that we are entering the new era of designing de novo enzymes for desired metabolites (Marcos et al., 2018; Y. P. Lim et al., 2018).

To analyze the network consisting metabolites and enzymes from the perspective of systems theory, several models have been used for varied applications from steady-state analysis to solving stochastic chemical dynamics (P.-W. Chen et al., 2017). Flux balance analysis (FBA) is a steady-state model that is often used to optimize fluxes (edges) in a pathway for a given objective such as maximizing ATP production. FBA uses known stoichiometry of reactions as the main constraints as well as the lower and upper boundaries of fluxes. Synthetic biologists can control the flux by genetic parts such as promoter libraries (Mutalik et al., 2013), inducible promoters (Ruegg et al., 2018), and ribosome binding site (RBS) libraries (Salis et al., 2009) to modify the enzyme concentration in the cell, or add external processes to mitigate inhibitory effects (Xu et al., 2017). Due to its simplicity, FBA is convenient to sketch the entire metabolic network of the host organism for the basic node and edge characterization. For metabolic modeling approaches to engineering, various tools have been developed such as those from the Maranas research group for cutting and amplifying pathway branches and dealing with co-factor reuse (Burgard et al., 2003; Pharkya et al., 2004).

1.1.2 Design at Genetic Level

Now, we take a step back and look at the network in a different design level – the genetic level. For cell survival, enzymes and other proteinous

cell components need to be expressed within an appropriate range at all time (M. Scott et al., 2010) for "survival" levels of component activity to be met. Thus, the gene regulatory network (GRN) becomes crucial for the sustainable cell replication and adaptation to the dynamically changing environment (Ma et al., 2009). Using a graph theory representation, the edges are then gene regulators such as transcription factors. The nodes here are genes or DNA sequences. The graph can contain multiple transcriptional cascades, feedback, and feed-forward loops, which complicates the analysis and exploitation of the functional modules.

A transcription-factor-edge is typically characterized by a response function, or an input-output curve, and expressed mathematically with a Hill function. With the vast availability of output measurement technology including reporter expression systems such as *sfGFP*, liquid chromatography-mass spectrometry (LC-MS), antibody labeling, microfluidics, microtiter plate assay, and HT flow cytometry, this area of synthetic biology has attracted much attention and made tremendous advancement in the past two decades. Some researchers have focused on the parts generation and characterization, where the generated parts could belong to a novel strategy such as optogenetics (Olson et al., 2017) or expansion of the existing library (Stanton et al., 2014; Chappell et al., 2017). Other researchers have focused on the construction of more complex circuits (Hu et al., 2018; Andrews et al., 2018). As the assembly of genetic parts with defined response functions closely resembles the assembly of logic gates in digital electronic circuits, these man-made intracellular networks have been named synthetic biological circuits, or genetic circuits.

The complexity of a genetic circuit may be increased via feedback loops or assembly of parts. In graphical representation, addition of a feedback loop translates to an extra edge, and addition of a part is equivalent to an extra node connected by minimum number of edges. Assembly of simple parts such as logic gates have been intensively explored by Chris Voigt's group to construct combinatorial logics (A. A. K. Nielsen et al., 2016). Feedback loops usually incorporate dynamics into circuits, thus requiring time-related analyses such as stability in addition to the standard input-output characterization for circuit function evaluation. Higher-order circuits require combination of both, which can be considered as assembly of modules (H. Zhang et al., 2014). High-order circuit construction benefits from well-characterized modules since the overall circuit behavior can be predicted from small subsystems of a circuit. In many higher-order circuits applications, modularity of the subsystems can be taken for granted, but in

some cases, we have to be cautious about the compatibility of the modules as these biological modules have context-dependent shortcomings such as delays and host-cell resource consumption upon gene expression (Del Vecchio, 2015). As the number of genetic parts used in a circuit increases, the risk of unexpected interactions between circuit-host and among modules increases. To this end, the subjects of context-dependency and resource allocation are attracting much attention (Yeung et al., 2017). Some ensure modularity by insulators (Gorochowski et al., 2017; Mishra et al., 2014), incorporation of host processes into the design (Weiße et al., 2015; Ceroni et al., 2018), or even recreate the central dogmas (replication, transcription and translation) outside cells to exert tighter control on the context (Tayar et al., 2017; Halleran et al., 2018; Jaroentomeechai et al., 2018). One increasingly popular way to resolve the risk is to differentiate cells with specific subtasks, just like how natural living systems evolved from unicellular to multicellular systems to perform more complex tasks. Compartmentalization of the modules into different members allows distribution of production load, reuse of parts, and better utilization of space and growth. (see Chapter 3 on using MazF for modulating synthetic circuit resource allocation)

1.1.3 Design at the Population Level

Interaction among individual cells (compartmentalized sets of genes) is referred to as population networks. The nodes consist of cells or species, and the edges are composed of communication channels which can take many variable forms. Even among non-communicating microbes, competition for physical space (Mousa et al., 2016; Hwang et al., 2018) and nutrition (Momose et al., 2008) are considered as important interactions observed in natural ecological system. The physical barrier is, for example, utilized in artificial symmetry-breaking (Nuñez et al., 2017). The molecular basis of bacterial communications is crucial for chemical therapeutics discovery (Fischbach, 2018) and modular assembly of cell-types as chemicals may provide orthogonal communication channels due to difference in chemical structures and specificity of sensory proteins. To this end, acyl homoserine lactone (AHL), a class of diffusible molecules used in Gram negative quorum-sensing (QS) systems, is a favored choice of signaling molecules due to its programmability (Y. Chen et al., 2015; S. R. Scott et al., 2016; Grant et al., 2016). Other modes of communication that have been discovered in natural systems include autoinducing-peptides (AIPs) (Piewngam et al., 2018), antimicrobial peptides (AMPs), metabolites, contact-dependent signaling

(E. C. Garcia et al., 2016; D. Chen, 2014), and G protein-coupled receptors (GPCRs) (Ostrov et al., 2017). Although the modularity aspect needs to be improved for broader applications, these channels have been used in engineered cells for synthetic applications (Kong et al., 2017), showing their strong potential to be utilized in future synthetic consortia research.

Recently, commercial applications of above-mentioned consortia have increased awareness that microbial communities have much room for engineering interventions to implement new ecological functions. For example, BioConsortia (Davis, California) has one of the world's largest pre-screened collections of microbes at over 45000 and uses directed selection to discover beneficial teams of microbes to improve agricultural crop performance. The transition from single cell engineering to multicellular engineering started as a hierarchical design of distributing loads among different clones of *E. coli*, our model organism (Brenner et al., 2007); however, as our ability to observe the microbial communities has exponentially improved, it has become increasingly apparent that the choices of microbes that compose microbiota are crucial to the outcome in the target environment. To this end, genetic manipulation frameworks have been developed for many ecologically important microbes (Shepherd et al., 2018; Brophy et al., 2018) identified in observational studies (Thompson et al., 2017; Banerjee et al., 2018). The knowledge of synthetic biology does not only apply to the de novo construction of multicellular circuits, but also can be utilized in designing and analyzing an assortment of unmodified microbes inspired from natural probiotic consortia. Systematic characterization of a community member per communication channel (multiple orthogonal chemicals act as separate communication channels) can be done similarly to genetic parts, as long as the input(s) and the output of the communication channel are identified. This field of synthetic biology is still nascent and demonstration of applications is limited to well-studied fields such as biofuel production (Poszytek et al., 2016). Yet, early systematic and quantitative studies on microbial interactions and ecological outcomes strongly suggest that rational assembly of functional community is possible, and can be utilized in a broad range of applications from agriculture (Herrera Paredes et al., 2018) to human gut (Ophelia S Venturelli et al., 2018). This new area of *synthetic ecology* is very exciting and promising. The multicellular networks with additional dimensions to consider are much more complicated than intracellular GRN, but their analysis would unravel higher-order design principles of functional organization. The outcome of the network can be more impactful

than the single cellular engineering case, as structured and diverse communities have been shown to be more efficient and resilient than single-strain populations (Boetius et al., 2000; Volke et al., 2018; Hays et al., 2017).

1.1.4 Interconnection between Design Levels

Design Levels One and Two –the metabolite and gene regulatory networks– both occur at intracellular level. Many implementations of intracellular circuits mix some components of gene regulations and some enzymatic pathways, which causes confusion in mechanistic understanding of the network as the design levels of the components are different. Dissection of gene regulatory network into biophysical processes is possible with careful substrate incorporation (H. G. Garcia et al., 2011; Phillips, 2017). For perfectly modular processes, we can analyze gene regulations and enzymatic pathways independently. For other cases, we can also interconnect different design levels via ad hoc functions. For example, the gene response function $y = f(\alpha, x)$ can have a metabolite-dependent parameter $\alpha = h(a)$, where a is the intracellular metabolite concentration that affects the gene y expression. The obvious limitation of these ad hoc functions is the empiricality of the analysis due to the lack of biophysical mechanism. Whether you analyze your system at the chemical or genetic design level depends on your objectives. Biophysical model is preferred for extensible analysis and genetic level for application-specific analysis.

Between Design Levels Two and Three, the well-characterized QS systems have been conveniently used to connect intercellular GRNs (Basu et al., 2005; Tamsir et al., 2011; Y. Chen et al., 2015). Analysis of intercellular network introduces new dimensions to analyze GRN–cell density and cell arrangement. The effects of cell density and growth competition can be implemented using enzyme-substrate kinetics for nutrient (Pai et al., 2009) or Generalized Lotka-Volterra (GLV) model (Ophelia S. Venturelli et al., 2017). The effects of cell arrangement and community structure can be incorporated by either full or derived models of partial differential equations (PDEs).

Finally, global platforms such as KBase (Arkin et al., 2018) would aid designing synthetic network at any design level as they encompass data and analysis tools at scales raging from the biomolecular to the ecological.

1.2 Rationale of this Dissertation

Engineering biological systems using synthetic genetic circuits has many promising applications in bioenergy, medicine, and natural computing. For my Ph.D., I designed and implemented various biological circuits in microbes, and evaluated functions of the systems both theoretically and experimentally through computational modeling and quantitative measurement of cellular dynamics.

First, I have engineered a tunable delay-generating circuit for biocomputing purposes. Delay is an important building block to construct dynamic regulatory networks such as pulses and oscillations which find utility in cellular timers and delivering controlled doses of synthesized materials. Delays can also can be applied for time derivative calculation which is useful for gradient sensing, such as in bacterial chemotaxis. Notably, it can be helpful to delay production of a valuable synthetic product until there are enough cells in a population to redirect resource allocation to maximize production without sacrificing yield.

This links into the next project where we look into redirection and optimization of cellular resource allocation (doi:10.1038/ncomms15128) towards bioenergy applications. Synthetic circuits embedded in host-cells compete with cellular processes for limited intracellular resources. We showed that funneling of cellular resources to a synthetic circuit, after global transcriptome degradation by the sequence-dependent endoribonuclease MazF, can increase production of a high-valued metabolite.

These technologies, especially the resource allocation strategy with MazF, are readily available to test out in the commercial processes. Beyond this thesis, one of my future research goals is to optimize MazF expression in industrial strains to improve the yield of high-valued chemicals produced in current processes, given the characteristics of MazF to halt host-cell biomass synthesis and increase production from recombinant genes. Further, I aim to apply the model-guided framework of engineering single cell gene circuit to design synthetic ecology, where cells work in cooperation to achieve more complex social behavior.

As a first step towards my future goals, I demonstrate in this thesis synthetic pattern formation using quorum sensing-based lateral inhibition in an engineered microbial community. Pattern formation is fundamental to developmental processes in complex systems, and cell fates are determined by both their memory and cues from the environment during embryo. Lateral inhibition, which occurs when spatially adjacent units inhibit each other,

facilitates diverging cell fates among neighbors in multicellular organisms. Mathematical and biological implementation of lateral inhibition system in simple microbial communities elucidates the theoretical and experimental design principles of diverging pattern formation and cell fate decisions.

To expand the diversity of the members of possible synthetic communities, better gene expression tools are required to engineer microbes outside model organisms. For my final part of my thesis (Chapter 5), I am developing well-characterized gene regulatory elements for predictable heterologous gene expression in diverse bacterial hosts, including plant pathogens, plant-growth promoting, and bioremediation bacteria. Utilizing fluorescence-activated cell sorting (FACS) and HTS, I aim to provide extensive registry of gene regulatory elements with gradient levels of target gene expression, which is currently only available for limited organisms such as E. coli. Through generation of the registry and statistical modeling on the high-throughput sequencing data, I also intend to uncover nature's design principles for gene expression in host-cell processes.

Overall, I am passionate about engineering synthetic microbial communities in diverse applications in agriculture, medicine, and bioenergy. I believe the resource allocation tool has further utility in bioenergy production when incorporated in polyclonal or differentiated population. Engineering a large synthetic pathway in a single host has risks of high mutation rates and escape mutants due to technical difficulty of gene-editing and selection pressure from high metabolic load. The differentiated cells in a population each can have a part of overall biofuel production pathway, and passage the intermediates to neighboring cells for complete production. Similarly, I would like to optimize probiotic microbes for production of high-utility metabolites in plant symbiosis and human gut microbiota for applications in agriculture and health. I believe my expertise in these different projects gives me a unique perspective that is the fine integration of assorted experiences to meet the current and future trend in modern science and to generate great discoveries in synthetic biology.

Part I

Single Cell Engineering

Chapter 2

Design and optimization of a programmable genetic timer

2.1 Introduction

A genetic timer is a module that controls the delay-to-onset of the target gene expression upon receiving the input signal. Due to its controllable delay, it is a fundamental building block that can be used to construct and tune dynamic regulatory circuits. For instance, oscillatory behavior can be generated by combining a constant delay, set by the timer with a negative feedback loop, which works by modifying the stability of the equilibrium. Another example of this behavior can be seen when a pulse whose interval equals to the duration of the delay can be generated when a delay is combined with AND and NOT gates with step-input (Mangan et al., 2003).

Other than being an important component in dynamic circuits, pure delays themselves also have applications in synthetic biology. Temporal regulation of gene expression resolves tasks such as organization of hierarchical structures and separation of incompatible processes. For instance, delayed self-destruction of genetically modified strains will be useful for biocontainment purposes (Chan et al., 2016). The genetic timer can be optimized to allow the cells just enough time to complete its work and self-destruct subsequently, minimizing risk of harmful mutations or over-proliferation. Controlled delays enable sensing-independent dynamic regulation, thus the usage can be more generic compared to sensing-dependent system such as transcriptionally regulated promoters that respond to specific metabolites (Dahl et al., 2013).

Similarly, in nature, genetic timers have been evolved to benefit the species, as exemplified in phage infection of bacteria. In the lambda phage, a combination of sensing-dependent and sensing-independent timers is used to transition from a lysogenic state to a lytic state, and then to a final

host cell lysis. Lambda phage has a mechanism to detect the host cell DNA damage and switches from the lysogenic to the lytic state, via a transcription factor cI; once the phage transitions into the lytic state, an autonomous timer system delays the timing of host cell lysis, which involves the expression of lysis-inducing protein called holin (S105) and its antagonist (S107) (Oppenheim et al., 2005; Chang et al., 1995). S107 dimerizes with S105 and inhibits the formation of S105 oligomers that makes "holes" on bacterial cell membrane for lysis. Why do natural systems employ autonomous timers in some cases and sensing systems in the other cases? Engineering a synthetic timer will provide insight into the advantages and limitations that distinguish these two classes of timing mechanisms and why nature has selected a particular circuit for controlling the temporal response of gene expression.

The previous efforts in synthetic biology community to implement a timer include lengthening the transcriptional cascade (Hooshangi et al., 2005), tuning the mutually inhibitory network to monostable regime (Ellis et al., 2009), and tuning the AHL accumulation rate of recombinant quorum sensing systems (Gupta et al., 2017). However, for the cases of transcriptional cascade and mutually inhibitory network, the duration of delay is fixed for a given circuit motif. Only for the quorum sensing-based timer is it possible to replace the promoter of AHL synthase with an inducible one to control AHL accumulation rate; however, this implementation is dependent on increasing cell density to generate a delay, which may not be applicable for some dynamic regulatory system such as cells in chemostat cultures that maintain the same density.

As exemplified in monostable mutually inhibitory switch, a few networks can generate delays where duration of the delay can be tuned by parameters. Rather than optimizing a specific timer network, theoretical exploration of the full space of possible networks enables identification of the solution space, which would provide insight in the mathematical principles of genetic timer design (W. A. Lim et al., 2013).

In this chapter, we demonstrate the optimization of genetic timer network by enumerating circuit motifs and computationally simulating with biologically reasonable parameter sets. The systematic search for the controllable timer revealed three types of delay generating mechanisms, and we identified sequestration-based timer to generate the most controllable and extensible delays out of the entire network motifs investigated computationally. Then we implemented the sequestration-based timer using orthogonal σ and anti-σ factors in *E. coli* and demonstrated that the onset

timing of the target gene expression can be regulated using chemical inducers.

2.2 Materials and Methods

2.2.1 ODE simulation

The simulation of time-responses was coded in MATLAB. The details of model development are discussed in 2.3. The units were normalized so that steady-state expression becomes 1 for the reporter and 10 K_D for the intermediate genes. Protein dilution rate was set to be a typical *E. coli* growth rate in 37 degree Celsius, which is around half an hour. Input variable was set to be 100 logarithmically spaced points ranged over five orders of magnitude.

2.2.2 Strain and plasmid construction

E. coli strain DH10β (NEB) was used for cloning. PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (Thermo) and oligonucleotides (IDT). BsaI (NEB) and T7 DNA ligase (NEB) were used to construct plasmids using parts obtained from the MIT Registry of Standard Biological Parts, JBEI registry T. Lee et al., 2011, or synthesized gBlocks (IDT). The list of strains and plasmids can be found in Table 2.1.

TABLE 2.1: Plasmids used for the sequestration-based timer circuits

Name	Resistance	Origin	Description
pN565	Strep	$oriV$	Attenuated T7 RNA polymerase, IPTG inducible (Rhodius et al., 2013)
pVRa03_1198	Amp	$ColE1$	ECF3 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa11_987	Amp	$ColE1$	ECF11 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa14_1324	Amp	$ColE1$	ECF14 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa15_436	Amp	$ColE1$	ECF15 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa16_3622	Amp	$ColE1$	ECF16 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa20_992	Amp	$ColE1$	ECF20 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa22_4450	Amp	$ColE1$	ECF22 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa27_4265	Amp	$ColE1$	ECF22 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa33_423	Amp	$ColE1$	ECF33 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRa38_1322	Amp	$ColE1$	ECF38 under T7 promoter, IPTG inducible (Rhodius et al., 2013)
pVRb03_up1198	Kan	$pSC101$	GFP under ECF3 promoter (Rhodius et al., 2013)
pVRb11_3726	Kan	$pSC101$	GFP under ECF11 promoter (Rhodius et al., 2013)
pVRb26_up601	Kan	$pSC101$	GFP under ECF14 promoter (Rhodius et al., 2013)
pVRb15_up436	Kan	$pSC101$	GFP under ECF15 promoter (Rhodius et al., 2013)
pVRb16_3622	Kan	$pSC101$	GFP under ECF16 promoter (Rhodius et al., 2013)
pVRb20_992	Kan	$pSC101$	GFP under ECF20 promoter (Rhodius et al., 2013)
pVRb22_up1147	Kan	$pSC101$	GFP under ECF22 promoter (Rhodius et al., 2013)

pVRb33_375	Kan	$pSC101$	GFP under ECF33 promoter (Rhodius et al., 2013)
pVRb38_up1322	Kan	$pSC101$	GFP under ECF38 promoter (Rhodius et al., 2013)
pVRc03_1198	Chlor	$p15A$	AS3 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc11_987	Chlor	$p15A$	AS11 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc14_1324	Chlor	$p15A$	AS14 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc15_436	Chlor	$p15A$	AS15 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc16_3622	Chlor	$p15A$	AS16 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc20_992	Chlor	$p15A$	AS20 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc22_4450	Chlor	$p15A$	AS22 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc27_4265	Chlor	$p15A$	AS27 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc33_423	Chlor	$p15A$	AS33 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pVRc38_1322	Chlor	$p15A$	AS38 under $plux$ promoter, AHL inducible (Rhodius et al., 2013)
pMT9_03	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF03; $tetR$; $pLtetO$ →AS03
pMT9_11	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF11; $tetR$; $pLtetO$ →AS11
pMT9_14	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF14; $tetR$; $pLtetO$ →AS14
pMT9_20	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF20; $tetR$; $pLtetO$ →AS20
pMT9_22	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF22; $tetR$; $pLtetO$ →AS22
pMT9_33	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF33; $tetR$; $pLtetO$ →AS33
pMT9_38	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF38; $tetR$; $pLtetO$ →AS38

pMT13c_11	Chlor	$p15A$	$lacI(W220F)$; $pLlacO$ →ECF11; $tetR$; $pLtetO/phlFO$ →AS11
pMT15g_11	Kan	$pSC101$	GFP and PhlF under ECF11 promoter

TABLE 2.2: Strains used for the sequestration-based timer circuits

Name	Background	Plasmids
OV-52	BW27783	pVRa03_1198+pVRb03_up1198+pVRc03_1198+pN565
OV-53	BW27783	pVRa11_987+pVRb11_3726+pVRc11_987+pN565
OV-54	BW27783	pVRa16_3622+pVRb16_3622+pVRc16_3622+pN565
OV-55	BW27783	pVRa22_4450+pVRb22_up1147+pVRc22_4450+pN565
OV-56	BW27783	pVRa27_4265+pVRb31_34+pVRc27_4265+pN565
OV-66	BW27783	pVRa20_992+pVRb20_992+pVRc20_992+pN565
OV-67	BW27783	pVRa14_1324+pVRb26_up601+pVRc14_1324+pN565
OV-68	BW27783	pVRa33_423+pVRb33_375+pVRc33_423+pN565
OV-69	BW27783	pVRa38_1322+pVRb38_1322+pVRc38_1322+pN565
MT264	BW27783	pMT9_03+pVRb03_up1198
MT265	BW27783	pMT9_11+pVRb11_3726
MT266	BW27783	pMT9_14+pVRb26_up601
MT267	BW27783	pMT9_20+pVRb20_992
MT268	BW27783	pMT9_22+pVRb22_up4450
MT269	BW27783	pMT9_27+pVRb31_34
MT270	BW27783	pMT9_33+pVRb33_423
MT271	BW27783	pMT9_38+pVRb38_1322
MT375	BW27783	pMT13c_11+pMT15g_11

2.2.3 Plate reader assay

Overnight cultures of cells in LB were diluted to fresh LB culture to 1:100 ratio and were either pre-induced by 100 ng/mL aTc or 0 for 8 hours in the incubated shaker at 37 °C. Cells were washed three times and diluted to OD600 of 0.025 in a 96 well flat clear bottom black polystyrene microplate (Corning) containing 192 μL LB and appropriate concentrations of aTc (Sigma-Aldrich) dissolved in 4 μL of ethanol for and IPTG (Sigma-Aldrich) in 4 μL water. Synergy 2 (Biotek Instruments) was used to measure cell density (OD600) and fluorescence of growing culture every 8 min for 18 hours at 37 °C. The BioTek excitation and emission wavelengths were 485 nm, 528 ± 20 nm for sfGFP and 560 nm, 620 ± 20 nm for mRFP1.

2.3 Results

2.3.1 A long and tunable delay is the hallmark of an ideal genetic timer

First, we define the criteria for good genetic timer to facilitate quantitative assessment of each circuit topology for delay-generating performance. Figure 1 represents an example expression of a target gene over time in response to a step input signal at time $t = 0$. A timer works to alter the duration in off state of the target gene expression after receiving the step input.

The measurable criteria for a good timer are defined as long and tunable delay (δ), tight control on off state during the delay (y_{off}), tunable gain (y_{on}), fast rise time ($\frac{dy}{dt}$), small noise ratio in delay ($\frac{N}{\delta}$), and minimal energetic burden (Fig. 2.1). Multi-criteria optimization can be achieved by defining the weight for each of these criteria and obtain a scalar objective function. For our search, we focused on the length and controllability of δ, as the delay generation activity is the core of a timer circuit, while the rest of criteria can be optimized subsequently.

In order to compare multiple motifs, we defined and fixed the input for the delay to be an inducible promoter. Unlike the prior demonstrations of timer by the depth of cascade or different constitutive promoter pairs (Hooshangi et al., 2005; Ellis et al., 2009), inducible promoter provides dynamic tuning of the delay without changing the circuit configuration. Hence, the input of every circuit is defined to be the fixed range of inducer concentrations, and the output to be measured is the duration for the target gene expression to be in the off state. The objective functions to be optimized are both the range and tunability of the delay. The range $\|\delta\|$ can be expressed as $\|\delta\| \equiv \delta_{max} - \delta_{min}$ as the difference between the maximal and minimal delays δ_{max} and δ_{min}. Tunability of delay with respect to input inducer concentration can be assessed by a measure of input sensitivity, such as the the maximal slope $\frac{d\delta}{dx}$, where x is the input inducer variable. The objectives are to satisfy both large $\|\delta\|$ and small $\max \frac{d\delta}{dx}$ at the same time.

2.3.2 Three-nodes-motif spans the feasible basis of regulatory networks

Complex cellular circuits might be abstracted into simpler core networks (W. A. Lim et al., 2013). A complex network could potentially be composed

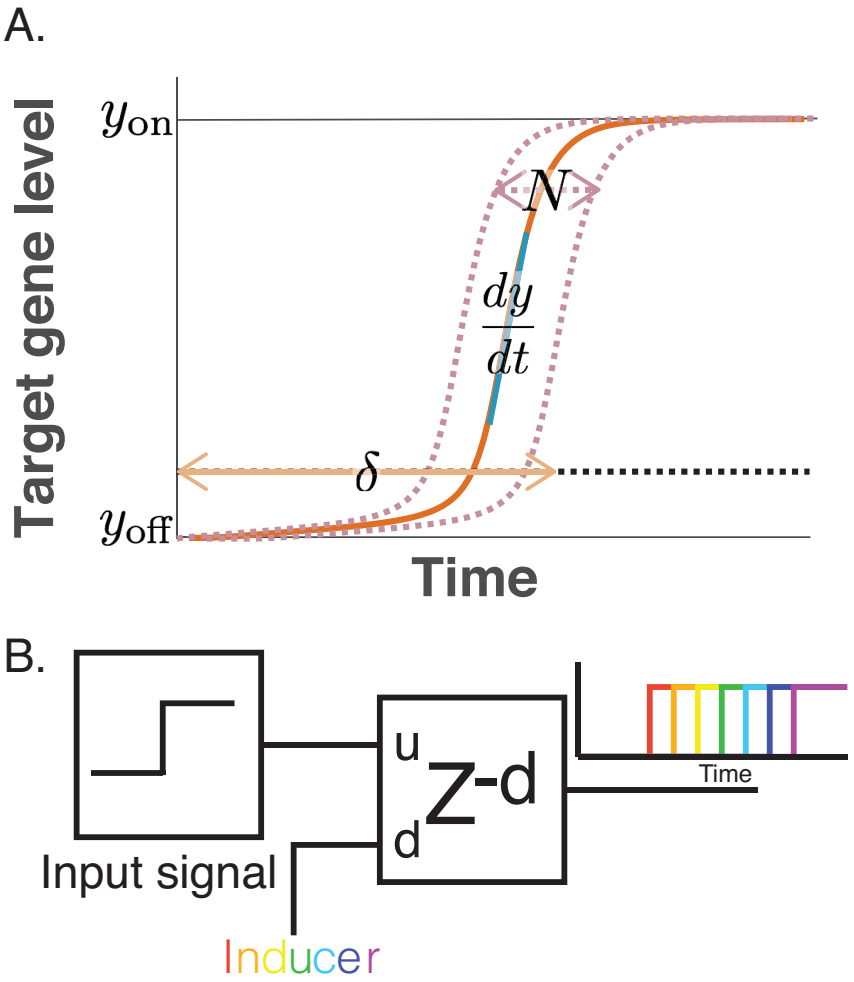

FIGURE 2.1: **Definition of a timer module described in a time-response curve of a target gene expression and a block diagram. (A)** The measurable criteria for a good timer depicted on an example time-response curve. The target gene expression delayed by δ. **(B)** A timer module in a block diagram. The part takes two inputs; one input (u) is connected to an ON/OFF switch and the other (d) is to a variable concentration of external inducer. The inducer level controls the delay, which is represented by different colors in the time-response of the output gene expression.

of several sub-network modules, each with a simpler core function. For our computational search, we limit our design space to be motifs that can be expressed less than three regulators. Network enumeration provides a solution close to global optimum. We enumerated a broad class of circuit motifs and modeled with ordinary differential equations (ODEs). Dynamics

of the target gene expression is simulated using biologically realistic values as model parameters to elucidate the mechanism of the delay (Table 2.3). We then evaluated the timer function based on the aforementioned criterion (programmability and length of the delay).

In detail, for our design space, we limit the network nodes to be regulators, which can be implementable by a synthetic genetic circuit, and the number of regulators to be in the range of one to three. The synthetic circuit should be orthogonal to host physiology to be extensible for applications that require host-independent processes. We focus on well-characterized parts to be the choice of regulators, including transcriptional factors, small RNAs, and activator/sequestrator pairs.

For the parameters, we assume the host has the same protein dilution rate and the same mRNA degradation rate regardless for any gene. Furthermore, to remove part-specific bias during comparison of network motifs, we assumed all the regulators to have the same steady-state concentrations, and all the transcriptional factors to have the same Hill coefficients and dissociation constant of K_D. The off state, thus the length of the delay, is defined to be the time between the point of induction of the circuit and the point where the reporter gene expression exceed $\frac{K_D}{10}$ of the steady-state value.

The transcriptional activation and repression are described by a Hill function as follows:

$$\frac{d[m_x]}{dt} = \alpha_m \frac{[p_A]^n}{[p_A]^n + K_D^n} - \beta_m [m_x] \tag{2.1}$$

for activation where m_x is the mRNA of gene x regulated by a transcriptional activator p_A, α_m is the maximal transcription rate, K_D is the dissociation constant of p_A to bind to the operator of the regulon, n is the apparent Hill coefficient, and β_m is the mRNA degradation rate, and

$$\frac{d[m_x]}{dt} = \alpha_m \frac{1}{[p_R]^n + K_D^n} - \beta_m [m_x] \tag{2.2}$$

for transcriptional repressor p_R. A linear ordinary differential equation (ODE) is used for protein concentration

$$\frac{d[p_x]}{dt} = \alpha_p [m_x] - \beta_p [p_x], \tag{2.3}$$

TABLE 2.3: Biologically acceptable values of parameters chosen for ODE simulations

Parameter	Value
Maximum transcription rate (α_m)	$10(\beta_m + \beta_p)$
mRNA degradation rate (β_m)	0.25 min^{-1}
Translation rate (α_p)	β_p
Protein dilution rate (β_p)	0.025 min^{-1}
Dissociation constant (K_D)	1 nM
Hill coefficient (n)	2
Diffusion-limited rate constant (k_{on})	100 nM^{-1} min^{-1}
Reverse rate constant (k_{off})	1 min^{-1}

where α_p and β_p are the translation and dilution rate of protein p_x, respectively. Mass action kinetics are used for the case where concentrations of two molecules binding and unbinding are known, such as for sequestration and induction processes:

$$\begin{cases} \frac{d[p_A]}{dt} & = \alpha_p[m_A] - k_{\text{on}}[p_A][p_B] + k_{\text{off}}[p_A p_B] - \beta_p[p_A], \\ \frac{d[p_B]}{dt} & = \alpha_p[m_B] - k_{\text{on}}[p_A][p_B] + k_{\text{off}}[p_A p_B] - \beta_p[p_B], \\ \frac{d[p_A p_B]}{dt} & = k_{\text{on}}[p_A][p_B] - k_{\text{off}}[p_A p_B] - \beta_p[p_A p_B]. \end{cases} \tag{2.4}$$

2.3.3 Computational search reveals three mechanisms of tunable delay generation

ODE simulations of circuit motifs consisting one to three regulators showed two types of timers based on the intermediate regulator behaviors. One is count-up timer, where the activating regulator of the target gene needs to accumulate above certain threshold, and the other is countdown timer, where the repressive regulator of the target gene needs to decrease below certain threshold (Levine et al., 2014).

Motifs of two nodes with activator, and three nodes with double activators or double repressors cascade lead to overall activation of the target gene expression with increasing inducer concentrations. Here, the external inducer controls the rate of accumulation of the molecule in charge

of the activation module, and the delay is generated as the threshold concentration divided by the rate of accumulation of the regulator (Fig. 2.2**A**).

Motifs with a repressor regulating the target gene require pre-induction of the repressors. Unlike the count-up timers where the delay was controlled by the rate, since we assumed all conditions to have the same dilution and degradation rates of proteins or RNAs across all motifs, the delay of this countdown timer is controlled by the concentration of pre-induced repressor (Fig. 2.2**B**). Sequestration-based timers fit both definitions of count-up and countdown timers. The delay generating mechanism is similar to the repressor case where the concentrations of pre-induced sequestration molecules are varied; hence the delay can be generated independent of the rate of activator accumulation rate. Here, the concentration of free sequestration molecules decreases upon accumulation and binding of activator molecules, resembling to the case of diluting repressor (Fig. 2.2**B, C**).

The feedback loop of nodes creates a toggle-switch-like timer. The pre-induction of the first repressor switches the timer to one state. The variable induction of the second repressor turns off the expression of the first repressor as well as directly induces the target gene expression. The delay is due to compounded effects of the second repressor accumulation to repress transcription of first repressor and dilution of the first repressor when induction level is low, and essentially no delay when induction level is high as excess inducer completely inactivates the first repressor that regulates the target gene directly (Fig. 2.2 **D**).

2.3.4 IO characterization enables circuit motif discovery with optimal range and tunability of the delay

The computer modeling and simulation result of each circuit motif is shown in Fig. 2.3. Input-output (IO) responses were characterized by recording the delay δ at given input inducer concentration ranging across five orders of magnitude and the fixed parameter set listed in Table 2.3). The performance of various circuit motifs is evaluated by $\|\delta\|$ (range) and $\frac{d\delta}{dx}$ (tunability) (Fig. 2.3).

The countdown timers based on sequestration and repressor-dilution, as well as simple activation exhibited similar range and tunability of the delay. The bistable activator had the broadest range of delay, but was susceptible

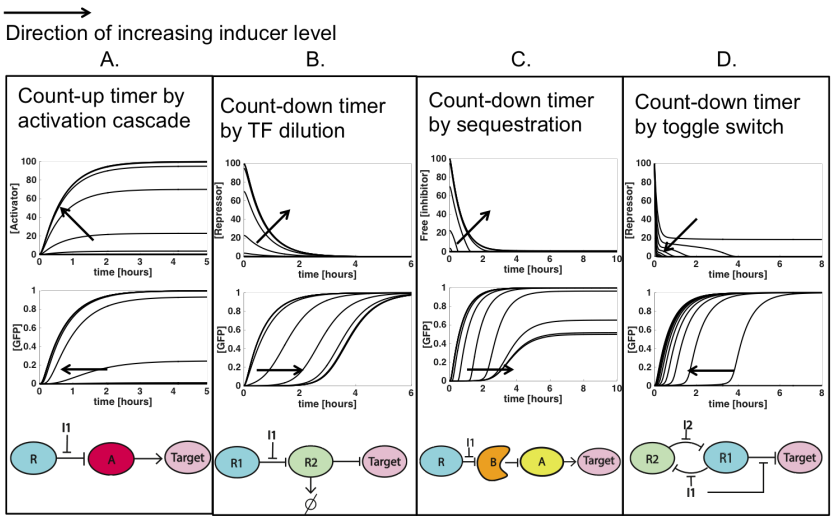

Figure 2.2: Four mechanisms of delay generation identified from circuit topology enumeration. (A) Count-up timer by activator accumulation. **(B)** Count-down timer by repressor dilution. **(C)** Sequestration-based timer. Count-down timer of free sequestration molecule. **(D)** Count-down timer with feedback regulation.

to input variation. Long transcriptional cascades were also identified with noise propagation property (Hooshangi et al., 2005). Between the sequestration and repressor-dilution based timers, the repressor-dilution based timer has advantage over sequestration for the tunable gain y_{on}, but is sensitive to stochasticity due to low repressor concentration toward the end of delay. Protein-protein sequestration timer is most promising architecture of delay generation not only because it has best range and tunability of delay, but also it has additional tuning knob of activator accumulation rate besides the sequestration molecule induction for delay generation.

We then further tested the delay generation of sequestration systems with different parameter ranges. For the original simulation, the dissociation constant between the sequestration molecule and the activator is one order of magnitude lower than the dissociation constant between the activator to the operator. The higher affinity of the sequestration molecule to the activator improves the maximum delay to nearly 8 hours, with the trade-off with the steady state target gene expression (Fig. 2.4**A**). With an additional regulator, we can incorporate a positive feedback loop on the target gene to control y_{on} independently from sequestrator induction level (Fig. 2.4**B**). Although the maximal attainable delay appear shorter as the threshold concentration of $\frac{y_{\text{on}}}{10}$ is reached at earlier stage of target gene expression,

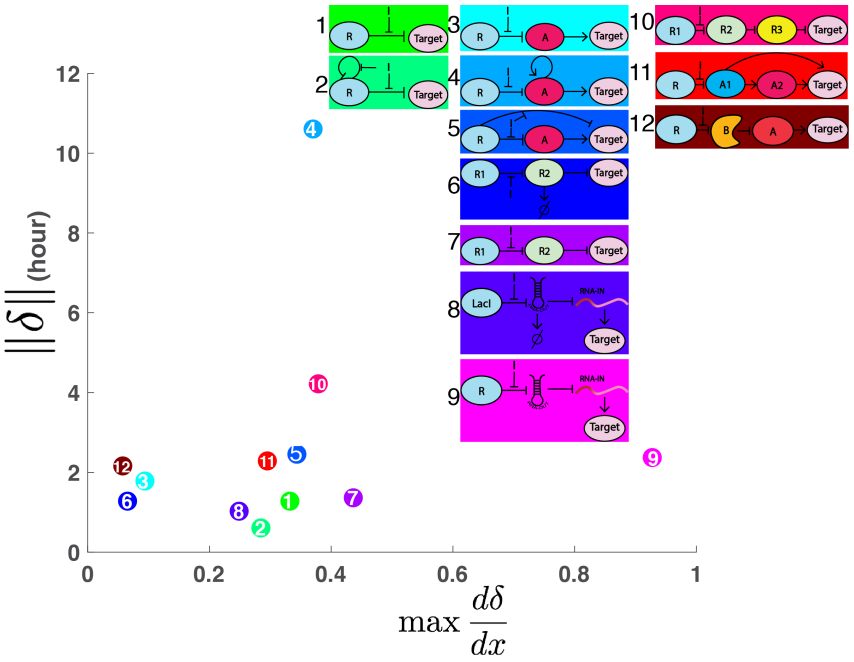

Figure 2.3: The delay performance evaluation plot. The delay performance of circuit motifs listed 1 to 12 are simulated and plotted for maximal slope $\frac{d\delta}{dx}$ in the horizontal axis and range of delay $\|\delta\|$ in the vertical axis.

the delay generated by sequestration remains unchanged by incorporation of the positive feedback loop.

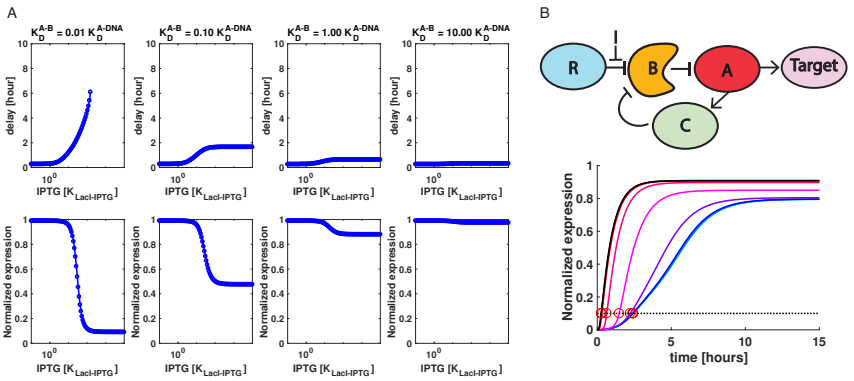

Figure 2.4: Optimization of the sequestration-based timer. (A) Sequestration-based timers evaluated with varying dissociation constants between the activator and sequestrator. The rest of parameters are fixed as provided in Table 2.3. **(B)** Circuit diagram **(top)** and simulation result **(bottom)** of a sequestration-based timer with a positive feedback loop.

2.3.5 Sequestration-based timer is implemented using orthogonal sigma and anti-sigma factors

Among many known activator and anti-activators known in bacteria, orthogonal pairs of σs, anti-σs, and promoters have been identified from genomes of diverse bacteria and tested in *E. coli* (Rhodius et al., 2013). In their study, five sets of extracytoplasmic function (ECF) σs anti-σs, and promoters showed orthogonal expression to native *E. coli* σ factors and to each other. The availability of multiple sets of orthogonal parts expands the possibility of dynamic gene regulation and builds toward more complex circuit construction consisting multiple timers.

Nine pairs of ECFs in groups 03, 11, 14, 16, 20, 22, 27, 33, and 38, controlled by AHL and IPTG as shown in Fig. 2.5**A**, were first tested in plate reader assay for delay-generating ability. At time $t < 0$, varying concentrations of AHL were added to the media to pre-induce the cells with anti-σ factors. At time $t = 0$, IPTG was added as the onset signal for σ-factor accumulation. The target gene sfGFP was regulated by the corresponding σ factor promoter. The fluorescence over time exhibited promising time-response curves (Fig. 2.5**C**).

Next, we transferred eight of the ECF pairs in groups 03, 11, 14, 20, 22, 27, 33, and 38 in a single plasmid expression system, and replaced the regulator of anti-σ factor from $luxR$ to $tetR$ (Fig. 2.5**B**), as we suspected the leaky expression from the $luxR$-regulated promoter would restrict the attainable concentrations of anti-σ factor and thus narrow the range of delays. The single plasmid expression system caused loss of sequestration in ECF groups of 20, 22, 27, and 38. ECF 11 showed the best time-response of sfGFP with long tunable delays and high y_{on}. ECF 11 was further implemented with a feedback system to decouple y_{on} from the delay and enable the independent control on the gain of the timer module (Fig. 2.6). Unfortunately, this implementation showed significant delay in growth and unpredictable time responses such as uncorrelated y_{on} and the inducer levels. This is likely due to cyto-toxicity of anti-σ factor observed in the prior study (Rhodius et al., 2013).

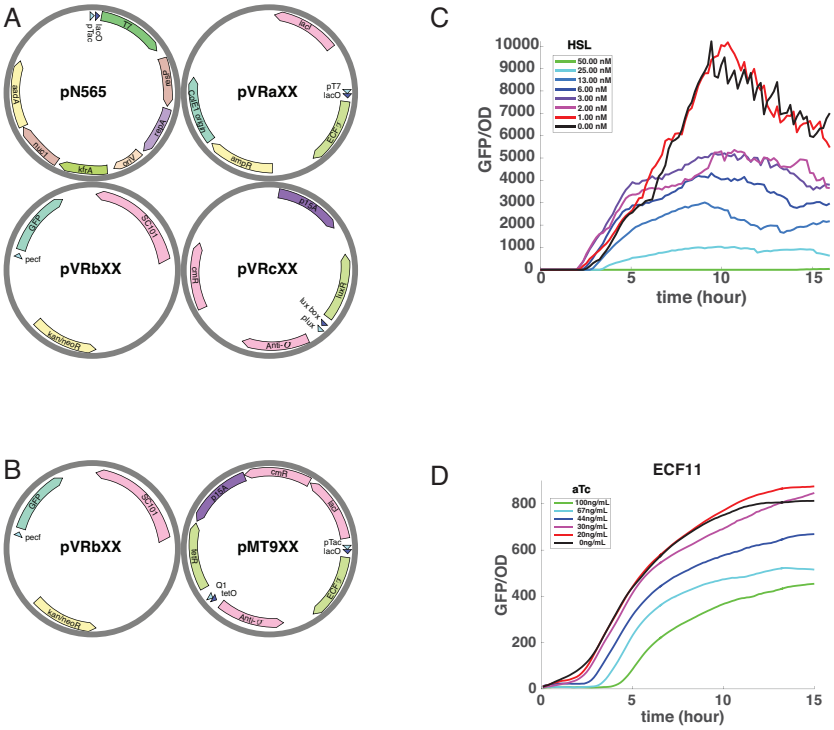

Figure 2.5: Comparison of three-plasmids and single-plasmid ECF expression systems. **(A)** Circuit diagram of the ECF expression system from (Rhodius et al., 2013). **(B)** Circuit diagram of the ECF expression system of single plasmid system. **(C,D)** Plate reader fluorescence readings over time from expression system of **A** and **B**, respectively. Different inducer concentrations for anti-sigma expression were added to the media 5 hours before the experiment. At $t = 0$, 500 μM IPTG is added to the media to start sigma factor accumulation.

2.4 Discussion

The work shown here can be expanded on by [investigating XXX future work] (Commenting on promise and excitement about work on timers). Among the many possible modules in a genetic circuit and in synthetic biology, the timer function and associated frequency-selectivity remains a fundamental building block. This is similar to electronics where the basic parts such as inductors and capacitors serve as frequency-selective or delay components to perform critical functions such as noise filtering, resonant

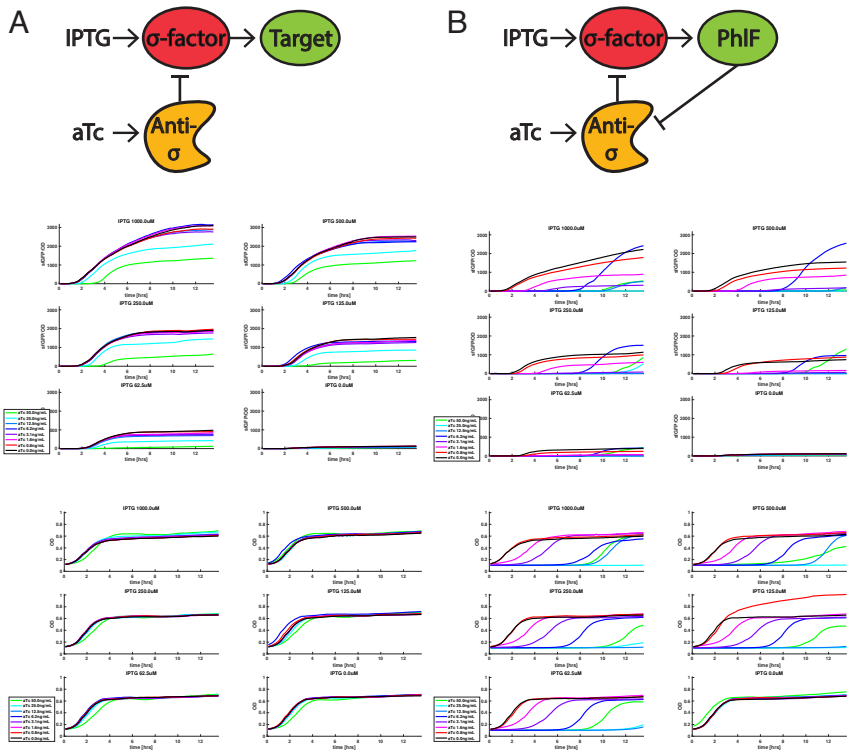

Figure 2.6: Comparison of the circuit diagram, GFP expression, and growth measurements of the two ECF expression systems. Different inducer concentrations were added to the media at time 0. Fluorescence and OD600 were measured by plate reader. **(A)** The original ECF expression system. **(B)** ECF expression system with feedback loop.

amplification as well as for the smoothing of signals. We believe that continued research into designing and improving genetic timers will be invaluable for future genetic circuits.

2.5 Conclusions

The circuit motif enumeration allows optimization of a genetic circuit over different mechanisms toward a desired outcome within the search space. For the delay-generating objective, we found sequestration-activation system to constitute the best timer structure within the tested motif of three or less regulators. Implementation of the sequestration-based timer using ECF σ and anti-σ factors from various microbes demonstrated tunable delays. The optimization of the sequestration timer was deadlocked as the

time response and steady state expression became unpredictable at higher anti-σ factor induction levels. Strong correlation of anti-σ factor induction level and host growth repression suggests undesirable host-circuit interaction contributing to the unpredictability and loss of modularity of the timer circuit.

2.6 Acknowledgments

I would like to thank Drs. Ophelia Venturelli and David Chen for their initial testing and mentorship on this work. I would not have been able to refine the theory this well without critical insights from Ophelia and David.

Chapter 3

Programming mRNA decay to modulate synthetic circuit resource allocation

3.1 Attribution

This chapter is based on a co-authored publication in Nature Communications:
Venturelli OS, Tei M, Bauer S, Chan LJG, Petzold CJ, Arkin AP. Programming mRNA decay to modulate synthetic circuit resource allocation. Nature Communications. 2017 Apr 26;8:15128. Available from: http://dx.doi.org/10.1038/ncomms15128
Author contributions are detailed as follows:
O.S.V. and A.P.A. designed the research. O.S.V. and M.T. carried out the experiments. O.S.V. and M.T. performed the computational modelling. O.S.V., M.T. and A.P.A. discussed data analyses and O.S.V. and M.T. performed the analyses. O.S.V., M.T. and A.P.A. wrote the manuscript. S.B. performed gluconate measurements and L.J.G.C. and C.J.P. implemented shotgun proteomics.

3.2 Introduction

When designing synthetic biological circuits, it must not be forgotten that these circuits exist within a biological entity that needs basic cellular processes to keep alive. As such, a core challenge is to optimize the distribution of the limited energy or resource quota between synthetic circuit functions and basic cell functions (Ophelia S Venturelli et al., 2016; Gyorgy et al., 2015; Dong et al., 1995; Shachrai et al., 2010). However, unlike gene expression

which can be controlled by many techniques like CRISPR (Qi et al., 2013; Zalatan et al., 2015; Gertz et al., 2009) or engineered promoters (Mutalik et al., 2013), there are limited methods to globally redistribute resources and reprogramme cellular state.

In this chapter, we investigate the use of RNA decay for synthetic circuit resource redistribution in *E. coli*. To achieve this, we repurposed a sequence-specific ribonuclease MazF15 (Y. Zhang et al., 2003) whose recognition site 'ACA' is present in 96% of *E. coli* coding sequences. The MazF recognition site can be eliminated from the synthetic circuit while preserving the amino acid content, allowing cellular resources to be reallocated towards synthetic gene expression by eliminating competing processes.

3.3 Materials and Methods

3.3.1 Cloning and strain construction

mazF was deleted from the *E. coli* BW25113 strain using lambda-red recombination and verified by colony PCR. MazF was introduced into an intergenic region referred to as SafeSite 1 (chromosomal position 34715) under control of an aTc-inducible promoter (PTET). PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (NEB) and oligonucleotides (IDT) for cloning and strain construction were obtained from Integrated DNA Technologies. Standard cloning methods were used to construct plasmids. Plasmids were derived from previously generated construct library (T. Lee et al., 2011). A list of plasmids and strains used in this study can be found in Table 3.1.

3.3.2 Growth conditions and plate reader experiments

For plate reader experiments, cells were grown at 37 ^{o}C for $\sim$6–8 h, and then diluted to OD600 of 0.01 in a 96-well plate (Corning) in LB Lennox media (Sigma). In 96-well plates, cells were grown in 200 ml volumes at 37 ^{o}C covered by a gas-permeable seal (Fisher Scientific) in a M1000 (Tecan) or Synergy 2 (BioTek) plate reader. Cells were cultured for 1–2 h in the plate reader before inducer administration. The method measured cell density (OD600) and fluorescence every 10 min for 15 h. The M1000 excitation and emission wavelengths were 485, 510 nm for GFP and 587, 610 nm for RFP (5 nm bandwidth). The BioTek excitation and emission wavelengths were 485, 528nm for GFP and 560, 620 for RFP (20 nm bandwidth). The M1000 and Synergy 2

TABLE 3.1: List of strains and plasmids used.

List of strains and plasmids

Strain or plasmid	Description
S1	KTS022IG[1]; mazF::Δ
S2	BW25113; mazF::Δ; SafeSite1::tetR-P_{TET}-mazF
S3	BW25113; mazF::Δ
P1	pBbS2k-P_{TET}-mazF-U
P2	pBbS2k-P_{TET} mazF-P
P4	pBbA8k-P_{BAD}-mCherry-P
P5	pBbA8k-P_{BAD}-mCherry-U
P6	pBbA6c-P_{LAC}-mCherry-P
P7	pBbA6c-P_{LAC}-mCherry-U
P8	pBbA6c-P_{LAC}-gdh-U
P9	pBbA6c-P_{LAC}-gdh-P
P10	pBbA6c-P_{LAC}-sfGFP-P-gdh-U
P11	pBbA6c-P_{LAC}-sfGFP-P-gdh-P
P12	pBbS6c-P_{LAC}-T7-P
P13	pBbS6c-P_{LAC}-T7-U
P14	pVRa-P_{T7}-mCherry-P-gdh-P
P15	pVRa-P_{T7}-mCherry-P-gdh-U
P16	pBbS6c-lacI-W220F[2]-P_{LAC}-rnr-P
P17	pBbS6c-lacI-W220F[2]-P_{LAC}-tsf-P
P18	pBbS6c-lacI-W220F[2]-P_{LAC}-tsf-P; UNS4[3]-$P_{apFAB321}$[4]-rnr-P-UNS6
P19	pVRa-P_{T7}-mCherry-U
P20	pVRa-P_{T7}-mCherry-P
P21	pBbS8k-P_{BAD}-mCherry-U
P22	pBbS8k-P_{BAD}-mCherry-P
P23	pBbS8k-P_{BAD}-mCherry-26
P24	pBbS8k-P_{BAD}-mCherry-65
P25	pBbS8k-P_{BAD}-mCherry-142
P26	pBbS8k-P_{BAD}-mCherry-191
P27	pBbS8k-P_{BAD}-mCherry-237
P28	pBbS8k-P_{BAD}-mCherry-248
P29	pBbS8k-P_{BAD}-mCherry-339
P30	pBbS8k-P_{BAD}-mCherry-354
P31	pBbS8k-P_{BAD}-mCherry-501
P32	pBbS8k-P_{BAD}-mCherry-550
P33	pBbS8k-P_{BAD}-mCherry-557
P34	pBbS8k-P_{BAD}-mCherry-593
P35	pBbS8k-P_{BAD}-mCherry-619
P36	pBbS8k-P_{BAD}-mCherry-704
P37	pBbS2k-P_{TET}-mazF-1
P38	pBbS2k-P_{TET}-mazF-2
P39	pBbS2k-P_{TET}-mazF-3
P40	pBbS2k-P_{TET}-mazF-4
P41	pBbS2k-P_{TET}-mazF-5
P42	pBbS2k-P_{TET}-mazF-10_position_159
P43	pBbS2k-P_{TET}-mazF-11_position_159
P44	pBbA6c-$P_{LAC\text{-}CSPB}$-cspB(5'UTR)-(14AA)cspB-sfGFP-P
P45	pBbA6c-$P_{LAC\text{-}CSPG}$-cspG(5'UTR)-(14AA)cspG-sfGFP-P

TABLE 3.2: List of qPCR primers and probes

Gene	Forward	Reverse	Probe
rrsA	GTCAGCTCGTGTTGTGAAATG	CCCACCTTCCTCCAGTTTATC	ACGAGCGCAACCCTTATCCTTTGT
cysG	TCTACGACCGTCTGGTTTCT	CGCAGCAGGATCTGGTTAAT	TATGAATCTGGTACGCCGCGATGC
mazF	GGTATGTGTCTGTGTGTTCCTT	CCTTTCTTCGTTGCTCCTCTT	CGTGATGGCTAGCGTTAGCTGAT
mCherry	CAGGATGGCGAGTTCATCTATAA	GGATACATGCGTTCGCTAGA	AAAGACGATGGGTTGGGAGGCG

measured absorbance at 600 nm (OD600) to quantify total biomass. For each experiment, the minimum value of fluorescence or OD600 across all conditions was subtracted from fluorescence or OD600 measurements. Normalized fluorescence was computed by dividing by the maximum value across conditions. Normalized fluorescence divided by total biomass (OD600) was computed by dividing total fluorescence by OD600 and then normalizing to the maximum value across conditions. For plate reader experiments, biological replicates consisted of cells inoculated into different wells in a 96-well plate that were exposed to equivalent inducer concentrations.

3.3.3 qPCR measurements

Oligonucleotides for quantitative real-time PCR (sequences are listed in Table 3.2) were designed using Integrated DNA Technologies. Total RNA of 500 ng was reverse transcribed using the iScript complementary DNA (cDNA) synthesis kit (Bio-Rad). The reaction mix contained 5 ml of SsoAdvanced Universal Probes Supermix (Bio-Rad), 0.5 ml primer and probe corresponding to 250nM primers and 125nM probe ($20\times$ stock) and 0.5 μl of cDNA. qPCR measurements were performed using a CFX96 real-time PCR machine (Bio-Rad). The relative expression levels were determined by a $2^{\Delta\Delta G}$ method. Each sample was normalized by the cycle threshold geometric mean using reference genes rrsA and cysG (Zhou et al., 2011). Biological replicates consisted of three *E. coli* cultures exposed to equivalent inducer concentrations (0 or 5 ng ml^{-1}). Three qPCR technical replicates were performed and averaged for each sample.

3.3.4 Gluconate measurements

KTS022IG mazF::Δ (strain S1 in Table 3.1) strains bearing pBbA6c-gdh-X (plasmid P8-9 in Table 3.1) and pBbS2k-mazF-U (plasmid P1) were grown in LB medium at 37^{o}C overnight and used to inoculate a 10ml culture the next morning at an OD600 of 0.05. At OD600 of 0.3, 1.5% glucose, 1mM IPTG and 5 or 0 ngml^{-1} were administered to the cultures. 1ml samples were collected at the specified times and centrifuged at 5,000g for 5min to isolate the supernatant. The supernatant samples were analysed for gluconic acid using a 1,200 Series liquid chromatography system (Agilent Technologies) coupled to an LTQ-XL ion trap mass spectrometer (Thermo Scientific) equipped with an electrospray ionization source. Aliquots of the diluted samples were injected onto a Rezex ROA-Organic Acid H + (8%) (150mm $\times$ 4.6 mm) column (Phenomenex) equipped with a Carbo-H + $(4 \times 3mm^2)$ guard column (Phenomenex). Gluconic acid was eluted at 55 ^{o}C at $\sim$3.5 min with an isocratic flow rate of 0.3 ml min1 of 0.5% (v/v) formic acid in water. Precursor ion m/z 195.1 was selected in negative ion mode using an isolation window of m/z 2 and was fragmented with a normalized collision energy of 35. Fragment ions were analysed in the range of m/z 50–200. m/z 128.5–129.5 was selected as pseudo-MRM transition for compound quantification. Resulting peak areas were compared to an external standard calibration in the range of 0.2–200 μM. The source parameters were ion spray voltage: 4 kV; capillary temperature: 350 ^{o} C; capillary voltage: 2V; tube lense voltage: 40 V; sheath gas flow: 60; auxiliary gas flow: 5; and sweep gas flow: 10 (all arbitrary units). Technical replicates were performed by measuring the sample three independent times. The experiment was repeated three independent times. These experiments showed that the MazF-induced cells expressing Gdh-P yielded the highest gluconate concentrations compared to uninduced cells and MazF-induced cells expressing Gdh-U.

3.3.5 Proteomics

BW25113 mazF::Δ, SafeSite1::tetR-PTET-mazF (strain S2 in Table 3.1) was grown overnight in LB at 37^{o}C and then diluted to an OD600 of 0.05 in a 500 ml LB culture. At OD600 of 0.5, cell populations were induced with 5 ngml^{-1} aTc and 40 ml of the cultures were collected approximately every hour and centrifuged for 5min at 4,300g. Proteomic samples were prepared for analysis by lysing the cell pellets and extracting the proteins using the chloroform/methanol precipitation method (González Fernández-Niño et al., 2015). The proteins were resuspended in 100mM AMBIC with 20%

methanol and reduced with tris(2-carboxyethyl) phosphine for 30 min, followed by addition of iodoacetamide (IAA; final conc. 10mM) for 30 min in the dark, and then digested overnight with MS-grade trypsin (1:50 w/w trypsin: protein) at 37^{o}C. Peptides were stored at 20^{o}C until analysis. Samples were analysed on an Agilent 1290 UHPLC—6550 QTOF liquid chromatography mass spectrometer (LC–MS/MS; Agilent Technologies) system and the operating parameters for the LC–MS/MS system were described previously (González Fernández-Niño et al., 2015). Peptides were separated on a Sigma-Aldrich Ascentis Express Peptide ES-C18 column ($2.1 \times 100 mm^2$, 2.7mm particle size, operated at 60^{o}C) and a flow rate of 0.4 ml min^{-1}. The chromatography gradient conditions were as follows: from the initial starting condition (98% buffer A containing 100% water, 0.1% formic acid and 2% buffer B composed of 100% acetonitrile, 0.1% formic acid) the buffer B composition was held for 2min then increased to 10% over 3min; then buffer B was increased to 40% over 117 min, then increased to 90% B over 3min and held for 8min, followed by a ramp back down to 2% B over 1min, where it was held for 6min to re-equilibrate the column to the original conditions. The data were analysed with the Mascot search engine version 2.3.02 (Matrix Science) and filtered and validated using Scaffold v4.3.0 (Proteome Software Inc.) (González Fernández-Niño et al., 2015). Replicates consisted of four aliquots of an *E. coli* culture exposed to 5 ngml^{-1} aTc for different lengths of time. Shotgun proteomics was performed independently on each sample.

3.3.6 RNA-seq library construction and sequencing

BW25113 mazF::Δ, Safe- Site1::tetR-PTET-mazF (strain S2 in Table 3.1) was grown overnight in LB at 37 ^{o}C and then diluted to an OD600 of 0.05 in a 10ml LB culture. At an OD600 of 0.5, cells were induced with 5 ngml^{-1} aTc. Samples were collected as follows: 200 ml of the cell cultures were added to 400 ml of RNAprotect (Qiagen) to stabilize the RNA, incubated for 5min at room temperature and then spun down for 10 min at 5,000g. Total RNA was isolated using RNeasy purification kit and treated with DNAase I (Qiagen). The Functional Genomics Lab (FGL), a QB3-Berkeley Core Research Facility at UC Berkeley, constructed the sequencing libraries. At the FGL, Ribo-Zero rRNA Removal Kits (Illumina) were used to remove ribosomal RNA and ERCC RNA Spike-In Control Mixes (Ambion by Life Technologies) were added to the samples. The library preparation was performed

on an Apollo 324 with PrepX RNAseq Library Prep Kits (WaferGen Biosystems, Fremont, CA), and 18 cycles of PCR amplification was used for index addition and library fragment enrichment. Biological replicates consisting of two *E. coli* culture aliquots exposed to 5 ng ml^{-1} aTc were collected at the specified times. RNA-seq libraries were constructed independently from each sample.

3.3.7 RNA-seq data analysis

The read counts were mapped onto the MG1655 genome using Bowtie 1 (Langmead et al., 2009) on the galaxy webserver (Afgan et al., 2018). Reads per kilobase of transcript per million (RPKM) was computed by multiplying the number of mapped reads by 109 and then dividing by the gene length and median number of total reads for each condition. For clustering analysis, the correlation coefficient (r = 0.9) between two biological replicates as a function of time was used as a threshold to remove genes that exhibited variability between replicates. The log2 fold change was partitioned into clusters using the K-means algorithm (MATLAB). To determine an optimal number of clusters, the sum of squared errors was computed for each data point from the corresponding cluster centroid across a range of K-values (1–10). The Elbow method was used as a heuristic to select the optimal number of partitions that minimizes the sum of squared errors. The Fisher's exact test (P = 0.05) was used to evaluate enrichment of genes based on TIGRFAM annotation (MicrobesOnline) or transcription factor network (RegulonDB). Table 3.3 contains a list of genes in the enriched categories.

3.3.8 Computational modeling

We used custom code for computational modelling and data analysis in MATLAB (Mathworks) and Python.

For the mathematical model, we constructed an ordinary differential equation (ODE) model that interrogated the role of MazF on ribosome competition. The model included mRNA and protein concentrations of all species. Major protein species (Table 3.1 contains a complete list of model species) included the unprotected cellular proteome (p), ribosomes (r), MazF monomer ($mazFp$) and dimer ($mazFpd$), MazE ($mazEp$) and a protected gene (FP). For simplicity, we assumed that nutrient levels, cellular energy and RNA polymerase concentrations were constant as a function of time. We assumed that MazE was not expressed (α_e = 0) for modeling

TABLE 3.3: Functional (TIGRFAM) and regulatory (Regulon DB) enrichments in RNA-seq clustering analysis. *Not statistically significant (p = 0.051)

Cell envelope (TIGRFAM) CLUSTER K1	murE(b0085),murF(b0086),mraY(b0087),murD(b0088),murC (b0091),lpxD(b0179),metQ(b0197),mrdA(b0635),pgaD(b102 1),lpxL(b1054),rfbC(b2038),rfbA(b2039),mreD(b3249),slp(b3 506),bcsG(b3538),rfe(b3784),rffH(b3789),rffM(b3794),murI(b 3967),murB(b3972)
Protein synthesis (TIGRFAM) CLUSTER K1	rpsT(b0023),ileS(b0026),rluA(b0058),gluQ(b0144),frr(b0172) ,tilS(b0188),proS(b0194),cysS(b0526),miaB(b0661),glnS(b0 680),rlmC(b0859),infA(b0884),serS(b0893),rpsA(b0911),prf A(b1211),tyrS(b1637),pheT(b1713),pheS(b1714),rpmI(b171 7),aspS(b1866),metG(b2114),yeiP(b2171),prmB(b2330),gltX (b2400),hisS(b2514),iscS(b2530),raiA(b2597),trmD(b2607),r psP(b2609),queD(b2765),lysS(b2890),rpsO(b3165),truB(b31 66),hpf(b3203),rpsD(b3296),rplF(b3305),rpsQ(b3311),rplP(b 3313),rplV(b3315),tufA(b3339),fusA(b3340),trpS(b3384),gly S(b3559),rpmH(b3703),mnmE(b3706),mnmG(b3741),typA(b 3871),trmA(b3965),tufB(b3980),efp(b4147),rpsR(b4202),rplI(b4203),valS(b4258),rimI(b4373)
Translation factors (TIGRFAM) CLUSTER K1	frr(b0172),infA(b0884),prfA(b1211),yeiP(b2171),raiA(b2597), hpf(b3203),tufA(b3339),fusA(b3340),typA(b3871),tufB(b398 0),efp(b4147)
tRNA aminoacetylation (TIGRFAM) CLUSTER K1	ileS(b0026),proS(b0194),cysS(b0526),glnS(b0680),serS(b08 93),tyrS(b1637),pheT(b1713),pheS(b1714),aspS(b1866),me tG(b2114),gltX(b2400),hisS(b2514),lysS(b2890),trpS(b3384) ,glyS(b3559),valS(b4258)
Energy metabolism - amino acids and amines (TIGRFAM) CLUSTER K2	putA(b1014),gcvP(b2903),gcvH(b2904),gcvT(b2905),ansB(b 2957),tdh(b3616),tnaA(b3708),aspA(b4139)
Energy metabolism - anaerobic (TIGRFAM) CLUSTER K2	narG(b1224),narH(b1225),narI(b1227),glpA(b2241),glpB(b2 242),glpC(b2243)
Energy metabolism (TIGRFAM) CLUSTER K2	aceF(b0115),acnB(b0118),ykgF(b0307),cyoD(b0429),cyoC(b0430),cyoB(b0431),pgm(b0688),sdhC(b0721),sdhD(b0722) ,sucB(b0727),sucC(b0728),sucD(b0729),cydB(b0734),ybgT(b4515),pflB(b0903),mgsA(b0963),wrbA(b1004),putA(b1014), narG(b1224),narH(b1225),narI(b1227),acnA(b1276),fdnG(b1 474),rsxA(b1627),rsxB(b1628),rsxG(b1631),rsxE(b1632),glo A(b1651),grxD(b1654),eda(b1850),edd(b1851),pykA(b1854) ,gatZ(b2095),gatY(b2096),ccmF(b2196),napG(b2205),mqo(

	b2210),glpA(b2241),glpB(b2242),glpC(b2243),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoI(b2281),nuoG(b2283),nuoE(b2285),nuoC(b2286),ackA(b2296),pta(b2297),glk(b2388),trxC(b2582),eno(b2779),sdaB(b2797),gcvP(b2903),gcvH(b2904),gcvT(b2905),rpiA(b2914),fbaA(b2925),tktA(b2935),ansB(b2957),tdcG(b4471),tdcE(b3114),garK(b3124),mdh(b3236),pck(b3403),malP(b3417),glgP(b3428),grxC(b3610),gpmM(b3612),tdh(b3616),tnaA(b3708),atpD(b3732),atpB(b3738),fdoI(b3892),tpiA(b3919),glpK(b3926),fsaB(b3946),aceB(b4014),nrfE(b4074),aspA(b4139),frdA(b4154),queG(b4166),deoC(b4381)
Nitrogen (TIGRFAM) CLUSTER K2	napA(b2206),nirB(b3365)
Protein modification (TIGRFAM) CLUSTER K2	pflA(b0902),prc(b1830),hypA(b2726),aslB(b3800),birA(b3973),nrdG(b4237),yjjW(b4379)
Energy metabolism – TCA* (TIGRFAM) CLUSTER K3	acnB(b0118),sdhC(b0721),sdhD(b0722),sucB(b0727),sucC(b0728),sucD(b0729),acnA(b1276),mqo(b2210),mdh(b3236),aceB(b4014),frdA(b4154)
ArcA (RegulonDB) CLUSTER K2	caiE(b0035),aceE(b0114),aceF(b0115),lpd(b0116),acnB(b0118),fadE(b0221),betT(b0314),cyoD(b0429),cyoC(b0430),cyoB(b0431),dcuC(b0621),sdhC(b0721),sdhD(b0722),sucB(b0727),sucC(b0728),sucD(b0729),cydB(b0734),cydC(b0886),pflB(b0903),ndh(b1109),prfA(b1211),oppB(b1244),oppC(b1245),oppD(b1246),acnA(b1276),tpx(b1324),gatD(b2091),gatC(b2092),gatA(b2094),gatZ(b2095),gatY(b2096),glpA(b2241),glpB(b2242),glpC(b2243),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoK(b2279),nuoI(b2281),nuoH(b2282),nuoG(b2283),nuoE(b2285),nuoC(b2286),nuoA(b2288),ackA(b2296),grcA(b2579),glcG(b2977),glcF(b4467),hybB(b2995),mdh(b3236),rpsQ(b3311),rplP(b3313),rplV(b3315),rplC(b3320),dctA(b3528),sodA(b3908),aceB(b4014),ubiC(b4039),cadA(b4131),treC(b4239)
ArgR (RegulonDB) CLUSTER K3	artJ(b0860),hisQ(b2308),hisJ(b2309),pnp(b3164),rpsO(b3165),truB(b3166),rbfA(b3167)
CRP (RegulonDB) CLUSTER K2	caiE(b0035),nadC(b0109),pdhR(b0113),aceE(b0114),aceF(b0115),lpd(b0116),acnB(b0118),dksA(b0145),yaeQ(b0190),nlpE(b0192),araJ(b0396),tsx(b0411),cyoD(b0429),cyoC(b0430),cyoB(b0431),hupB(b0440),entC(b0593),entE(b0594),cstA(b0598),nagA(b0677),sdhC(b0721),sdhD(b0722),sucB(b0727),sucC(b0728),sucD(b0729),fiu(b0805),pflB(b0903),aroA(b0908),ompF(b0929),ompA(b0957),ycdZ(b1036),hlyE(b118

	2),sohB(b1272),acnA(b1276),feaB(b1385),uxaB(b1521),marR(b1530),malY(b1622),sodB(b1656),manX(b1817),manY(b1818),manZ(b1819),gatD(b2091),gatC(b2092),gatA(b2094),gatZ(b2095),gatY(b2096),cdd(b2143),preA(b2147),mglC(b2148),mglA(b2149),mglB(b2150),galS(b2151),cirA(b2155),yeiP(b2171),gyrA(b2231),ubiG(b2232),nrdB(b2235),yfaE(b2236),glpT(b2240),glpA(b2241),glpB(b2242),glpC(b2243),dsdX(b2365),nupC(b2393),ptsH(b2415),ptsI(b2416),crr(b2417),guaB(b2508),grcA(b2579),pka(b2584),raiA(b2597),srlA(b2702),srlE(b2703),srlD(b2705),mazF(b2782),mazE(b2783),gcvP(b2903),gcvH(b2904),gcvT(b2905),serA(b2913),fbaA(b2925),galP(b2943),ansB(b2957),nupG(b2964),glcC(b2980),tdcG(b4471),tdcF(b3113),tdcE(b3114),tdcC(b3116),tdcB(b3117),tdcA(b3118),pnp(b3164),rpsO(b3165),truB(b3166),rbfA(b3167),mdh(b3236),dusB(b3260),nirB(b3365),nirC(b3367),cysG(b3368),pck(b3403),glgP(b3428),gntU(b4476),dctA(b3528),yiaM(b3577),mtlD(b3600),ilvN(b3670),ilvB(b3671),tnaC(b3707),tnaA(b3708),bglF(b3722),rbsA(b3749),rbsC(b3750),udp(b3831),glnL(b3869),glnA(b3870),rhaR(b3906),sodA(b3908),glpK(b3926),glpF(b3927),cytR(b3934),aceB(b4014),proP(b4111),dcuA(b4138),aspA(b4139),hflX(b4173),rpsR(b4202),rplI(b4203),treC(b4239),fecE(b4287),fecD(b4288),fecB(b4290),fecA(b4291),yjiY(b4354),osmY(b4376),deoC(b4381),deoB(b4383)
Cra (RegulonDB) CLUSTER K2	pdhR(b0113),aceE(b0114),aceF(b0115),lpd(b0116),acnB(b0118),betT(b0314),cyoD(b0429),cyoC(b0430),cyoB(b0431),cydB(b0734),adhE(b1241),acnA(b1276),marR(b1530),yeaD(b1780),manX(b1817),manY(b1818),manZ(b1819),eda(b1850),edd(b1851),fruB(b2169),glk(b2388),ptsH(b2415),ptsI(b2416),crr(b2417),eno(b2779),fbaA(b2925),glcC(b2980),nirB(b3365),nirC(b3367),cysG(b3368),pck(b3403),mtlD(b3600),gpmM(b3612),envC(b3613),tpiA(b3919),ppc(b3956),aceB(b4014)
DcuR (RegulonDB) CLUSTER K2	dctA(b3528),frdD(b4151),frdC(b4152),frdA(b4154)
FNR (RegulonDB) CLUSTER K2	caiE(b0035),pdhR(b0113),aceE(b0114),aceF(b0115),lpd(b0116),cyoD(b0429),cyoC(b0430),cyoB(b0431),dcuC(b0621),sdhC(b0721),sdhD(b0722),sucB(b0727),sucC(b0728),sucD(b0729),cydB(b0734),moaA(b0781),ompX(b0814),cydC(b0886),dmsA(b0894),dmsC(b0896),pflB(b0903),aspC(b0928),ndh(b1109),hlyE(b1182),prfA(b1211),narL(b1221),narK(b1223),narG(b1224),narH(b1225),narJ(b1226),narI(b1227),adhE(b1241),ompW(b1256),acnA(b1276),tpx(b1324),fdnG(b1474),ynfH(b1590),yecR(b1904),ccmH(b2194),ccmF(b2196),ccmC(b2199),ccmA(b2201),napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),glpT(b2240),glpA(b

	2241),glpB(b2242),glpC(b2243),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoK(b2279),nuoI(b2281),nuoH(b2282),nuoG(b2283),nuoE(b2285),nuoC(b2286),nuoA(b2288),ackA(b2296),grcA(b2579),trmD(b2607),rpsP(b2609),xdhC(b2868),gcvP(b2903),gcvH(b2904),gcvT(b2905),ansB(b2957),tdcG(b4471),tdcF(b3113),tdcE(b3114),tdcC(b3116),tdcB(b3117),tdcA(b3118),garK(b3124),acrF(b3266),rpsQ(b3311),rplP(b3313),rplV(b3315),rplC(b3320),nirB(b3365),nirC(b3367),cysG(b3368),feoB(b3409),malP(b3417),nikA(b3476),nikB(b3477),nikC(b3478),nikD(b3479),nikE(b3480),pitA(b3493),dppA(b3544),pstA(b3726),pstC(b3727),pstS(b3728),sodA(b3908),katG(b3942),ubiC(b4039),nrfE(b4074),dcuA(b4138),aspA(b4139),frdD(b4151),frdC(b4152),frdA(b4154),nrdG(b4237),nrdD(b4238),arcA(b4401)
Fis (RegulonDB) CLUSTER K2	lpd(b0116),acnB(b0118),thrW(b0244),hupB(b0440),glnX(b0664),valT(b0744),dmsA(b0894),dmsC(b0896),pflB(b0903),ndh(b1109),hlyE(b1182),narK(b1223),narG(b1224),narH(b1225),narJ(b1226),narI(b1227),tpr(b1229),hns(b1237),adhE(b1241),topA(b1274),marR(b1530),mglC(b2148),mglA(b2149),gyrA(b2231),nrdB(b2235),yfaE(b2236),glpT(b2240),glpA(b2241),glpB(b2242),glpC(b2243),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoK(b2279),nuoI(b2281),nuoH(b2282),nuoG(b2283),nuoE(b2285),nuoC(b2286),nuoA(b2288),gltX(b2400),guaB(b2508),grcA(b2579),mazG(b2781),mazF(b2782),mazE(b2783),fau(b2912),ansB(b2957),glcC(b2980),pnp(b3164),rpsO(b3165),truB(b3166),rbfA(b3167),dusB(b3260),nirB(b3365),nirC(b3367),cysG(b3368),mtlD(b3600),bglF(b3722),gltU(b3757),glnL(b3869),glnA(b3870),trmA(b3965),tufB(b3980),nrfE(b4074),proP(b4111),leuP(b4369),osmY(b4376),deoC(b4381),deoB(b4383)
FlhDC (RegulonDB) CLUSTER K2	gltK(b0653),gltI(b0655),flgC(b1074),flgE(b1076),flgJ(b1081),yecR(b1904),mglC(b2148),mglA(b2149),mglB(b2150),ccmH(b2194),ccmF(b2196),ccmC(b2199),ccmA(b2201),napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),glpA(b2241),glpB(b2242),glpC(b2243),recC(b2822),mdh(b3236),nrfE(b4074)
Fur (RegulonDB) CLUSTER K2	lpd(b0116),fhuA(b0150),fhuD(b0152),fhuB(b0153),cyoD(b0429),cyoC(b0430),cyoB(b0431),fepG(b0589),fepD(b0590),fepB(b0592),entC(b0593),entE(b0594),sdhC(b0721),sdhD(b0722),sucB(b0727),sucC(b0728),sucD(b0729),fiu(b0805),aspC(b0928),ompF(b0929),ndh(b1109),oppB(b1244),oppC(b1245),oppD(b1246),tonB(b1252),mntP(b1821),cirA(b2155),grcA(b2579),nrdF(b2676),exbD(b3005),exbB(b3006),garK(b3124),feoB(b3409),sodA(b3908),metJ(b3938),katG(b3942),fecE(b4287),fecD(b4288),fecB(b4290),fecA(b4291),fecI(b4293),fhuF(b4367),yjjZ(b4567)

GatR (RegulonDB) CLUSTER K2	gatD(b2091),gatC(b2092),gatA(b2094),gatZ(b2095),gatY(b2096)
GcvA (RegulonDB) CLUSTER K2	gcvP(b2903),gcvH(b2904),gcvT(b2905)
GlpR (RegulonDB) CLUSTER K2	glpT(b2240),glpA(b2241),glpB(b2242),glpC(b2243),glpK(b3926),glpF(b3927)
HypT (RegulonDB) CLUSTER K1	metQ(b0197),cydB(b0734),fecE(b4287),fecD(b4288),fecB(b4290),fecA(b4291)
IHF (RegulonDB) CLUSTER K2	sucB(b0727),sucC(b0728),sucD(b0729),glnP(b0810),glnH(b0811),dmsA(b0894),dmsC(b0896),pflB(b0903),ompF(b0929),ndh(b1109),prfA(b1211),narK(b1223),narG(b1224),narH(b1225),narJ(b1226),narI(b1227),pspE(b1308),sodB(b1656),ihfA(b1712),ompC(b2215),atoE(b2223),glpT(b2240),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoK(b2279),nuoI(b2281),nuoH(b2282),nuoG(b2283),nuoE(b2285),nuoC(b2286),nuoA(b2288),hypA(b2726),glcG(b2977),glcF(b4467),tdcG(b4471),tdcF(b3113),tdcE(b3114),tdcC(b3116),tdcB(b3117),tdcA(b3118),dcuD(b3227),dusB(b3260),nirB(b3365),nirC(b3367),cysG(b3368),uspA(b3495),dppA(b3544),yiaM(b3577),ibpB(b3686),pstA(b3726),pstC(b3727),pstS(b3728),ilvG_1(),ilvG_2(),ilvD(b3771),sodA(b3908),aceB(b4014),ubiC(b4039),nrfE(b4074),adiA(b4117),fimI(b4315),osmY(b4376)
IscR (RegulonDB) CLUSTER K2	napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),iscS(b2530),rnlA(b2630),rnlB(b2631),nrdF(b2676)
LexA (RegulonDB) CLUSTER K1	ftsI(b0084),murE(b0085),murF(b0086),mraY(b0087),murD(b0088),ftsW(b0089),murC(b0091),ftsQ(b0093),ftsA(b0094),uvrB(b0779),ftsK(b0890),yebG(b1848),recD(b2819),dnaG(b3066),lexA(b4043)
MazE-MazF (RegulonDB) CLUSTER K3	mazG(b2781),mazF(b2782),mazE(b2783)
Mlc (RegulonDB) CLUSTER K2	manX(b1817),manY(b1818),manZ(b1819),ptsH(b2415),ptsI(b2416),crr(b2417)
ModE (RegulonDB) CLUSTER K2	moaA(b0781),dmsA(b0894),dmsC(b0896),narL(b1221),oppB(b1244),oppC(b1245),oppD(b1246),ccmH(b2194),ccmF(b2196),ccmC(b2199),ccmA(b2201),napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),deoC(b4381),deoB(b4383)
MraZ (RegulonDB) CLUSTER K1	ftsI(b0084),murE(b0085),murF(b0086),mraY(b0087),murD(b0088),ftsW(b0089),murC(b0091),ftsQ(b0093),ftsA(b0094),mioC(b3742)

NarL (RegulonDB) CLUSTER K2	cydC(b0886),dmsA(b0894),dmsC(b0896),pflB(b0903),narK(b1223),narG(b1224),narH(b1225),narJ(b1226),narI(b1227),adhE(b1241),fdnG(b1474),ynfH(b1590),ccmH(b2194),ccmF(b2196),ccmC(b2199),ccmA(b2201),napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),nuoN(b2276),nuoM(b2277),nuoL(b2278),nuoK(b2279),nuoI(b2281),nuoH(b2282),nuoG(b2283),nuoE(b2285),nuoC(b2286),nuoA(b2288),hybB(b2995),nirB(b3365),nirC(b3367),cysG(b3368),nikA(b3476),nikB(b3477),nikC(b3478),nikD(b3479),nikE(b3480),ubiC(b4039),nrfE(b4074),dcuA(b4138),aspA(b4139),frdD(b4151),frdC(b4152),frdA(b4154)
NarP (RegulonDB) CLUSTER K2	fdnG(b1474),ccmH(b2194),ccmF(b2196),ccmC(b2199),ccmA(b2201),napC(b2202),napB(b2203),napH(b2204),napG(b2205),napA(b2206),napD(b2207),nirB(b3365),nirC(b3367),cysG(b3368),nrfE(b4074)
NikR (RegulonDB) CLUSTER K2	nikA(b3476),nikB(b3477),nikC(b3478),nikD(b3479),nikE(b3480)
RstA (RegulonDB) CLUSTER K2	ompF(b0929),narG(b1224),narH(b1225),narJ(b1226),narI(b1227)
TdcR (RegulonDB) CLUSTER K2	tdcG(b4471),tdcF(b3113),tdcE(b3114),tdcC(b3116),tdcB(b3117),tdcA(b3118)
TorR (RegulonDB) CLUSTER K2	tnaC(b3707),tnaA(b3708)
NtrC (RegulonDB) CLUSTER K3	glnP(b0810),glnH(b0811),cbl(b1987),hisQ(b2308),hisJ(b2309),yhdW(b3268),yhdX(b3269),yhdY(b3270),yhdZ(b3271),glnL(b3869),glnA(b3870)

analyses except Supplementary Fig. 16. Table 3.5 contain a description of model species and parameters and Supplementary Fig. 12 shows the molecular interactions in the model. The Supplementary Software contains MATLAB code for simulation of the model. Parameters were based on previous literature (see Table 3.5). Unknown parameters were estimated using physiologically relevant values for *E coli*. We focused on qualitative circuit behaviors, which were robust to small variations in free parameters. The growth rate function was based on a previous model of growth and gene expression. Here, growth rate is defined as $\lambda = k_{trans} P_{tot}^{-1}([rp] + [rr])$. For steadystate analyses, the model was first simulated to steady-state (t = 278 hr) from an initial condition corresponding to 1 nM for all species. Second, the steady-state species concentrations from the first simulation were used as the initial condition for a second simulation. This two-stage simulation mirrored our experimental design, whereby cells were grown for a period of time prior to inducer administration. The equations are detailed in Figure 3.4.

3.3.9 Characterization of cell viability

A BW25113 mazF::D strain (strain S3 in Table 3.1) transformed with pBbS2k-mazF-U or pBbS2k-mazF-P (plasmid P1-2 in Table 3.1) was grown overnight at 37 ^{o}C in LB media and then diluted to an OD600 of 0.01 in 5ml LB media. At an OD600 of 0.3, 5 ngml^{-1} aTc dissolved in 100% ethanol was used to induce the cells and an equivalent volume of 100% ethanol was administered to the uninduced cell populations. Following 0 and 7 h, cells were prepared for fluorescent microscopy using the LIVE/DEAD Baclight Bacteria Viability Kit (Thermo Fisher) to characterize the fraction of viable cells across the population. Microscope images were collected using a Zeiss Axio Observer D1 and Plan-Apochromat 63/1.4 Oil Ph3 M27 objective (Zeiss). Cells were imaged using excitation BP 470/40 and emission BP 525/50 (Filter Set 38 High Efficiency) or excitation 560/40 and emission BP 630/75 (Filter Set 45). Images were captured with a Hamamatsu ORCA-Flash4.0 using the ZEN Software (Zeiss). Cell Counter (Fiji)63 was used to analyse the images and quantify the number of viable and dead cells. Technical replicates consisted of aliquots of *E. coli* cultures that were independently prepared for microscopy using the LIVE/DEAD protocol.

TABLE 3.4: Equations used in the MazF resource allocation model

$$\frac{d[m_p]}{dt} = \alpha_p - k_f[m_p][r] + k_{rr}[\mathrm{rp}] + k_{trans}[\mathrm{rp}] - k_f[m_p][\mathrm{mazFpd}] + k_{rfp}[\mathrm{pf}] - \lambda[m_p] - k_d[m_p],$$

$$\frac{d[m_r]}{dt} = \alpha_r - k_f[m_r][r] + k_{rrR}[\mathrm{rr}] + k_{trans}[\mathrm{rr}] - k_f[m_r][\mathrm{mazFpd}] + k_{rfr}[\mathrm{rf}] - \lambda[m_r] - k_d[m_r],$$

$$\frac{d[m_f]}{dt} = \alpha_f - k_f[m_f][r] + k_{rr}[\mathrm{rmf}] + k_{trans}[\mathrm{rmf}] - k_f[m_f][\mathrm{mazFpd}] + k_{rff}[\mathrm{ff}] - \lambda[m_f] - k_d[m_f],$$

$$\frac{d[\mathrm{m_{FP}}]}{dt} = \alpha_{FP} - k_f[m_{FP}][r] + k_{rr}[\mathrm{rFP}] + k_{trans}[\mathrm{rFP}] - \lambda[m_{FP}] - k_d[m_{FP}],$$

$$\frac{d[m_e]}{dt} = \alpha_e \left(\frac{[\mathrm{mazFpd}]}{[\mathrm{mazFpd}] + K_{qe}} \right) - k_f[m_e][r] + k_{rr}[\mathrm{rme}] + k_{trans}[\mathrm{rme}] - k_f[m_e][\mathrm{mazFpd}] + k_{rfe}[\mathrm{fe}] - \lambda[m_e] - k_d[m_e],$$

$$\frac{d[p]}{dt} = k_{trans}[\mathrm{rp}] - \lambda[p] - k_{pdeg}[p],$$

$$\frac{d[r]}{dt} = -k_f[m_e][r] + k_{rr}[\mathrm{rme}] + k_{trans}[\mathrm{rme}] - k_f[m_p][r] + k_{rr}[\mathrm{rp}] + k_{trans}[\mathrm{rp}] - k_f[m_r][r] + k_{rrR}[\mathrm{r} + 2k_{trans}[\mathrm{rr}] - k_f[m_f][r] + k_{rr}[\mathrm{rmf}] + k_{trans}[\mathrm{rmf}] - k_f[m_{FP}][r] + k_{rr}[\mathrm{rFP}] + k_{trans}[\mathrm{rFP}] - k_{pdeg}[r] - \lambda[r],$$

$$\frac{d[FP]}{dt} = k_{trans}[\mathrm{rFP}] - \lambda[\mathrm{FP}] - k_{pdegFP}[\mathrm{FP}],$$

$$\frac{d[\mathrm{mazFp}]}{dt} = k_{trans}[\mathrm{rmf}] - 2k_f[\mathrm{mazFp}]^2 + 2k_{rd}[\mathrm{mazFpd}] - \lambda[\mathrm{mazFp}] - k_{pdeg}[\mathrm{mazFp}],$$

$$\frac{d[\mathrm{mazFpd}]}{dt} = k_f[\mathrm{mazFp}]^2 - k_{rd}[\mathrm{mazFpd}] - \lambda[\mathrm{mazFpd}] - k_f[m_p][\mathrm{mazFpd}] + k_{rfp}[\mathrm{pf}] + k_{deg}[\mathrm{pf}] - k_f[m_r][\mathrm{mazFpd}] + k_{rfr}[\mathrm{rf}] + k_{deg}[\mathrm{rf}] - k_f[m_f][\mathrm{mazFpd}] + k_{rff}[\mathrm{ff}] + k_{deg}[\mathrm{ff}] - k_{pdeg}[\mathrm{mazFpd}] - k_f[\mathrm{maze}][\mathrm{mazFpd}] + k_{rseq}[\mathrm{cef}] + k_{rfe}[\mathrm{fe}] + k_{deg}[\mathrm{fe}] - k_f[m_e][\mathrm{mazFpd}],$$

$$\frac{d[\mathrm{mazEp}]}{dt} = k_{trans}[\mathrm{rme}] - k_{pdegE}[\mathrm{mazEp}] - \lambda[\mathrm{mazEp}] - k_f[\mathrm{mazEp}][\mathrm{mazFpd}] + k_{rseq}[\mathrm{cef}],$$

$$\frac{d[\mathrm{rp}]}{dt} = k_f[m_p][r] - k_{rr}[\mathrm{rp}] - k_{trans}[\mathrm{rp}] - \lambda[\mathrm{rp}] - k_{pdeg}[\mathrm{rp}],$$

$$\frac{d[rr]}{dt} = k_f[m_r][r] - k_{rrR}[\mathrm{rr}] - k_{trans}[\mathrm{rr}] - \lambda[\mathrm{rr}] - k_{pdeg}[\mathrm{rr}],$$

$$\frac{d[\mathrm{rmf}]}{dt} = k_f[m_f][r] - k_{rr}[\mathrm{rmf}] - k_{trans}[\mathrm{rmf}] - \lambda[\mathrm{rmf}] - k_{pdeg}[\mathrm{rmf}],$$

$$\frac{d[\mathrm{rFP}]}{dt} = k_f[m_{FP}][r] - k_{rr}[\mathrm{rFP}] - k_{trans}[\mathrm{rFP}] - \lambda[\mathrm{rFP}] - k_{pdeg}[\mathrm{rFP}],$$

$$\frac{d[\mathrm{rme}]}{dt} = k_f[m_e][r] - k_{rr}[\mathrm{rme}] - k_{trans}[\mathrm{rme}] - \lambda[\mathrm{rme}] - k_{pdeg}[\mathrm{rme}],$$

$$\frac{d[\mathrm{pf}]}{dt} = k_f[m_p][\mathrm{mazFpd}] - k_{rfp}[\mathrm{pf}] - \lambda[\mathrm{pf}] - k_{pdeg}[\mathrm{pf}] - k_{deg}[\mathrm{pf}],$$

$$\frac{d[\mathrm{rf}]}{dt} = k_f[m_r][\mathrm{mazFpd}] - k_{rfr}[\mathrm{rf}] - \lambda[\mathrm{rf}] - k_{pdeg}[\mathrm{rf}] - k_{deg}[\mathrm{rf}],$$

$$\frac{d[\mathrm{ff}]}{dt} = k_f[m_f][\mathrm{mazFpd}] - k_{rff}[\mathrm{ff}] - \lambda[\mathrm{ff}] - k_{pdeg}[\mathrm{ff}] - k_{deg}[\mathrm{ff}],$$

$$\frac{d[\mathrm{fe}]}{dt} = k_f[m_e][\mathrm{mazFpd}] - k_{rfe}[\mathrm{fe}] - \lambda[\mathrm{fe}] - k_{pdeg}[\mathrm{fe}] - k_{deg}[\mathrm{fe}],$$

$$\frac{d[\mathrm{cef}]}{dt} = k_f[\mathrm{mazEp}][\mathrm{mazFpd}] - k_{rseq}[\mathrm{cef}] - \lambda[\mathrm{cef}] - k_{pdeg}[\mathrm{cef}].$$

TABLE 3.5: Parameters used in the MazF resource allocation model

Model species	Description
m_p	p mRNA
m_r	r mRNA
m_f	mazF mRNA
m_{FP}	FP mRNA
m_e	mazE mRNA
p	p protein
r	r protein
FP	FP protein
mazFp	MazF protein
mazFpd	MazF dimer
mazEp	Maze protein
rp	m_p-r complex
rr	m_r-r complex
rmf	m_f-r complex
rFP	m_{FP}-r complex
rme	m_e-r complex
pf	m_p-mazFpd complex
rf	m_r-mazFpd complex
ff	m_f-mazFpd complex
fe	m_e-mazFpd complex
cef	mazFpd-mazEp complex

Parameter	Description	Value	Unit	Reference(s)
α_p	m_p transcription rate	1200	nM min^{-1}	[5]
α_r	m_r transcription rate	6	nM min^{-1}	[6]
α_f	m_f transcription rate	0-60	nM min^{-1}	[6]
α_{FP}	m_{FP} transcription rate	0.06	nM min^{-1}	[6]
k_d	mRNA degradation rate	0.252	min^{-1}	[6,7]
k_{deg}	mRNA degradation rate catalyzed by mazFpd	2.52	min^{-1}	[8]
k_f	Forward binding rate	60	nM^{-1} min^{-1}	We assumed the on-rate is diffusion-limited.

3.3.10 Statistics

Statistical analyses and sample sizes for each experiment are described in the figure legends and Methods subsections. Data represent the mean ± 1 s.d., unless noted otherwise. P = 0.05 was considered significant.

3.3.11 Code Availability

The authors declare that all computer code supporting the findings of this study is available on request. Data availability. The RNA-seq data in this study have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) with accession code GSE94998.

3.4 Results

3.4.1 MazF allocates intracellular resources from host processes to synthetic pathways

To explore whether manipulation of resource allocation could predictably modulate circuit behaviour, we needed to develop a comprehensive reallocation mechanism that preserved core processes required for a target function, while downregulating competing pathways. MazF is a sequence-dependent and ribosome-independent endoribonuclease that cleaves the recognition site 'ACA' in single-stranded RNA (Y. Zhang et al., 2003; Y. Zhang et al., 2005). Approximately 96% of *E. coli* coding sequences contain at least one MazF recognition site (Figure 3.2a). Thus, induction of MazF should inhibit cellular processes other than those protected from its action.

We characterized the impact of MazF on expression of a target gene mCherry that contained nine recognition sites in the coding sequence (mCherry-U) or was recorded to not contain any sites using alternative codons (mCherry-P). mazF was introduced into an intergenic genomic site under control of an aTc-inducible promoter (P_{TET}) in an *E. coli* strain deleted for mazF (strain S2 in Table 3.1). The total fluorescence of mCherry-P and mCherry-U were similar in the absence of MazF, indicating that recoding the transcript did not modify expression (Fig. 3.2b). The MazF induction ratio is a metric used to quantify resource redistribution activity, and is defined as the ratio of total mCherry-P fluorescence in the presence to absence of MazF. Following 10 h of induction with 0 or 5 ng ml^{-1} aTc, the MazF

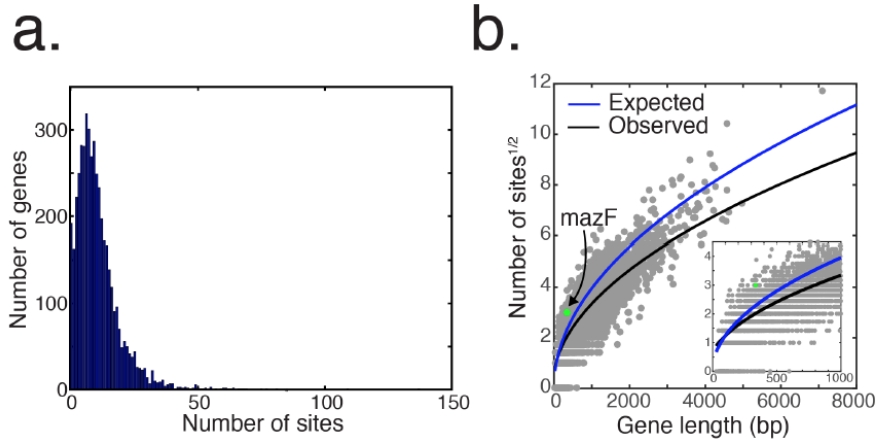

Figure 3.1: MazF recognition sites in *E. coli* coding sequences. (a) Distribution of MazF recognition sites in *E. coli* coding sequences. 96% of coding sequences contain at least one recognition site. (b) Relationship between gene length and the square root of the number of recognition sites for each transcript. Expected relationship based on the GC content of the *E. coli* genome (blue line) and observed relationship (black line) using robust linear regression. mazF is highlighted (green data point) and contains a larger number of MazF sites than predicted. Scatter plot of gene length vs. square root of the number of sites across a limited range of values (inset)

induction ratio was < 1 for mCherry-U and 5 for mCherry-P (Fig. 3.2c). The sequence protection ratio of total fluorescence, defined as the ratio of mCherry-P to mCherry-U, was ~ 1 or 19 in the absence or presence of MazF (Fig. 3.2d). Altogether, these data show that MazF significantly enhanced protected and inhibited unprotected gene expression.

To map the relationship between MazF expression and resource redistribution activity, growth and mCherry-X (X denotes U or P) expression were measured across a broad range of aTc concentrations. The total fluorescence of mCherry-U driven by an arabinose-inducible promoter (P_{BAD}) was reduced up to 4-fold in response to aTc (Figure 3.3). In the presence of aTc, the MazF induction ratio of total fluorescence was enhanced (Fig. 3.2e), whereas the total biomass was lower (Figure 3.4a). The MazF induction ratio of fluorescence divided by OD600 increased with aTc and arabinose (Fig. 3.4b). While the biomass normalization factor altered the quantitative value of the induction ratio, the qualitative relationship between MazF activity and protected gene expression was unmodified (Fig. 3.2e and Fig. 3.3 and Fig. 3.4b). These data highlight that mCherry-P expression and biomass synthesis were inversely correlated in response to MazF. In sum, our results

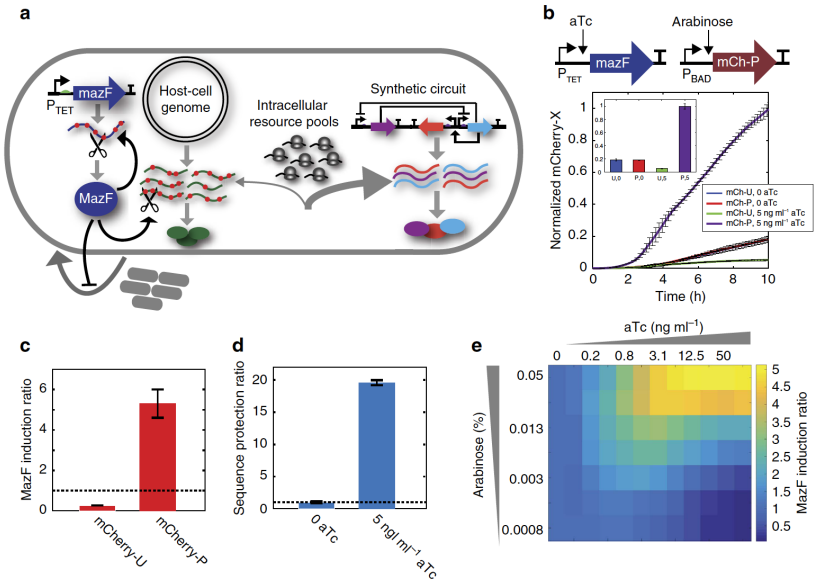

Figure 3.2: Redistributing resources in *E. coli* by programming mRNA decay. (a) Schematic diagram of the MazF resource allocator. Host-cell transcripts containing MazF recognition sites ('ACA') are targeted for cleavage. The MazF site can be removed from target genes while preserving the amino acid sequence. As such, MazF down-regulates transcripts that compete with the protected synthetic circuit for limiting resources, yielding an increase in protected gene expression. (b) MazF and protected mCherry (mCherry-P) were controlled by an aTc and arabinose-inducible promoter (top), respectively. Time-series measurements of total fluorescence normalized to the maximum steady-state value (t = 10 h) across conditions for cells expressing unprotected mCherry (mCherry-U) or mCherry-P in the presence (5 ngml^{-1} aTc) or absence (0 ng ml^{-1} aTc) of MazF. Cells were induced with 0.05% arabinose. Bar plot showing the steady-state normalized mCherry fluorescence (inset). (c) MazF induction ratio, defined as the total fluorescence of mCherry-X in the presence (5 ng ml^{-1} aTc) to absence (0 ng ml^{-1} aTc) of MazF. Cells were induced for 10 h with 0.05% arabinose. (d) Sequence protection ratio, defined as the total fluorescence ratio of mCherry-P to mCherry-U in the presence (5 ng ml^{-1} aTc) or absence (0 ng ml^{-1} aTc) of MazF. Cells were induced for 10 h with 0.05% arabinose. **(e)** Heat-map of MazF induction ratio of total fluorescence following 10 h of induction across a range of arabinose and aTc concentrations. Error bars represent 1 s.d. (n = 3).

suggest that the enhancement of the protected gene mCherry-P in MazF-induced cells is due to augmented synthesis.

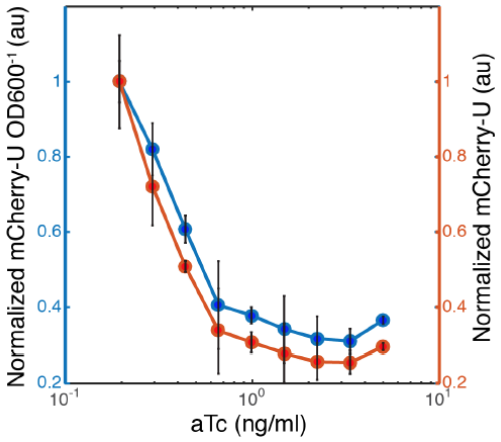

Figure 3.3: Relationship between the concentration of MazF and the expression of an unprotected fluorescent reporter (mCherry-U). mazF and mCherry-U were regulated by an aTc-inducible promoter (PTET) and arabinose-inducible promoter (PBAD), respectively. Total fluorescence or fluorescence divided by OD600 was normalized to the maximum value across conditions. Cells were induced with 0.05% arabinose for 14.2 hr. Error bars represent 1 s.d. (n = 4).

To interrogate the temporal variation in expression in MazF-induced cells, cell populations were induced with mCherry-P at three time points following exposure to MazF. To account for variability in biomass across conditions, we evaluated fluorescence divided by OD600 since the qualitative relationships were not altered by the biomass normalization factor (Fig. 3.2b,e; Figs 3.3 and 3.4b). To compare expression across conditions, fluorescence divided by OD600 was normalized to the maximum expression level across all conditions following 12 h of induction with 5 ngml^{-1} aTc. In the absence of MazF, delayed induction by 2 h reduced mCherry-P expression by 85% (Fig. 3.5a), whereas cells induced with MazF displayed a 34% decrease in mCherry-P expression (Fig. 3.5b). These data indicate that heterologous expression was significantly attenuated by delayed induction in the absence of MazF, presumably by the transition from exponential to stationary phase. By contrast, delays in the induction of mCherry-P reduced expression by a smaller magnitude in the presence of MazF, indicating that MazF-induced cells preserved high-metabolic activity for a period of time.

To distinguish whether transcriptional or translation activity dominated the enhancement of mCherry-P in response to MazF, mCherry-P mRNA

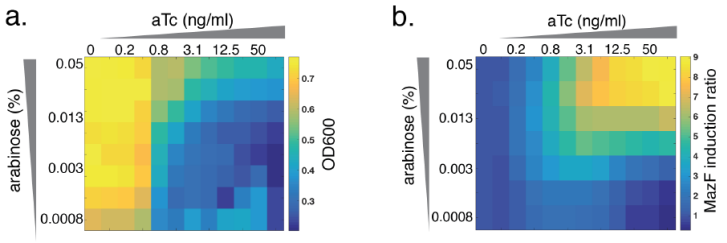

Figure 3.4: MazF activity inhibited growth and enhanced the expression of a protected fluorescent reporter (mCherry-P). MazF and mCherry-X were regulated by an aTc (PTET) and arabinose-inducible (PBAD) promoter, respectively. (a) OD600 at saturation across a range of aTc and arabinose concentrations. Cells were induced for 10 hr. (b) MazF induction ratio of mCherry-P fluorescence divided by OD600 across a range of arabinose and aTc concentrations. Cells were induced for 10 hr. We measured the expression of mCherry-P divided by OD600 in the presence and absence of MazF induction. The induction ratio is defined as the division of the former quantity by the latter.

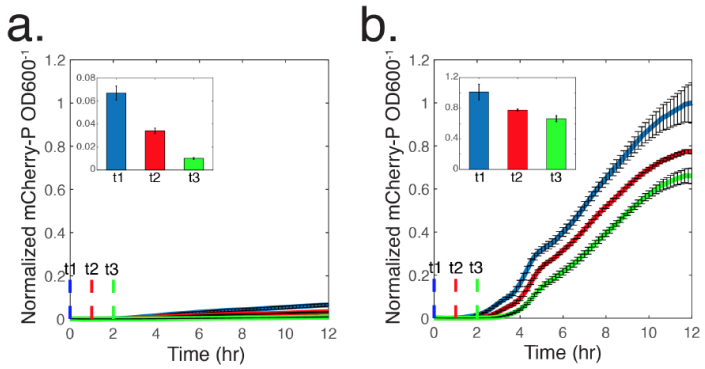

Figure 3.5: Relationship between the timing of induction of mCherry-P and timeresolved measurements of gene expression in the presence (5 ng/mL aTc) or absence (0 ng /mL aTc) of MazF. mCherry-P and MazF were controlled by an IPTG (PLAC) and aTc-inducible promoter (PTET). mCherry-P was divided by OD600 and normalized to maximum steady-state expression value across all conditions. (a) Normalized mCherry-P expression as a function of time for uninduced cells (0 ng/mL aTc). 1 mM IPTG was administered to different cell populations at three times highlighted by the dashed lines (t1 = 0 hr, t2 = 1 hr and t3 = 2 hr). Bar plot of normalized mCherry-P expression for the t1, t2 and t3 conditions following 12 hours of induction (inset). (b) Normalized mCherry-P expression as a function of time for cells induced with MazF (5 ng/mL aTc). Cells were induced with IPTG at three times highlighted by the dashed lines. Bar plot of normalized mCherry-P expression for the t1, t2 and t3 conditions following 12 hr of induction (inset). Error bars represent 1 s.d. (n = 4).

was measured using quantitative real-time PCR (qPCR). The mCherry-P mRNA fold change following 56 min of induction with 0 or 5 ngml^{-1} aTc relative to mCherry-P mRNA abundance at the beginning of the experiment (t = 0) was similar in the presence or absence of MazF (Fig. 3.6). These data show that MazF did not significantly alter the mCherry-P transcription rate over this period of time. Therefore, these results suggest that MazF activity augmented the translation rate of mCherry-P.

3.4.2 Gluconate activity is enhanced during MazF-induced resource allocation

The gluconate pathway competes directly with biomass synthesis by redirecting glucose into gluconate via glucose dehydrogenase (Gdh, Fig. 3.8a). To determine the impact of MazF on metabolic flux, biomass and gluconate were measured as a function of time (see Methods) in cells expressing protected Gdh (gdh-P) or unprotected Gdh containing 10 MazF recognition sites (gdh-U) controlled by a P_{LAC} promoter. These experiments were conducted in a strain background that contained genetic modifications to inhibit gluconate metabolism and decouple glucose phosphorylation and transport to efficiently utilize glucose as a substrate for target metabolic pathways (strain S1 in Table 3.1) (Solomon et al., 2012). As expected, cell growth was inhibited by MazF induction whereas the uninduced population continued to grow as a function of time (Fig. 3.8b). Cells bearing gdh-P driven by a P_{LAC} promoter displayed up to a three-fold higher gluconate concentration and five-fold higher gluconate per unit time in the presence of MazF compared to cells that were not induced with aTc (Fig. 3.8c; Fig. 3.7a). The gluconate titre was 85% higher for cells induced with MazF compared to cells that were not induced following 18.25 h (Fig. 3.8d). A protected fluorescent reporter sfGFP (sfGFP-P) N-terminally fused to Gdh-U or Gdh-P increased up to 3.3 and five-fold as a function of aTc (Fig. 3.7b). These data demonstrated that the MazF resource allocator could enhance metabolic flux by protecting genes in a target metabolic pathway.

3.4.3 Protection of host-factors to enhance resource allocation

Synthetic circuits depend on a dense network of host-genes including the transcriptional and translational machinery. Therefore, MazF-mediated decay of host factors could impact circuit functions. To investigate whether

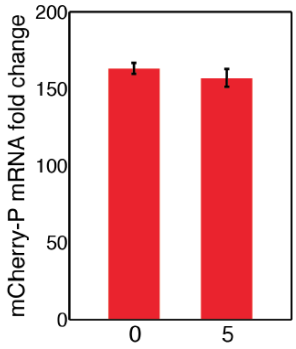

Figure 3.6: mCherry-P mRNA fold change measured by qPCR in the absence (0 ng/mL aTc) or presence (5 ng/mL aTc) of MazF following 56 min of induction. Cells were induced with 0.05% arabinose. MazF and mCherry-P were controlled by an aTc (PTET) and arabinose-inducible (PBAD) promoter, respectively. Error bars represent 1 s.d. (n = 3).

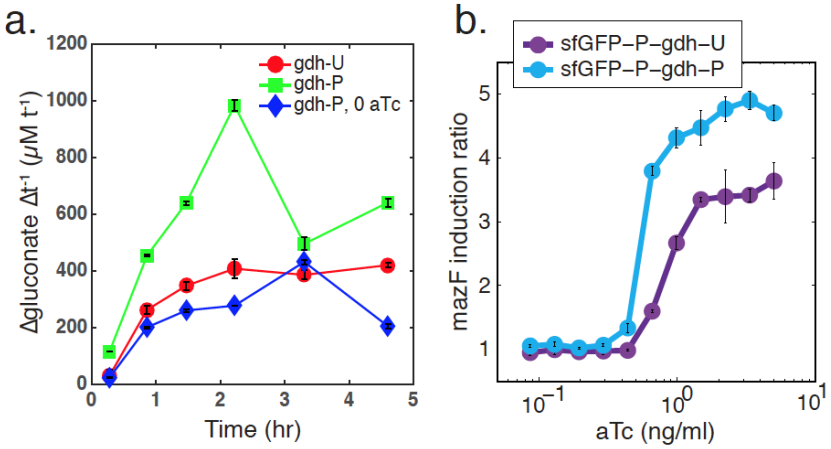

Figure 3.7: MazF resource allocator significantly enhanced gluconate flux and the expression of a protected fluorescent protein fusion (sfGFP-P) to glucose dehydrogenase (Gdh). MazF was regulated by an aTc-inducible promoter (PTET). The protected and unprotected versions of Gdh and sfGFP-Gdh were controlled by an IPTG-inducible promoter (PLAC). (a) Change in gluconate concentration per unit time for cell populations expressing an unprotected version of Gdh (gdh-U) or protected Gdh (gdh-P) driven by an IPTG-inducible promoter (PLAC) in the presence (5 ng/mL aTc) or absence of MazF (0 ng/mL aTc). The media was supplemented with 1.5% glucose and 1 mM IPTG. (b) MazF induction ratio of an N-terminal fusion of sfGFP-P to Gdh-U or Gdh-P as a function of aTc following 12.2 hr of induction with 1 mM IPTG. The media was supplemented with 1.5% glucose. We measured the expression of sfGFP-P divided by OD600 in the presence and absence of MazF induction. The induction ratio is defined as the division of the former quantity by the latter. Error bars represent 1 s.d. from the mean of three technical replicates (n = 3).

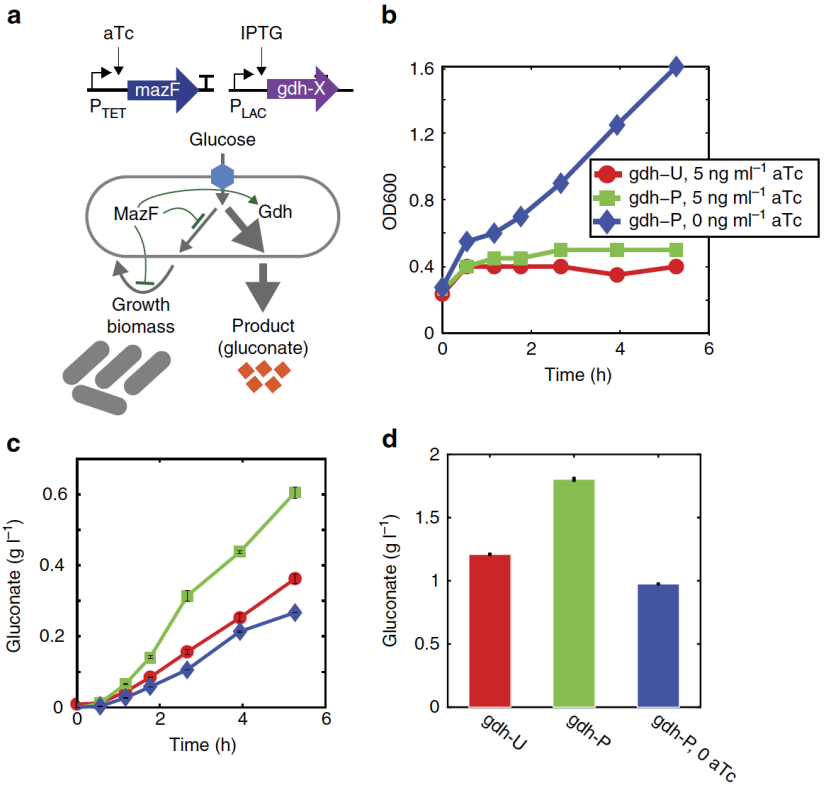

Figure 3.8: The MazF resource allocator enhanced gluconate production. (a) Schematic diagram of the circuit design (top) and gluconate metabolic pathway (bottom). Glucose dehydrogenase (Gdh) transforms glucose into gluconate and competes directly with biomass synthesis. MazF and glucose dehydrogenase (gdh) were controlled by an aTc (PTET) and IPTG-inducible (PLAC) promoter, respectively. (b) OD600 as a function of time for cells expressing Gdh that contained 11 (Gdh-U) or 0 recognition sites (Gdh-P) in response to 5 or 0 ng ml^{-1} aTc (below). All cultures were induced with 1mM IPTG and supplemented with 1.5% glucose. (c) Gluconate titre as a function of time. (d) Gluconate titre following 18.25 h of induction. Error bars represent 1 s.d. from the mean of technical replicates (n = 3).

protection of support genes could improve the performance of the resource allocator, we tested whether protection of an orthogonal RNA polymerase T7 could enhance the circuit output. A protected (T7-P) or unprotected T7 RNA polymerase (T7-U containing 50 MazF sites) controlled by an IPTG-inducible promoter (P_{LAC}) was used to drive the expression of mCherry (Fig. 3a). The combination of T7-P and mCherry-P yielded a 21 or 7.6-fold higher expression level of mCherry compared T7-P, mCherry-U or T7-U, mCherry- P in the presence of MazF (5 ng ml^{-1} aTc) and 1mM IPTG. T7- P

regulating an N-terminal fluorescent protein fusion of mCherry-P to Gdh-P (mCherry-P-Gdh-P) exhibited a 1.4 and 15-fold higher expression compared to T7-P, mCherry-P-Gdh-U or T7-U, mCherry-P-Gdh-P (Fig. 3.10). The mCherry expression level of the T7-X, mCherry-X (Fig. 3a) and T7-X, mCherry-X-Gdh-X (Fig. 3.10) circuits were differentially enhanced by protection of T7 RNA polymerase or the reporter gene (mCherry-X or mCherry-X-gdh-X) in the presence of MazF. Thus, the quantitative value of the enhancement by protection of specific genes in a circuit depended on the circuit composition.

Defining translation factors in need of protection is challenging since the basic translation machinery consists of 78 factors including ribosomal proteins and aminoacyl-tRNA synthases (Gil et al., 2004). To identify candidates, the proteome of MazF-induced cells was measured as a function of time. The majority of the proteome (216 measured proteins) and 91% of 35 detectable ribosomal proteins varied by $< 10\%$ following 5 h of induction, demonstrating that highly abundant proteins were stable for hours following exposure to MazF (Fig. 3.11a). Ribosomal protein subunits S9, S20 and L17 decreased by $B20\%$ and an essential elongation factor EF-Ts decreased by approximately 80% following 5 h of induction with MazF (Supplementary Fig. 8b). In the presence of MazF, a protected version of EF-Ts (EF-Ts-P) driven by an IPTG-dependent promoter (P_{LAC}) significantly enhanced the expression of mCherry-P compared to cells that were not induced with EF-Ts-P (Fig. 3.9b). These results indicated that genome-wide measurements could be used to discover support genes in need of protection to augment resource redistribution activity.

Global mRNA decay could generate imbalances in the expression levels of genes in a regulatory network. For example, high concentrations of truncated mRNA fragments could saturate exonucleases that process these fragments into mononucleotides (Deutscher, 2006). Further, mRNA cleavage generates ribosome stalling at the 30 end of the mRNA, referred to as non-stop complexes, which require the action of ribosome recycling factors to rescue the ribosomes (Keiler, 2015). RNase R is a multifunctional protein that exhibits ribonuclease and ribosome recycling factor activities (Richards et al., 2006). Co-expression of MazF and protected version of RNase R (RNase R-P) significantly enhanced the expression of mCherry-P compared to cells expressing only MazF (Fig. 3.9b). However, co-expression of EF-Ts-P and RNase R-P did not yield an additional enhancement in the level of mCherry-P in the presence of MazF compared to cells expressing either of the single support genes, RNase R or EF-Ts-P (Fig. 3.12). These results

suggested that epistasis among support genes could potentially limit incremental improvement of resource redistribution activity.

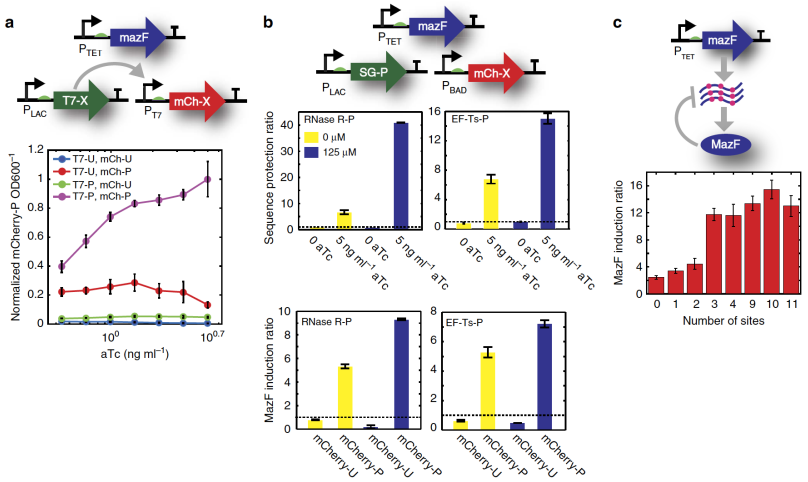

Figure 3.9: Improvement in resource redistribution activity via protection of key support genes and evaluation of the role of the MazF mRNA-decay negative feedback loop. (a) Schematic of the orthogonal T7 RNA polymerase resource allocator circuit (top). MazF, T7 RNA polymerase (T7-X) and mCherry-X were controlled by an aTc (PTET), IPTG (PLAC) and T7 (PT7) regulated promoter, respectively. Normalized fluorescence divided by OD600 as a function of aTc for cells expressing combinations of T7-U or T7-P and mCherry-U or mCherry-P following 8.3 h of induction with 1 mM IPTG (bottom). Error bars represent 1 s.d. (n = 3). (b) Schematic of support gene (SG-P) circuit (top). The support genes included protected host factors RNase R-P and EF-Ts-P. MazF, protected support genes and mCherry-X were controlled by PTET, PLAC and PBAD, respectively. We measured the expression of mCherry-P divided by OD600 in the presence and absence of MazF induction. The induction ratio is defined as the division of the former quantity by the latter. The sequence protection ratio is defined as the ratio of mCherry-P $OD600^{-1}$ to mCherry-U $OD600^{-1}$ in the presence or absence of MazF. Sequence protection ratio (middle) and MazF induction ratio (bottom) in the presence (5 $ngml^{-1}$ aTc, 125uM IPTG) or absence (0 ng ml^{-1} aTc, 0 $ngml^{-1}$ IPTG) of IPTG or aTc. Cells were induced with 0.05% arabinose for 8.3 h. Error bars represent 1 s.d. (n = 4). (c) Schematic of MazF mRNA-decay feedback loop (top). MazF induction ratio of fluorescence divided by OD600 for cells expressing mazF transcripts that varied in the number of recognition sites (P37-43 in Table 3.1). mCherry-P was regulated by a IPTG-inducible promoter (PLAC). Cells were induced with 0 or 5 ng ml^{-1} aTc and 1 mM IPTG for 9.2 h. Error bars represent 1 s.d. (n = 4).

3.4.4 MazF mRNA-decay feedback loop is important in stabilizing cell vitality

The mazF transcript is enriched for recognition sites (Fig. 3.1b), establishing an mRNA-decay negative feedback loop. We suspected that protection of MazF could enhance circuit performance. However, the feedback loop may modulate the regulatory dynamics of MazF and therefore influence resource redistribution activity. To investigate this possibility, we probed the role of the mRNA-decay feedback in the MazF resource allocator. Cells (strain S3 in Table 3.1) bearing mazF-U on a low copy plasmid (plasmid P1 in Table 3.1) controlled by an aTc-inducible promoter (P_{TET}) and induced with 5ng/mL aTc exhibited a lower steady-state mazF mRNA level compared to cells expressing mazF-P (Fig 3.13a), demonstrating that the feedback loop was actively regulating the abundance of the mazF transcript. Corroborating this result, a 35% lower threshold of aTc was required to inhibit growth in a strain expressing MazF-P compared to MazF-U (Fig 3.13b), suggesting that protection of mazF mRNA yielded a higher MazF protein level. The Hill coefficients of OD600 as a function of aTc following 11.2 h of induction were 2.6 and 5.9 for cells induced with MazF-U or MazF-P, revealing an ultrasensitive relationship between MazF activity and biomass synthesis that was significantly increased in the absence of the MazF mRNA decay feedback loop.

Contrary to expectation, cells expressing MazF-U displayed significantly higher mCherry-P expression compared to cells expressing MazF-P across a broad range of aTc concentrations, highlighting that the negative feedback loop was a critical regulatory feature for the MazF resource allocator (Fig. 3.13c). To further investigate the quantitative relationship between feedback loop strength and resource redistribution activity, we examined growth and protected reporter gene expression in cells (strain S3 in Table 3.1) bearing mazF sequences that varied in the number of recognition sites (Fig. 3.9c; Fig. 3.15). The MazF induction ratio of fluorescence divided by OD600 increased with the number of sites and the wild-type mazF sequence (nine sites) generated nearly the highest output expression level (Fig. 3.9c). In sum, these results indicated that the activity of the mRNA-decay feedback loop was a tunable knob that could be used to modulate circuit performance.

A mechanistic computational model of cellular resource allocation was constructed to provide insight into the role of the mRNA-decay negative feedback loop on circuit behavior. The dynamic model represented the

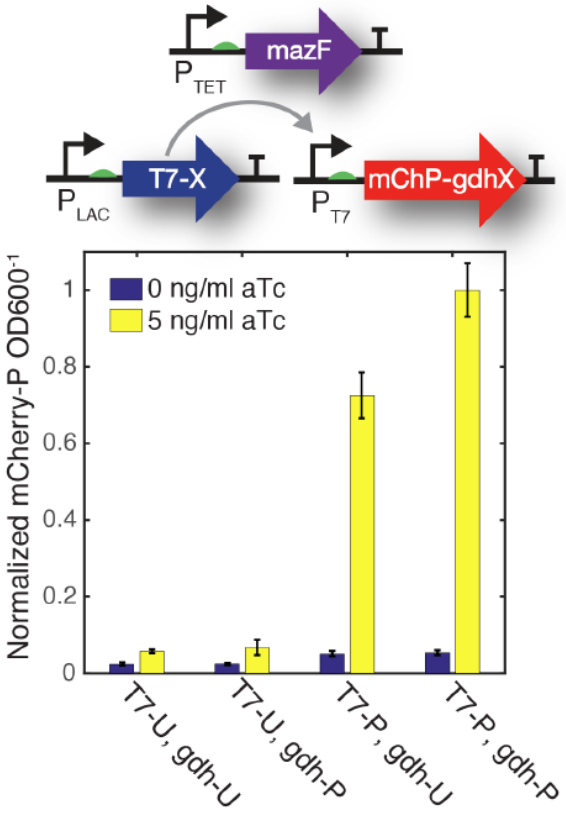

Figure 3.10: Protection of T7 RNA polymerase (T7-P) enhanced the expression of mCherry-P N-terminally fused to Gdh-X (X represents U or P) following MazF induction. Fluorescence divided by OD600 was normalized to the maximum expression across all conditions. Cells were induced with 0 or 5 ng ml^{-1} aTc and 1 mM IPTG for 10.8 hr. The media was supplemented with 1.5% glucose. Error bars represent 1 s.d. (n = 6). Strain S2 (Table 3.1) was used for this experiment. T7-X and mCherry-X were controlled by an IPTG (PLAC) and T7 (PT7) regulated promoter, respectively.

mRNA and protein levels of key species involved in the MazF resource allocator (Fig. 3.16), which compete for limiting ribosome pools including ribosomes (r), unprotected proteome (p), MazF (mazFp) and a protected reporter gene (FP). The growth rate (λ) function was based on a previous coarsegrained mechanistic model of gene expression and growth (Weiße et al., 2015). A detailed description of the model and parameters are in Tables 3.2 and 3.3.

The relationship between the mazF transcription rate α_f and the FP

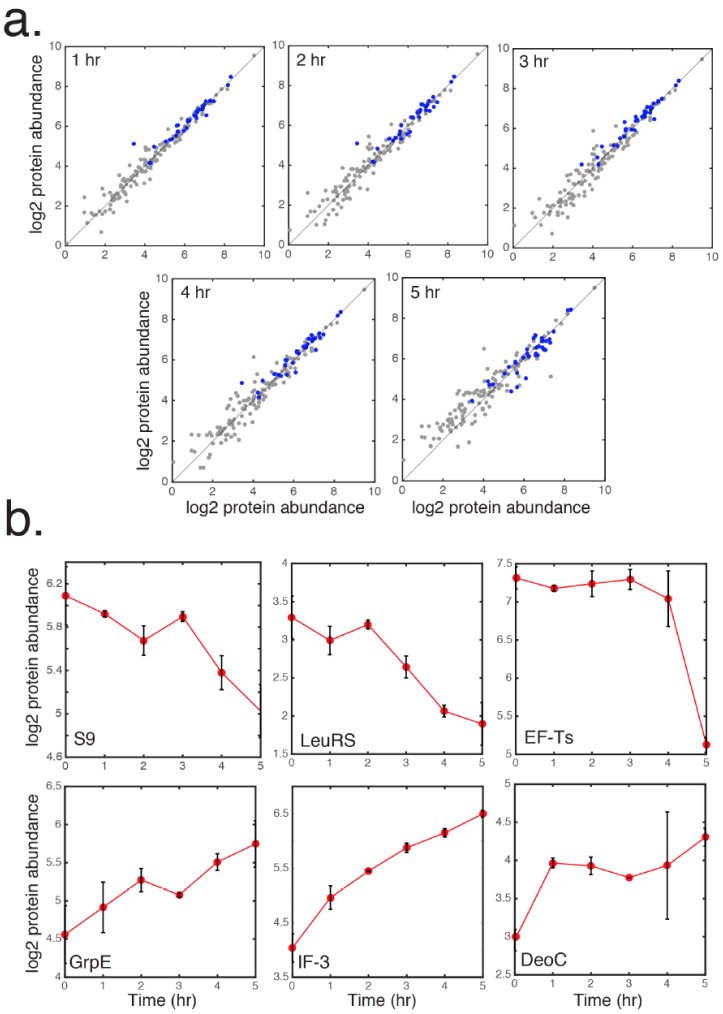

Figure 3.11: Time-series proteomics measurements of cell populations (strain S2 in Table 3.1) induced with MazF (5 ng ml^{-1} aTc). (a) Scatter plots of log2 transformed protein abundance (normalized total spectra) of 216 proteins prior to induction with MazF (x-axis) vs. exposure to MazF for 1-5 hr (y-axis). Ribosomal proteins are highlighted in blue. (b) Representative proteins that significantly decreased (top row) or increased (bottom row) in response to aTc administration (5 ng ml^{-1}). Error bars represent 1 s.d. (n = 4).

translation rate ($k_{trans}FP = k_{trans}[rFP]$) is non-monotonic (Supplementary Fig. 13a), indicating that there is an optimal expression level of MazF to maximize resource redistribution activity. The model shows that the strength of the feedback loop, represented by the dissociation constant of

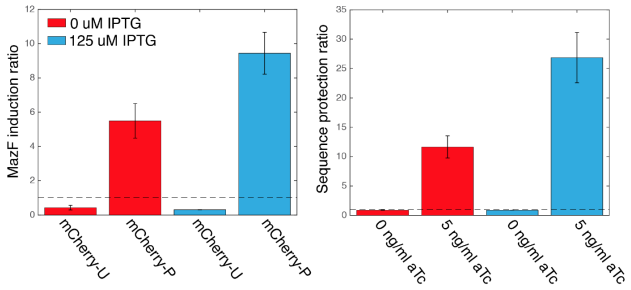

Figure 3.12: Co-expression of support genes protected RNAse R (RNAse R-P) and EF-Ts (EF-Ts-P) did not improve the resource redistribution activity compared to cells induced with a single protected support gene RNAse R-P or EF-Ts-P (Figure 3.9b).

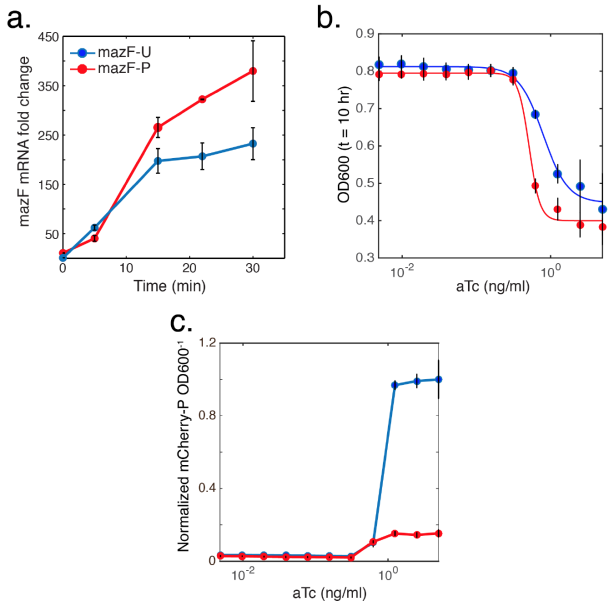

Figure 3.13: Characterization of aTc-inducible MazF-U or MazF-P on mazF mRNA dynamics, growth and mCherry-P expression. MazF and mCherry-P were regulated by an aTc (PTET) and arabinose-inducible promoter (PBAD). (a) mazF mRNA fold change measured by qPCR as a function of time for cells induced with MazF-P or MazF-U. Error bars represent 1 s.d. (n = 2). (b) Saturating cell density (OD600) as a function of aTc following 11.2 hr of induction. Error bars represent 1 s.d. (n = 4). (c) Normalized expression of mCherry-P as a function of aTc. Cells were induced with 0.05% arabinose and a range of aTc concentrations for 11.2 hr. Strain S3 and plasmids P1-2,4 were used for this experiment. Error bars represent 1 s.d. (n = 4).

MazF dimer ($mazFpd$) to the mazF transcript $m_f(KD_f = k_{rff}k_f^{-1})$, is inversely correlated with the dose-response ultrasensitivity of total steady-state MazF concentration ($mazF_T = 2 \times [pf]_{ss} + 2 \times [rf]_{ss} + 2 \times [ff]_{ss} + 2 \times [fe]_{ss} + 2 \times [mazFpd]_{ss} + [mazFp]_{ss}$, where ss denotes steadystate) as a function of α_f (Fig. 4a,b). Molecular mechanisms that realize ultrasensitivity include MazF dimerization (G.-Y. Li et al., 2006), molecular sequestration (Buchler et al., 2008; Ophelia S Venturelli et al., 2012) of mRNAs by ribosomes (De Vos et al., 2011) or positive feedback (Ferrell Jr et al., 2014). In addition, thresholded control of l by mazFT, which was observed in our experimental and modelling data (Fig. 3.14d; Fig. 3.13b), could contribute to ultrasensitivity in the network. For high KDf corresponding to the open loop system, the model exhibits bistability manifesting as two stable steady states across a range of α_f values (Supplementary Fig. 13b). Since m_p and m_r compete for limiting ribosome pools (Supplementary Fig. 13c), bistability could arise via positive feedback (Ophelia S Venturelli et al., 2012) established by an increase in the synthesis rate of r as a consequence of MazF-dependent m_p decay. The MazF mRNA-decay negative feedback loop enables proportional adjustment of the mazFT (Nevozhay et al., 2009) and reduces the potential for bistability by abolishing ultrasensitivity (Ophelia S Venturelli et al., 2012; Tiwari et al., 2012) (Fig. 3.14b). As such, mazFT concentration could be tuned to operate in the regime that maximized resource redistribution activity.

For a fixed value of α_f, ktransFP is inversely related to KD_f (Fig. 4c), qualitatively recapitulating the increase in mCherry-P with the number of binding sites in the mazF transcript (Fig. 3c). λ and the total concentration of the unprotected gene p decrease as a function of α_f, mirroring the experimental data that showed lower OD600 and mCherry-U in the presence of aTc (Figs 3.3, 3.4a and 3.13b; Fig. 3.14d,e). The increase in ultrasensitivity of the dose response of mazFT versus λ as a function of KD_f (Fig. 3.14d) qualitatively reflected the enhanced ultrasensitivity of the steady-state dose response of aTc versus biomass (OD600) for cells expressing MazF-P compared to MazF-U (Fig. 3.13b). The negative feedback loop strength is inversely related to the range of α_f values that enhance total steady-state r concentration (rT, Fig. 3.14f). Above a threshold value of KD_f, r_T decreases monotonically with α_f. The mRNAdecay negative feedback has important implications for resource allocator design by enabling precise tuning of the MazF operating point by establishing a proportional relationship between α_f and mazFT. Indeed, this negative feedback may provide an evolutionary

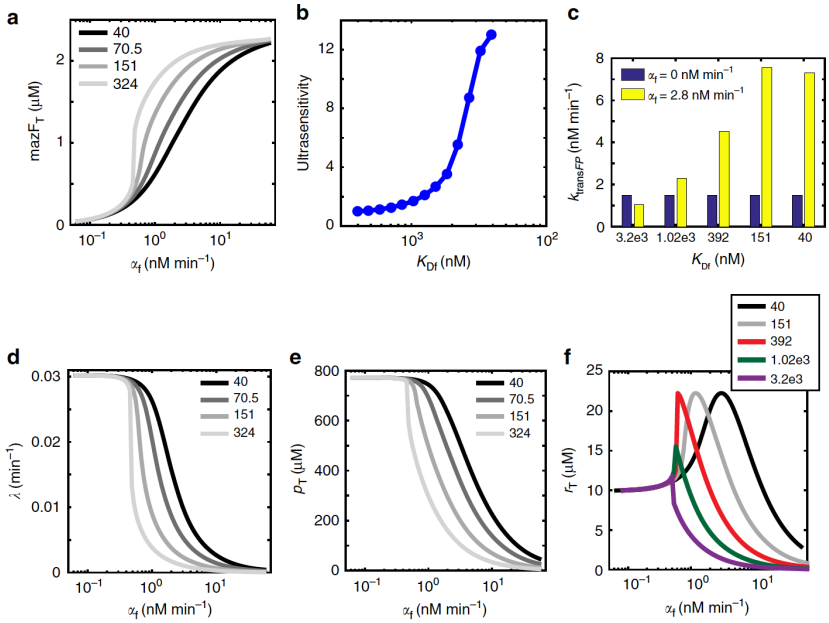

Figure 3.14: Probing the role of the MazF negative feedback loop in a dynamic computational model of resource allocation. This model demonstrates that the MazF mRNA-decay feedback loop established proportional control of MazF in the absence of MazE ($\alpha_e = 0$). (a) Total MazF concentration at steadystate (mazFT, t=278 h) as a function of the transcription rate of mazF (α_f) across a range of dissociation constants (KDf) in units of nM of MazF to mazF mRNA (m_f). Here, mazFT = $2 \times [pf]_{ss} + 2 \times [rf]_{ss} + 2 \times [ff]_{ss} + 2 \times [fe]_{ss} + 2 \times [mazFpd]_{ss} + [mazFp]_{ss}$, where ss denotes steady-state. (b) Maximum logarithmic sensitivity (ultrasensitivity) of the dose response of af versus mazFT across a range of KDf values. (c) Steady-state translation rate of a protected gene FP (ktransFP = ktrans[rFP]ss) as a function of KDf in the presence ($\alpha_f = 2.8$ nMmin^{-1}) or absence ($\alpha_f = 0$ nMmin^{-1}) of MazF. (d) Steady-state growth rate (l) as a function of α_f for different values of KDf. (e) Steady-state total unprotected proteome (pT) concentration as a function of α_f for different values of KDf. (f) Steady-state total ribosome concentration (rT) as a function of α_f for different values of KDf.

advantage for cells by preventing the deleterious effects of MazF overexpression that accelerated cell death (Fig. 3.18).

3.4.5 MazE feedback loop impacts growth and circuit properties.

Transcriptional profiling and proteomics measurements of MazF-induced cells (strain S2 in Table 3.1) revealed that the mazE transcript (Fig. 3.20a) and MazE protein (Fig. 3.20b) were up-regulated by aTc administration.

MazE is a stoichiometric inhibitor of MazF activity by sequestering MazF into an inactive complex (Kamada et al., 2003). Stimulation of MazE synthesis in response to MazF activity establishes a molecular sequestration negative feedback loop. The protein abundance of MazF significantly exceeded MazE, explaining the lack of MazF inhibition in these conditions (Supplementary Fig. 3.20b). Since MazE could be used to control the activity of the MazF resource allocator, we examined the impact of MazE activity on growth and circuit properties in the model.

The transcription rate of mazE was a function of active MazF (mazFpd) in the model to capture the coupling between MazF induction and MazE synthesis. Increasing the maximum mazE transcription rate α_e reduced the total active MazF concentration (total active MazF concentration was defined as $[pf]_{ss} + [rf]_{ss} + [ff]_{ss} + [fe]_{ss} + [mazFpd]_{ss}$, where ss denotes steady-state; Fig. 3.21a). As a result, a higher α_f was required to fully inhibit cell growth in the presence of MazE (Fig. 3.21b). Increasing α_e shifted the regime of maximum resource redistribution activity towards higher α_f values (Fig. 3.21c). Ultrasensitivity in the steady-state dose response of α_f versus total MazF ($mazFT = 2 \times [pf]_{ss} + 2 \times [rf]_{ss} + 2 \times [ff]_{ss} + 2 \times [fe]_{ss} + 2 \times [mazFpd]_{ss} + 2 \times [cef]_{ss} + [mazFp]_{ss}$) was moderately enhanced by up to $\sim 23\%$ in a narrow parameter regime corresponding to high KD_f and intermediate α_e values, presumably via molecular sequestration (Fig. 3.21d) (Buchler et al., 2008). However, ultrasensitivity was significantly reduced across a broad range of α_e values. The range of α_f that mapped to high resource distribution activity could be adjusted by modulating both the MazE and MazF mRNA-decay feedback loops. However, in contrast to the mRNA-decay feedback, increasing the strength of the MazE feedback moderately reduced the parameter range that mapped to optimal circuit performance (Fig. 3.21c). In sum, MazE is a key control parameter for the MazF resource allocator that can be used to rapidly modulate growth and resource redistribution activity (Amitai et al., 2004).

3.4.6 Transcriptional profiling of MazF-induced cells show suppression on replication activity

To evaluate the genome-wide variation in transcript abundance following MazF exposure, RNA-seq measurements of MazF-induced cells were collected every 2 min for a total of 8 min using strain S2 induced with 5 ng ml^{-1} aTc (Supplementary Table I). The majority of the 192 endogenous protected genes increased or remained constant following induction with MazF

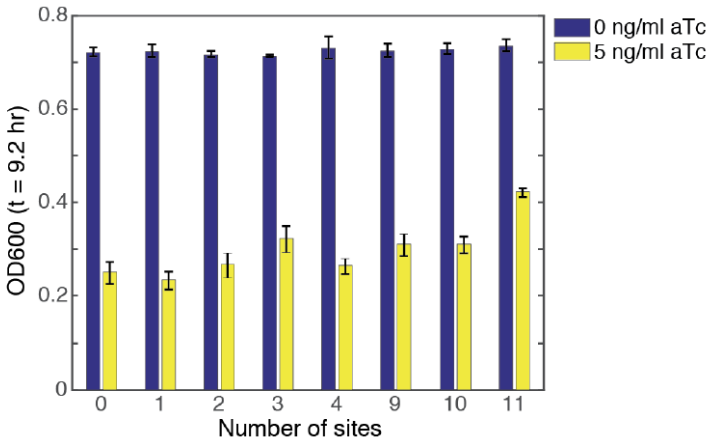

Figure 3.15: Saturating cell densities (OD600) of cell populations in the absence (0 ng ml^{-1} aTc) or presence (5 ng ml^{-1} aTc) of induction with a set of mazF variants. MazF was regulated by an aTc-inducible promoter (P_{TET}). The x-axis indicates the number of MazF recognition sites in the mazF mRNA sequence (plasmids P37-43 in Table 3.1). Cells were induced with 0.05% arabinose for 9.2 hr. Error bars represent 1 s.d. (n = 4).

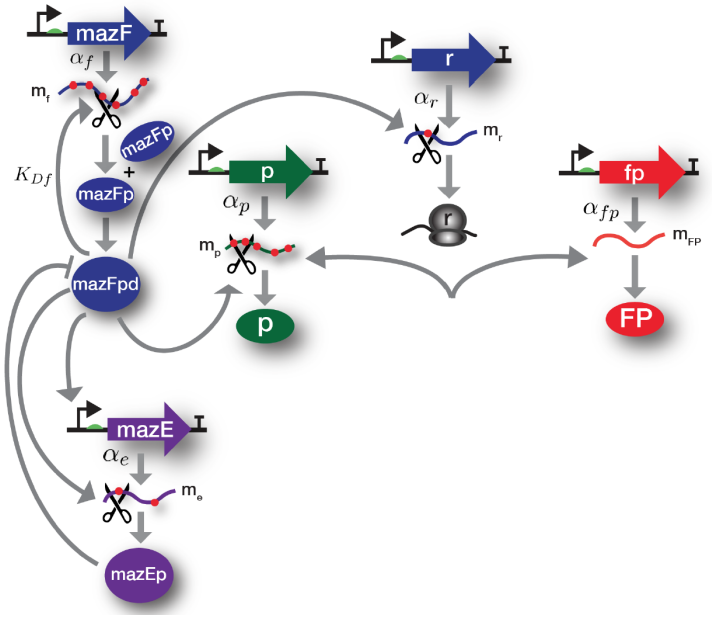

Figure 3.16: Schematic of the MazF resource allocator circuit.

for 8 min (Fig. 3.19a). A balance between synthesis and decay catalysed by RNases and MazF determines transcript abundance. Therefore, it is challenging to directly decipher the MazF-dependent transcript decay rates. Nevertheless, the number of MazF sites was negatively correlated with the mean log2 fold change of transcript abundance following 8 min of induction with aTc, indicating that on average the number of MazF sites predicted the fold change across the transcriptome (Fig. 3.19b,Fig. 3.22).

Partitioning the transcriptome fold change dynamics into three clusters (see Methods) revealed three temporal patterns in transcript abundance in response to MazF induction: downregulation (K1, 460 genes), pulsatile response characterized by an increase in transcript abundance at early times and decrease following a delay (K2, 148 genes) or up-regulation (K3, 331 genes, Fig. 3.19c). We evaluated functional or regulatory enrichment ($P < 0.05$ using the Fisher's exact test) in each cluster to provide insights into the physiological impact of MazF exposure (Table 3.5). Cell envelope and genes regulated by Fur, MraZ and LexA were enriched in the K1 cluster (Fig. 3.19c; Fig. 3.23). MraZ is a transcriptional repressor that controls many genes involved in cell division and cell wall biosynthesis (Mengin-Lecreulx et al., 1998). In addition, the cell division regulator minE mRNA decreased significantly in the RNA-seq data (Fig. 3.19a), corroborating a link between MazF activity and inhibition of cell division (Hale et al., 2001; Eraso et al., 2014). The K2 cluster was enriched for genes regulated by NikR, GlpR, GcvA, IHF, IscR and RstA and amino acid and anaerobic metabolism (Fig. 3.23). K2 contained numerous regulatory categories (Table 3.5), suggesting that the pulsatile transcript dynamics could be established by an early increase in synthesis rates and delayed down-regulation due to mRNA-decay at a threshold concentration of MazF. Genes regulated by ArgR were enriched in the up-regulated cluster K3. In addition, 11 TCA cycle enzymes were up-regulated in the RNA-seq data (P = 0.051 enrichment in K3), suggesting that MazF-induced cells exhibited high metabolic activity (Fig. 3.24; Table 3.5).

Previous work has demonstrated that fumarate production increased the frequency of persister cells following antibiotic exposure (J.-S. Kim et al., 2016). A closer examination of the catabolic pathway revealed that fumarate producing enzymes were significantly induced, illustrating a connection between MazF activity and persistence via enhancement of fumarate flux (Mok et al., 2015; Tripathi et al., 2014) (Fig. 3.24).

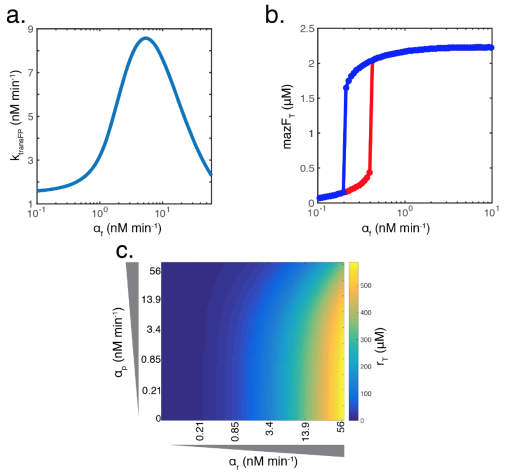

Figure 3.17: Non-monotonic circuit performance, trade-offs and bistability in the dynamic resource allocation model in the absence of MazE.

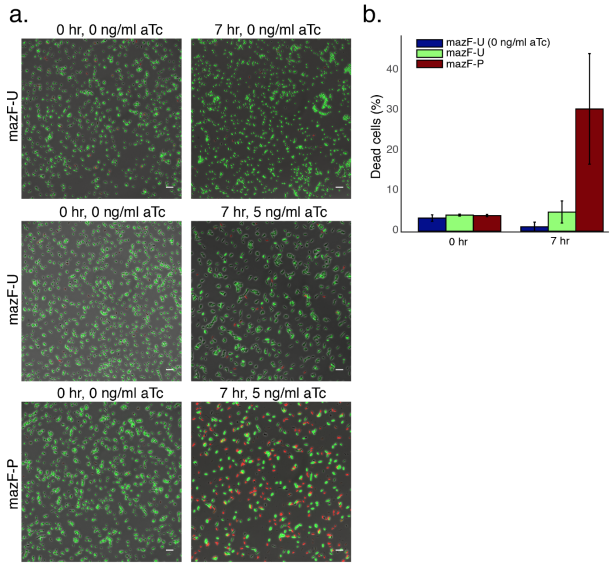

Figure 3.18: Characterization of cell viability using the LIVE/DEAD assay. Strain S3 bearing plasmid P1 or P2 were used in this experiment (Table 3.1). (a) Representative fluorescence microscopy images of cells uninduced (0 ng ml^{-1} aTc) or induced (5 ng ml^{-1} aTc) with MazF-U or MazF-P for 0 or 7 hr. Red (propidium iodine) and green (SYTO 9) represents cells with compromised and intact membranes, respectively. Scale bars represent 10 μm. (b) Ratio of the number of cells with compromised membranes over the total number of cells for each condition following 0 (n = 3502, 3413, 4844 cells from left to right) or 7 hr of induction (n = 2455, 1706, 1653 cells from left to right). Error bars represent 1 s.d. (n = 2).

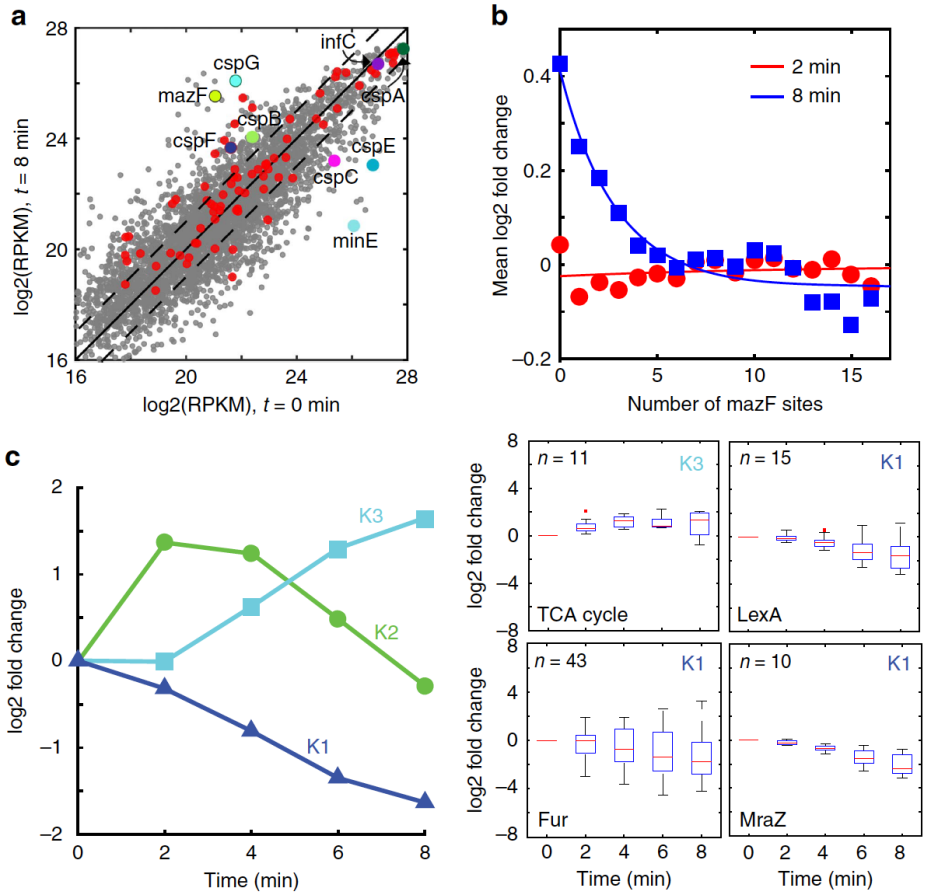

Figure 3.19: Time-series RNA-seq measurements of MazF-induced cells. The mean RPKM value (n = 2) was log2 transformed. (a) Scatter plot of log2 transformed RPKM measurements before induction with MazF versus induction with MazF (5 ng ml^{-1} aTc) for 8 min. Grey and red data points denote unprotected or protected transcripts larger than 80 nucleotides, respectively. Dashed lines represent a two-fold threshold in transcript abundance. cspABCGEF, mazF and minE transcripts are highlighted. (b) Scatter plot of the number of mazF sites for each gene versus mean log2 fold change following induction with 5 ngml^{-1} aTc for 2 or 8 min. A 5-point moving average was applied to the data. Lines represent fitted exponential functions to the data. (c) K-means clustering of log2 fold change of 939 genes (left) that exhibited correlated dynamics between biological replicates. Box plots (right) of representative functional or regulatory enrichments in the K1 and K3 clusters according to the Fisher's exact test (Po0.05). On each box, the red line indicates the median, the bottom and top edges represent the 25th and 75th percentiles and 'þ' denote outlier data points. n represents the number of genes in each category (Supplementary Table V)

Cold-shock genes are selectively expressed in response to cold stress

and perform diverse functions including unwinding of RNA secondary structures, modulation of ribosome and DNA/RNA chaperone activity (Gualerzi et al., 2003). The transcriptional profiling data revealed significant shifts in cold-shock cspBCEFG and associated rbfA, rhlB, rhlE and deaD transcript abundance as a function of time (Fig. 3.25). IF-3, one of the major translation factors in *E. coli*, has been shown to mediate cold shock translational bias in response to cold stress (Giuliodori et al., 2007; Gualerzi, 2011). IF-3 increased over four-fold in the proteomics data (Fig. 3.11b) following 5 h of MazF induction, whereas the abundance of infC mRNA did not change significantly in response to MazF activity (Fig. 3.19a). Future work should interrogate the molecular mechanisms and functional connection among MazF activity, up-regulation of IF-3, and significant shifts in cold-shock transcript abundance.

As cold-shock transcripts were up-regulated in response to MazF activity, these sequences were promising candidates for engineering MazF-responsive promoters. To test the modularity of cold-shock induction by MazF, we constructed a tandem promoter composed of PLAC upstream of the cspB or cspG promoter, UTR and the first 14 amino acids of CspG or CspB N-terminally fused to sfGFP-P. MazF induction increased sfGFP-P by a maximum of 20 or 80-fold, demonstrating that the cspB and cspG regulatory sequences are modular control elements that directly respond to MazF activity as an input.

3.4.7 MazF cleavage efficiency is position and frequency dependent of the recognition sequence

A quantitative understanding of the mapping between MazF site placement and cleavage efficiency could enable tuning of the timing and degrees of protection to inform resource allocator design. Previous work demonstrated that MazF activity was inhibited by strong secondary structures and ribosomes enhanced cleavage efficiency by unwinding mRNA secondary structures during translation (Christensen-Dalsgaard et al., 2008). To explore the dominant parameters that influence MazF cleavage efficiency, we varied the number and position of MazF recognition sites in the mCherry transcript (plasmids P21-36) in the S2 background strain (Table 3.1).

To map the relationship between position and cleavage efficiency, a single MazF site was inserted at 14 positions in mCherry-P (Fig. 3.26). These mCherry sequences exhibited a broad range of expression levels in response

to MazF (Fig. 3.26a). The output was correlated with the predicted secondary structure Gibbs free energy (ΔG) 38–47 bp upstream of the recognition site (ρ ranged between -0.7 to -0.5, P < 0.05 using the Student's t-test) computed using NUPACK (Fig. 3.26b,c). For sequences spanning upstream and downstream of the MazF site, mCherry expression was correlated (ρ = -0.6, P = 0.05 using the Student's t-test) with ΔG (39–40 bp, Fig. 3.26d). However, the ΔG of the sequence downstream of the recognition site was not correlated with the expression level of mCherry across a broad range of window sizes (Fig. 3.26e). Therefore, MazF cleavage efficiency could be predicted using the folding energy of the local mRNA secondary structure upstream or across the recognition site.

To provide insight into the programmability of MazF cleavage efficiency, we interrogated whether measurements of mCherry variants containing a single MazF site (Fig. 3.26a) could predict the expression of mCherry sequences containing combinations of sites. mCherry expression decreased as a function of the number of recognition sites in the presence of MazF (Fig. 3.26f). The product of the single site mCherry expression levels could predict the expression of the multi-site variants (P <4e-6 using the Student's t-test), suggesting that combinations of MazF recognition sites could be used to modulate the degree of transcript protection.

3.5 Discussion

A major goal of synthetic biology and metabolic engineering is to develop strategies to control the resource economy of cells for switching between modes of growth and production (Venayak et al., 2015). During a production phase, cellular energy and resources are focused on specific pathways, while minimizing resource expenditure towards nonessential cellular operations. Towards these objectives, previous work leveraged tunable enzymatic degradation of a metabolic hub that determines the direction of metabolic flux to augment the yield and titre of a metabolic pathway twofold (Brockman et al., 2015). While this strategy provided localized control of metabolic flux, it does not modulate the global allocation of subsystems such as transcription and translation. On a larger scale, inducible regulation of RNA polymerase subunits was recently used to control *E. coli* growth (J.-S. Kim et al., 2016). However, this mechanism cannot be generally applied to redirect resources towards engineered networks.

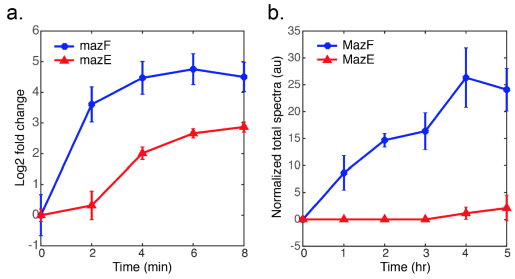

Figure 3.20: Temporal variation in the transcript and protein levels of MazE and MazF in cells (strain S2 in Table 3.1) induced with 5 ng ml^{-1} aTc. MazF is regulated by an aTc-inducible promoter (P_{TET}).

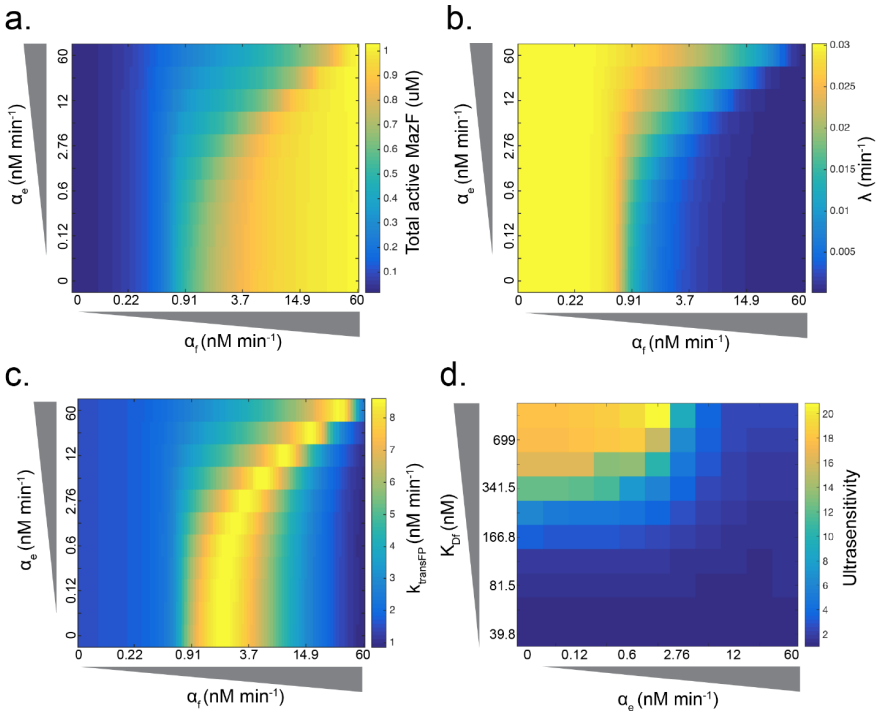

Figure 3.21: The MazE negative feedback loop influences circuit properties and growth. (a) Steady-state total active MazF. The binding affinity of MazF to the mazF transcript (KDf) was equal to 116.6 nM. (b) Steady-state growth rate (λ) as a function of α_f and α_e for KD_f = 116.6 nM. (c) Steady-state translation rate of a protected gene $F_P(k_{trans}F_P)$ across a range of α_f and α_e values for KD_f = 116.6 nM. (d) Maximum logarithmic sensitivity (ultrasensitivity) of the dose response of α_f vs. total steady-state MazF (mazFT) concentration across a range of α_e and KD_f values.

Here, we showed that synthetic circuits could exploit shifts in cellular physiological state due to MazF activity, suggesting that intracellular resources could be diverted via programmable mRNA decay. This approach could be harnessed for diverse applications by protecting genes in an engineered network and systematically discovering key support factors beyond the engineered pathway in need of protection. Recent advancements in DNA synthesis technologies will facilitate large-scale recoding of support genes to protect from MazF activity. A utility of this approach is to enhance target functions that compete directly with biomass synthesis, such as exploiting microbes as 'cell factories' to synthesize chemicals or biomolecules of interest. Further, MazF activity could potentially minimize unintended environmental impact due to cell proliferation, while allowing engineered cells to carry out a desired function in a complex environment. Coupling this strategy to dynamic regulation of MazE would enable periodic resuscitation of cellular sub-systems and maintain metabolic activity over longer time scales. MazF regulates orders of magnitude more genes simultaneously compared to other technologies such as CRISPRi (Peters et al., 2016; Yao et al., 2015). Homologues of MazF that recognize 3, 5 and 7-bp recognition sites have been identified in diverse bacterial species (Yamaguchi et al., 2011; J. H. Park et al., 2011; Yamaguchi et al., 2012).

Active site mutations have been shown to modify the MazF sequence specificity, suggesting that protein engineering could be used to expand the diversity of MazF recognition site sequences (Ishida et al., 2013). The variation in recognition sequence specificity could be used to tune the number of genes targeted by MazF. In addition to the unknown myriad effects of MazF-induction on network activities, there are several limitations to optimizing the MazF resource allocator. MazF activity increased the abundance of a set of host-cell transcripts (cluster K3 in Fig. 5c), which sequesters resources away from engineered circuits. However, this activation programme could be exploited by repurposing regulatory elements that respond to MazF activity to expand the resource allocator design. In addition, MazF activity has been shown to yield a heterogeneous ribosome pool by targeting a specific site of the 16S rRNA (Vesper et al., 2011), which could manifest as translation bias for selected transcripts (Sauert et al., 2016). Decay of the unprotected proteome occurs on the time scale of hours, thus limiting the time scale required to shift metabolic flux. To rapidly manipulate metabolic flux, induction of MazF could be coupled with proteases (Cameron et al., 2014) for targeted control of protein abundance. As

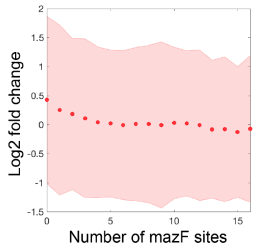

Figure 3.22: Scatter plot of the number of mazF sites vs. mean log2 fold change in transcript abundance following administration of 5 ng ml^{-1} aTc for 8 min to strain S2

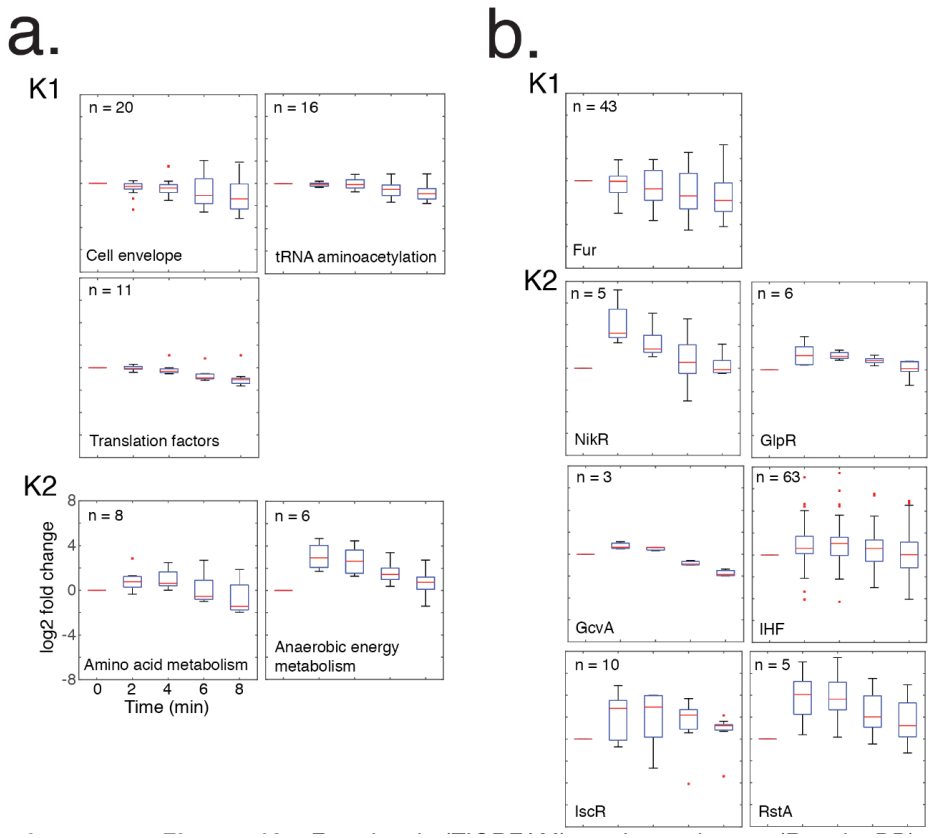

Figure 3.23: Functional (TIGRFAM) and regulatory (RegulonDB) gene enrichment using a Fishers exact test in RNA-seq clusters shown in Figure 3.19c. (a) Box plots showing the RNA-seq log2 fold change (y-axis) as a function of time (x-axis) for each category. On each box, the red line indicates the median, the bottom and top edges represent the 25th and 75th percentiles and '+' denote outlier data points. n denotes the number of genes in each category. (a) Functional enrichments in clusters ($p < 0.05$ using the Fisher's exact test). (b) Regulatory enrichments in clusters ($p < 0.05$ using the Fisher's exact test).

the proteome decays, stoichiometric relationships required for protein activity must be maintained (Amitai et al., 2004). Further, MazF has been shown to establish a futile cycle of continuous RNA synthesis and decay, resulting in energy dissipation (Mok et al., 2015). To minimize an energy deficit, orthogonal T7-P could be used to drive the engineered pathway, while at the same time inactivating native RNA polymerases.

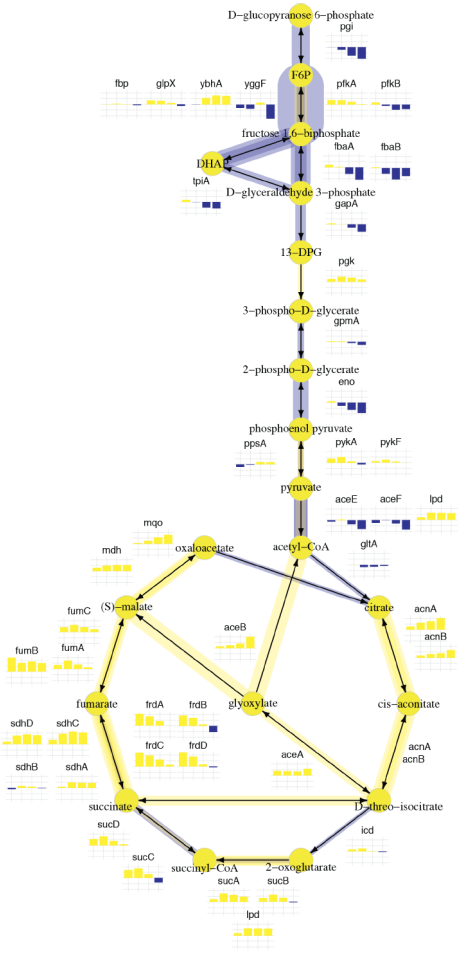

Figure 3.24: Schematic of the glycolysis and TCA cycle metabolic network highlighting the RNA-seq log2 fold-change following 8 min of induction with MazF

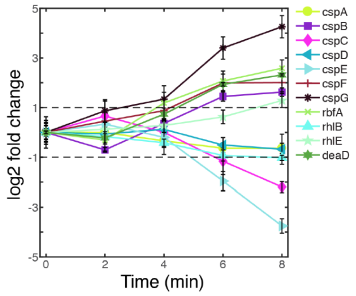

Figure 3.25: RNA-seq log2 fold change of transcript abundance of cold-shock associated genes as a function of time.

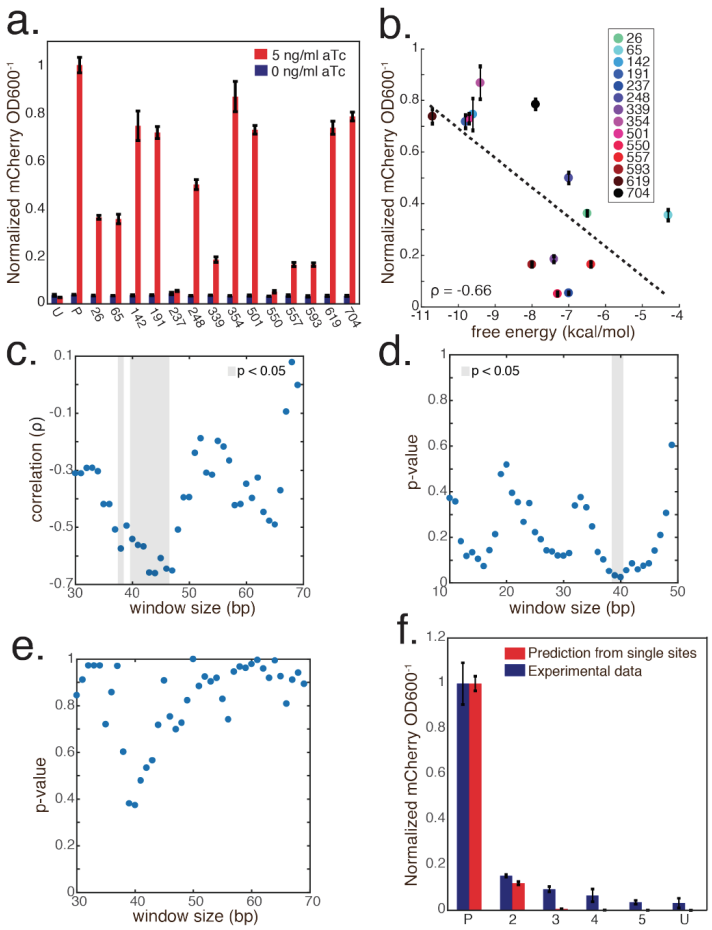

Figure 3.26: The expression levels of 14 mCherry sequences containing a single MazF recognition site at different positions were correlated with local secondary structure upstream or across the recognition site. P and U denote mCherry-P or mCherry-U. MazF and respectively. (a) mCherry variants containing a single MazF recognition site at different positions exhibited a broad range of expression levels in response to MazF. (b) The expression levels of mCherry variants were correlated with the Gibbs free energy (ΔG) of the predicted RNA secondary structure calculated using NUPACK (c) Scatter plot of the correlation coefficient between ΔG vs. mCherry expression across a range of upstream window sizes. Shaded regions denote a statistically significant correlation coefficient ($p < 0.05$ using the Student's t-test). (d) Scatter plot of the p-values of the correlation between ΔG and mCherry expression across a range of window sizes spanning upstream and downstream of the MazF recognition site. (e)Scatter plot of the p-values of the correlation between ΔG and mCherry expression across a set of window sizes downstream of the MazF recognition site. (f) Comparison of experimental measurements and predicted expression levels of mCherry sequences containing multiple MazF recognition sites.

Cells have evolved numerous feedback mechanisms to optimize ribosome concentrations to match changes in environmental conditions, including nutrient quality and abundance (Klumpp et al., 2009; M. Scott et al., 2010). These growth-rate dependent couplings to cellular processes including transcription, translation and replication can influence the behaviour of synthetic circuits. In MazF-induced cells, the consequences of growth rate inhibition on cellular sub-systems remain unresolved. The stringent response is not activated in MazF-induced cells, which allows cells to maintain ribosome synthesis and cellular maintainence (Germain et al., 2015). A detailed understanding of network activities and resource partitioning in MazF-induced cells will allow for exploitation of this unique physiological state for diverse biotechnological applications. Top-down approaches such as MazF could be used to discover host factors that preserve high metabolic activity in the absence of growth. Genome engineering could be used to protect these pathways from MazF activity (Bassalo et al., 2016). Optimal regulatory strategies should be designed to balance enhancement of resource redistribution activity and degradation of cellular support subsystems over long time scales. For example, MazF could be transiently induced until energy degrades to a threshold that triggers rapid inhibition of MazF activity via MazE and allows rebalancing of the proteome (Rebbapragada et al., 1997). Altogether, advances in regulatory control strategies and large-scale recoding may enable the design of protected and unprotected orthogonal sub-genomes that dynamically switch between cellular operations.

3.6 Conclusions

In conclusion, we demonstrated that MazF activity induces a global cellular physiological shift that can be exploited to enhance synthetic circuit expression. These results suggest that the MazF resource allocator controllably redistributed core cellular subsystems to support a synthetic circuit and an engineered metabolic pathway. The former is further enhanced by protection of specific host-cell factors and use of the orthogonal RNA polymerase from T7 bacteriophage (T7 RNA polymerase) to transcribe genes in the synthetic circuit. We used shotgun proteomics to identify a host factor in need of protection to prevent loss of translational efficiency following MazF induction.

Our results demonstrate that the activity of the mRNA-decay feedback loop is a critical parameter for the resource allocator. In sum, these results suggest a platform for global manipulation of resource pools as a key parameter for modulating synthetic circuit behaviour. This will be a step forward in expanding the synthetic biology toolbox towards achieving optimal resource distribution in synthetic circuits.

3.7 Acknowledgments

We would like to thank Karen Lundy for constructing the RNA-seq libraries and Kristala Prather (MIT) for providing an E. coli strain for the gluconate measurements. This work was supported by the US Department of Energy (Grant DE-SC0008812) and used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. O.S.V. was supported by the Simons Foundation at the Life Sciences Research Foundation postdoctoral fellowship.

Part II

Multi-strains Engineering

Chapter 4

Designing spatially distributed gene regulatory networks to elicit contrasting patterns

4.1 Attribution

This chapter is based on a first-authored publication in ACS Synthetic Biology:
Tei M, Perkins M, Hsia J, Arcak M, Arkin AP. Designing spatially distributed gene regulatory networks to elicit contrasting patterns. ACS Synthetic Biology 2019.

4.2 Introduction

Over the last two decades, synthetic biologists have sought to engineer microbes and their consortia to execute ever more complex tasks. These range from the relatively straightforward production of valuable chemicals to the computation of complex logics that allow the microbes to make sophisticated decisions to optimize such production (Bokinsky et al., 2011; Dahl et al., 2013) or release therapeutics *in situ* (J. C. Anderson et al., 2006; C. J. Anderson et al., 2007). There has been sustained interest in both expanding the intrinsic size of these "circuits" for implementation of even more ambitious functions and spreading these circuits among different members of a population that distributes the production load (Klavins, 2014; Tsoi et al., 2018), allows reuse of components (Tamsir et al., 2011), or better utilizes space (Fernandez-Rodriguez et al., 2017; Prindle et al., 2012) and growth (Kong et al., 2018). Certainly, in nature, microbial populations and consortia spatially arrange themselves to form specialized structures that have

mechanical, developmental, and chemical advantages over homogeneous distributions (Mousa et al., 2016).

Ideally, spatial patterning arises in a self-organized fashion from control among individual cells, but shaping communities of cells remains a challenge to engineer. In natural systems, this organization often relies on highly specific communication among cells and/or among highly spatially constrained signaling such as cell contact-mediated mechanisms (Toda et al., 2018). In bacteria, intercellular signaling systems like AHL used in QS (Federle et al., 2003) tend to be fairly non-specific and long range, while contact-mediated systems like Cdi-A/B (Morse et al., 2012; E. C. Garcia et al., 2016) are not easily programmable. However, it is possible to experimentally explore the principles of spatial organization among microbial populations by imposing external constraints on communication.

Here, we formulate a simple extensible modeling framework that provides a systematic way to represent the dynamics of multi-strain communities, and design a compartmentalized culturing platform to control the spatial arrangement of bacterial colonies harboring different genetic circuits constrained to communicate through specified channels on the device. We demonstrate the utility of the approach by applying the framework to the canonical example of lateral inhibition (Collier et al., 1996) to predict the emergence of stable contrasting patterns.

4.3 Materials and Methods

4.3.1 Bacterial strains, plasmid construction, and growth conditions

E. coli strain DH10B (NEB) was used for cloning. PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (Thermo) and oligonucleotides (IDT). BsaI (NEB) and T7 DNA ligase (NEB) were used to construct plasmids using parts obtained from the MIT Registry of Standard Biological Parts, JBEI registry (T. Lee et al., 2011), or synthesized gBlocks (IDT). RBS calculator (Salis et al., 2009) was used to generate balanced RBS strengths for $luxI$ and $lasI$. TR117 (gift of Thomas L. Ruegg) is a DH10B variant with genomically integrated mRFP1 driven by a constitutive promoter. MOPS EZ Rich Medium (Teknova) and MOPS with 1.5 % UltraPure

agarose (Thermo) were used for liquid and solid medium. When appropriate, 50 $\mu g/mL$ Kanamycin or 20 $\mu g/mL$ Chloramphenicol were added to medium.

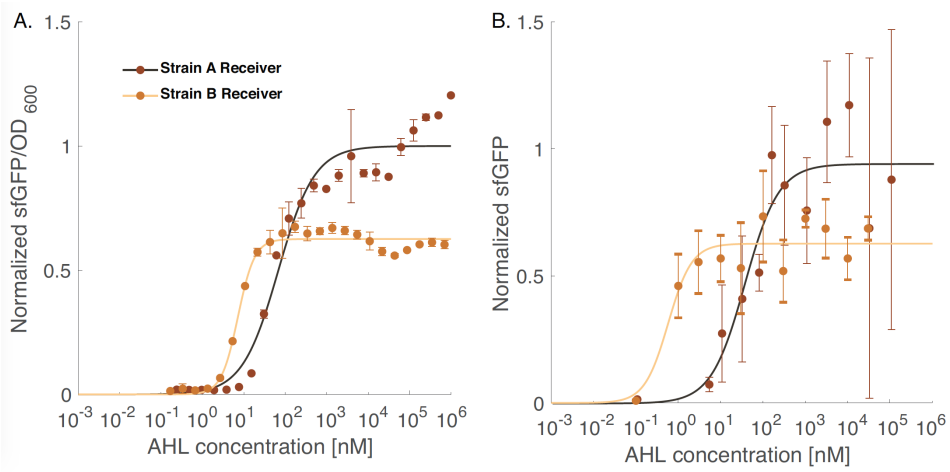

Figure 4.1: Plate reader assays of output steady-state TetR-sfGFP fluorescence in response to input AHL concentration for the reception modules of Strains A and B in liquid (a) and solid (b) medium. The strains used for reception module characterization lack $ptet \rightarrow luxI$ or $lasI$. Steady-state sfGFP fluorescence (t = 10 h) divided by OD600 in liquid medium and steady-state sfGFP fluorescence ($t = 10\ h$) at the center of colonies on solid medium were each normalized to the maximum steady-state values across strains and input AHL conditions. Strain A receiver showed similar threshold AHL concentrations regardless of medium and Strain B receiver showed more AHL-sensitive response in solid than in liquid medium. In both conditions, the approximated K_d of Strain A reception module was higher than Strain B, suggesting higher sensitivity of $plas$ in Strain B than $plux$ in Strain A. Error bars show standard deviation and solid circles show the average of the measurement (n = 2). The solid curves show best-fit models when measurements were fit to activation Hill functions in the form of $a\frac{[AHL]}{([AHL]+K_d^n}$, where a denotes the maximal production of TetR-sfGFP, K_d denotes the apparent dissociation constant of AHL binding to the promoter and n denotes the apparent Hill coefficient. The following equation parameters were used for the best-fit models: $a = 1$ (Strain A) and 0.771 (Strain B), $K_d = 47.5\ nM$ (A) and $8.33\ nM$ (B), $n = 1.91$ (A) and 1.83 (B) for liquid medium, $a = 0.9423$ (A) and 0.6269 (B), $K_d = 34.71\ nM$ (A) and $0.577\ nM4(B)$,n=4 1.088 (A) and 1.556 (B) for solid medium.

4.3.2 Plate reader assays

Overnight cultures of cells in MOPS were washed three times and diluted to fresh MOPS at OD600 of 0.3. After 8 hours in 30^{o}C at 750 rpm, cells were

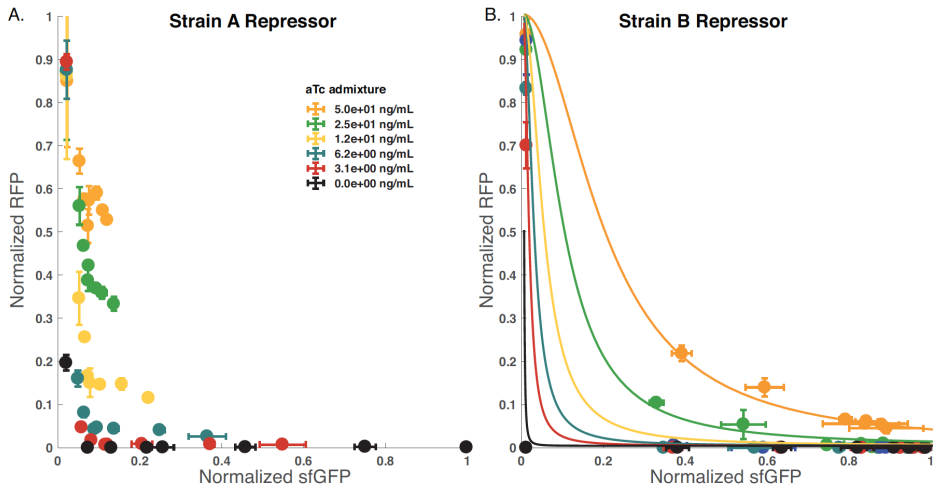

Figure 4.2: Plate reader assays of output $ptet \rightarrow mRFP1$ fluorescence in response to input TetR-sfGFP fluorescence for the repression modules of Strains A (a) and B (b). The strains used for repression module characterization contain $ptet$ driving mRFP1 instead of AHL synthases. Steady-state sfGFP fluorescence and steady-state mRFP1 fluorescence ($t = 10\ h$) were divided by OD600 and then each fluorescence was normalized to the maximum steady-state values across all conditions within the same strain. Different levels of input TetR-sfGFP were induced by AHL to repress mRFP1. As the leaky expression of $tetR - sfGFP$ caused significant repression of mRFP1 even in the absence of any AHL, a range of aTc was added to the medium to sequester basal level of TetR. Re-measurement of mRFP1 fluorescence showed a broad output range when at least $3.1\ ng/mL$ aTc was added to medium. Unexpectedly, aTc also repressed TetR-sfGFP production in Strain A at high concentrations, which constrained viable aTc concentration to be less than $10\ ng/mL$. Solid circles show the average and error bars show standard deviation of the measurements (n = 2). The solid curve represents the best-fit model when the measurements were fit to repressive Hill function in the form of $\frac{1}{1+x_{max}\frac{[TetR]/K_t}{1+[aTc]/K_{aTc}}^n}$, where $K_{aTc} = 0.098\ nM$, $K_{aTc} = 0.27\ nM$ as determined in Table 4.3. The resulting fit had a range of Hill coefficient between 1.7 to 2.8 and x_{max} ranged from $1100\ nM$ to $1621\ nM$ across different aTc concentrations.

washed three times and diluted to OD600 of 0.025 in a 96 well flat clear bottom black polystyrene microplate (Corning) containing 196 μL MOPS and appropriate concentrations of AHLs (Sigma) dissolved in 4 μL of dimethyl sulfoxide (DMSO) for Fig. 4.1, or 192 μL MOPS and appropriate concentrations of AHLs in 4 μL DMSO and anhydrotetracycline (aTc) (Sigma) in 4 μL ethanol for Fig. 4.2. Synergy 2 (Biotek Instruments) was used to measure cell density (OD600) and fluorescence of growing culture every 8 min for 12 hours at room temperature. The BioTek excitation and emission wavelengths were $485\ nm$, $528 \pm 20\ nm$ for sfGFP and $560\ nm$, $620 \pm 20\ nm$ for

mRFP1.

4.3.3 Flow cytometry

Overnight cultures of Strains A and B in MOPS were washed three times and diluted to OD600 of 0.03 in fresh MOPS added with 5 ng/mL aTc and 0 or 1 μM AHL for pre-induction. After 6 hours of shaking at 750 rpm in 30 ^{o}C, cells were washed three times and diluted to OD600 of 0.025 in a 96 well deep well plate (Green BioResearch) containing 196 μL MOPS, 5 ng/mL aTc, and appropriate concentrations of AHLs (Sigma). After 8 hours of shaking at 750 rpm in 30 ^{o}C, cells were analyzed using BD LSRFortessa (BD Biosciences). Blue (488 nm) and green (561 nm) lasers were used in combination with $530/30$ nm and $610/20$ nm filters.

4.3.4 Construction of DLI device

The compartments and channels in the patterns were cut into 1/8 inch acrylic sheet (McMaster Carr) using a laser cutter (Universal Laser Systems) and then filled with SYLGARD 182 Silicone Elastomer (Dow Corning). PDMS molds were attached to the bottom of 6 well clear flat bottom cell culture plate (Falcon), and 3.4 mL of MOPS solid medium was poured into each mold to create DLI devices as illustrated in Fig. 4.3. aTc in ethanol was added to the final concentration of 5 ng/mL.

4.3.5 DLI assays

Overnight cultures of Strains A and B in MOPS were washed three times and diluted to OD600 of 0.03 in fresh MOPS added with 5 ng/mL aTc and 0 or 1 μM AHL for pre-induction. After 8 hours of shaking at 750 rpm in 30^{o}C, cells were washed three times and rediluted to OD600 of 2.0 in fresh MOPS. 0.5 μL of the culture was seeded onto each compartment of the DLI devices. Gel Doc XR+ System (Bio-rad) was used to image bacterial colonies every 30 min for 12 hours in room temperature. The blue epi illumination at 488 nm and $530/28$ nm filter was used for sfGFP and the green epi illumination at 532 nm and $605/50$ nm filters were used for mRFP1. Camera exposure time of 100 ms was used for all images.

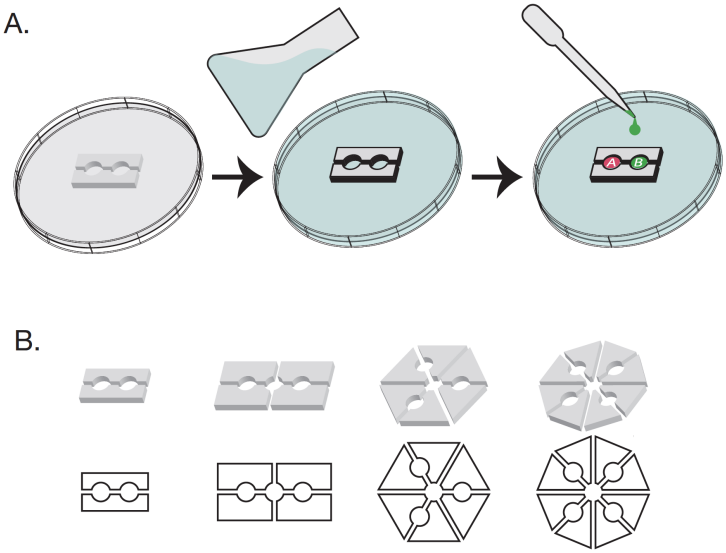

Figure 4.3: Sketch of DLI device preparation. (a) PDMS molds are attached to the bottom of a culture plate and $3.4\ mL$ of 1.5 % agarose-mixed medium is poured into the devices. After solidifying, $0.5\ \mu L$ of 2.0 OD600 cells are pipetted at the center of each compartment and allowed to grow for 10-12 hours in room temperature to study pattern formation. **(b)** Various spatial configurations are prepared as PDMS molds. A central compartment with different number of neighbors can be manufactured.

4.3.6 Computational modeling and simulation

We used custom code for computational modeling and data analysis in MATLAB (Mathworks). Details about the model construction are provided in *Supplementary Theory*. Model species and parameters are described in *Table* 4.2 and 4.3.

TABLE 4.1: Biochemical species (state space variables) whose behavior is represented by system ODEs.

Variable	**Description**
X_A	3OC12HSL in A (produced by A)
X_B	3OC12HSL in B (diffused from A)
Y_A	3OC6HSL in A (diffused from B)
Y_B	3OC6HSL in B (produced by B)
m_{R_A}	$tetR$ mRNA in A
m_{R_B}	$tetR$ mRNA in B
R_A	TetR in A
R_B	TetR in B
m_{I_A}	$lasI$ mRNA in A
m_{I_B}	$luxI$ mRNA in B
I_A	LasI in A
I_B	LuxI in B

TABLE 4.2: Parameter definitions and values used in the toy system with even parameters, i.e., corresponding parameters between types A and B are equal.

Parameter	**Description**	**Value**
$D_{X/Y}$	diffusion constant of AHLs	5×10^{-4}
l	length of channels	5×10^{-5}
d	edge weight constant	$:= \frac{D}{l^2}$
μ	dilution due to growth	0.1
$\gamma_{X/Y}$	decay rate of AHL + drainage	0.01
$\nu_{X/Y}$	production rate of AHLs	1
$V_{R_{A/B}}$	transcription rate of $tetR$	1
$N_{R_{A/B}}$	copy number of $tetR$	1
C	concentration constant	1
K_d	AHL dissociation constant	10
$n_{X/Y}$	Hill coefficient for AHL	2
$l_{R_{A/B}}$	leakiness of $plux/plas$	0
$\gamma_{m_{R_{A/B}}}$	decay rate of $tetR$ mRNA	0.105
$\epsilon_{R_{A/B}}$	translation rate of TetR	1
$\gamma_{R_{A/B}}$	decay rate of TetR	μ
$K_{R_{A/B}}$	dissociation constant of TetR to $ptet$	10
$n_{R_{A/B}}$	Hill coefficient for TetR to $ptet$	4
$l_{I_{A/B}}$	leakiness of $ptet$	0
$\gamma_{m_{I_{A/B}}}$	decay rate of $lasI/luxI$ mRNA	0.01
$\epsilon_{I_{A/B}}$	translation rate of LasI/LuxI	1
$\gamma_{I_{A/B}}$	decay rate of LasI/LuxI	μ

TABLE 4.3: Parameter definitions and values used in the ODEs for an experimental implementation using the orthogonal QS systems LuxI/LuxR and LasI/LasR.

Parameter	Description	Value	Reference
D	diffusion constant of AHLs	4.9×10^{-4} mm^2 s^{-1}	(Cunningham et al., 2011)
l	length of channels	4.5 mm	Section 4.8.7
d	edge weight constant	2.42×10^{-5}	$\frac{D}{l^2}$
λ	doubling time for *E. coli*	1.65×10^3 s	(M. Scott et al., 2010)
μ	dilution rate due to growth	4.20×10^{-4} s^{-1}	$\frac{\log 2}{\lambda}$
γ_X	decay rate of 3OC12HSL	3.07×10^{-5} s^{-1}	(Kaufmann et al., 2005)
γ_Y	decay rate of 3OC6HSL	4.77×10^{-6} s^{-1}	(Kaufmann et al., 2005)
ν_X	production rate of 3OC12HSL	1×10^{-5} s^{-1}	Fig.4.1
ν_Y	production rate of 3OC6HSL	8×10^{-5} s^{-1}	Fig.4.1
V_{R_A}	max transcription rate from $plux$	0.0308 s^{-1}	(Gotta et al., 1991)
V_{R_B}	max transcription rate from $plas$	0.0238 s^{-1}	Fig.4.9
N	plasmid pSC101* copy number	4 copies	(Lutz et al., 1997)
C	concentration constant	1	-
K_X	apparent K_d of $plas$	47.5 nM	Fig.4.9
K_Y	apparent K_d of $plux$	8.33 nM	Fig.4.9
n_X	apparent Hill coefficient of $plas$	1.83	Fig.4.9
n_Y	apparent Hill coefficient of $plux$	1.91	Fig.4.9
l_{R_A}	apparent leakiness of $plux$	0.0208	Fig.4.9

l_{R_B}	apparent leakiness of $plas$	0.0261	Fig.4.9
ϵ_R	translation rate of TetR	0.0320 s^{-1}	(Dalbow et al., 1975)
γ_R	decay rate of TetR	3.07×10^{-4} s^{-1}	(Andersen et al., 1998)
V_{I_A}	max transcription rate of $lasI$	0.0690 s^{-1}	(Gotta et al., 1991)
V_{I_B}	max transcription rate of $luxI$	0.0722 s^{-1}	(Gotta et al., 1991)
K_R	K_d of TetR for $ptet$	30 nM	$K_{[TetR-O1]} \cdot \frac{[aTc]}{K_{[aTc-TetR]}}$
n_R	Hill coefficient of $ptet$	2	Fig.4.3
$l_{I_{A/B}}$	leakiness of $ptet$	$1.98 \times 10^{10-4}$	(Lutz et al., 1997)
$\gamma_{m_{R_{A/B}, I_{A/B}}}$	decay rate of mRNA	0.002 s^{-1}	(H. Chen et al., 2015)
ϵ_{I_A}	translation rate of LasI	0.0718 s^{-1}	(Dalbow et al., 1975)
ϵ_{I_B}	translation rate of LuxI in B	0.0751 s^{-1}	(Dalbow et al., 1975)
$\gamma_{I_{A/B}}$	decay rate of LasI or LuxI	4.20×10^{-4} s^{-1}	μ
$K_{[aTc-TetR]}$	K_d of TetR and aTc	0.098 nM	(Scholz et al., 2000)
$K_{[TetR-O1]}$	K_d of TetR and tet operator	0.27 nM	Estimated from (Bintu et al., 2005)

4.4 Results

4.4.1 Spatial distribution introduces a new control knob to the conventional gene regulatory network

To explore the effects of spatial configuration on a gene regulatory network, we propose a structured bacterial communication device consisting of compartments and channels (Fig. 4.6**a**). Bacterial colonies harboring diffusible signal sensing/producing circuits grow in segregated compartments, and channels establish specific cell-cell communication between the connected colonies. In this section, we develop a mathematical model of this system that permits analysis of the dynamic gene regulation distributed across colonies. A classical intracellular gene regulatory network with two interacting biochemical molecules x and y can be modeled by:

$$\begin{cases} \frac{dX}{dt} &= -\gamma_x X + f_1(X,Y), \\ \frac{dY}{dt} &= -\gamma_y Y + f_2(X,Y), \end{cases} \tag{4.1}$$

where γ_x, γ_y are the linear decay rates of x and y, and functions f_1, f_2 characterize inhibition or activation of a species' production.

Intercellular communication between compartments for our case introduces spatially descriptive control parameters such as diffusivity, channel length, and the configuration of compartments ("geometry"). Cells secrete diffusible molecules that transmit along constrained paths to reach other cells, creating biological reaction networks that span multiple cells. For this application, we model physiologically responsive non-growing populations of cells whose physical locations are constrained within compartments and only signaling molecules diffuse through the channels between the compartments. Each compartment is assumed to consist of a single strain that produces only a single species of diffusible molecule, either x or y. While this work focuses on single-channel communication, the model can also be extended to multi-channel communication, where multiple species of diffusible molecules are present.

In a two-compartment system, we describe the one-dimensional transit of signaling molecules by introducing three new parameters: the channel length l and the diffusivities D_x, D_y for molecules x and y, respectively. We approximate the full continuous-space diffusion model to be a compartmental model, with the concentration of biochemical species assumed to be constant at all points within the same compartment (*Supplementary Theory*

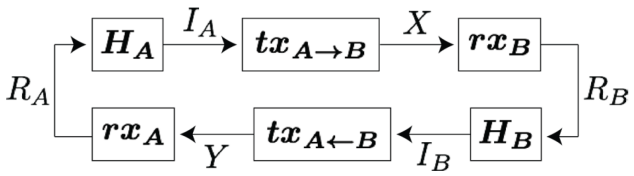

Figure 4.4: Modularized representation of the two-compartment system. Individual modules are represented by boxes, and their steady-state responses can be experimentally characterized. The subscript A or B indicates whether the box represents an event in cell type A or B. The "transmission modules" $\boldsymbol{tx_{A\to B}}$ and $\boldsymbol{tx_{B\to A}}$ encapsulate AHL synthesis and diffusion. The "reception modules" $\boldsymbol{rx_A}$ and $\boldsymbol{rx_B}$ encapsulate TetR production activated by cognate AHL. The "repression modules" $\boldsymbol{H_A}$ and $\boldsymbol{H_B}$ encapsulate AHL synthase production repressed by TetR. Y is the vector denoting concentrations of 3OC6HSL, and X denotes 3OC12HSL concentrations. Rs denote intracellular TetR concentrations. Is denote intracellular AHL synthase concentrations.

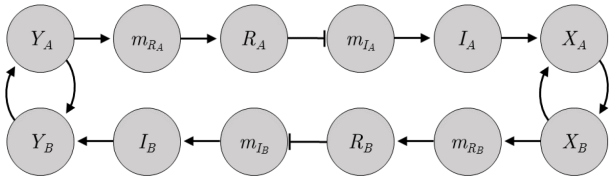

Figure 4.5: Network of the biochemical species in the two-compartment system. Sharp arrows indicate activation, or positive parity; flat arrows indicate inhibition/repression, or negative parity. A cycle is any undirected sequence of edges and nodes beginning at one element and ending at the same element (i.e., ignoring direction of arrows). The parity of a cycle is the parity of the product of the signs of all edges traversed to complete the cycle. Since every cycle in the graph is positive in parity, the closed-loop system is monotone. To ensure that the corresponding open-loop system is strongly monotone we require a directed path (i.e., following the arrows) to exist between the input node and every other node, and between every node and the output node. The input and output nodes are determined by where the feedback loop is broken (see Section 4.8.34.8.3). In this system every element is reachable from every other element, so these conditions will be satisfied regardless of where the cycle is broken. Hence the open-loop (input-output) system is also monotone.

Section 1). In doing so, we trade off accuracy in spatial and temporal dynamics for a set of analytically tractable equations that are reasonable when channels are narrow and the volume of compartments and channels is negligible. Then the equations for x and y produced in separate compartments A and B, respectively, are given by

$$\begin{cases} \frac{dX_A}{dt} = -\gamma_X X_A + \frac{D_x}{l^2}(X_B - X_A) + f_1(X_A, Y_A), \\ \frac{dX_B}{dt} = -\gamma_X X_B + \frac{D_X}{l^2}(X_A - X_B), \\ \frac{dY_A}{dt} = -\gamma_Y Y_A + \frac{D_Y}{l^2}(Y_B - Y_A), \\ \frac{dY_B}{dt} = -\gamma_Y Y_B + \frac{D_Y}{l^2}(Y_A - Y_B) + f_2(X_B, Y_B), \end{cases} \tag{4.2}$$

where X_A, Y_A and X_B, Y_B designate the concentrations of x and y in compartments A and B. Increasing diffusivity or decreasing channel length increases the practical strength and rate of communication between neighboring compartments.

Channeled diffusion permits the design of arbitrary networks whose dimensions increase with each added compartment by the number of biochemical species present within that compartment. To model high-dimensional networks, we use matrices to represent connections between compartments. Given a system with N_A compartments producing x and N_B compartments producing y, we define the vector $X_A \in \mathbb{R}^{N_A}$ for the concentrations of x in compartments A, and the vector $X_B \in \mathbb{R}^{N_B}$ for the concentrations of x in compartments B. The vectors Y_A and Y_B are defined similarly for molecule y. We then introduce the Laplacian matrix L to represent connections between compartments (Fig. 4.6**a**). The entries of L are

$$[L]_{ij} = \begin{cases} \frac{1}{l_{ij}^2}, & \text{if there is a channel,} \\ 0, & \text{if there is no channel,} \\ -\sum_{k \neq i}[L]_{ik}, & \text{if } i = j, \end{cases}$$

where $l_{ij} = l_{ji}$ is the length of the channel between compartments i and j.

The full biochemical dynamics are therefore described by

$$\begin{cases} \begin{bmatrix} \frac{dX_A}{dt} \\ \frac{dX_B}{dt} \end{bmatrix} & = -\gamma_X \begin{bmatrix} X_A \\ X_B \end{bmatrix} + D_X L \begin{bmatrix} X_A \\ X_B \end{bmatrix} + \begin{bmatrix} f_1(Y_A) \\ 0 \end{bmatrix}, \\ \begin{bmatrix} \frac{dY_A}{dt} \\ \frac{dY_B}{dt} \end{bmatrix} & = -\gamma_Y \begin{bmatrix} Y_A \\ Y_B \end{bmatrix} + D_Y L \begin{bmatrix} Y_A \\ Y_B \end{bmatrix} + \begin{bmatrix} 0 \\ f_2(X_B) \end{bmatrix}. \end{cases} \tag{4.3}$$

Formulation of the L matrix and the simulated gene expression behavior of an example network are shown in Fig. 4.6**a&b**.

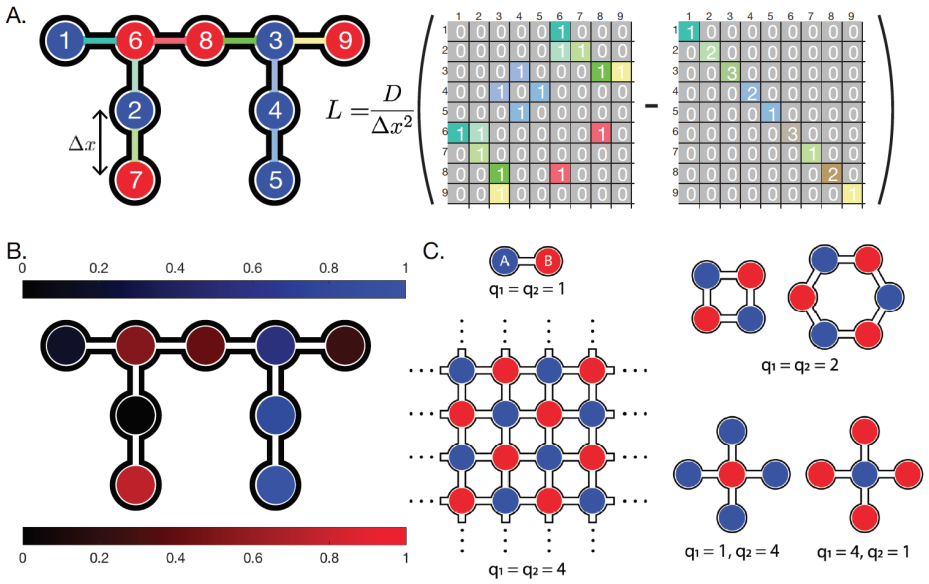

Figure 4.6: A spatially distributed gene regulatory network established by intercellular communication between bacterial subpopulations separated by compartments. (a) Example of the Laplacian matrix constructed from an arbitrary network. Δx refers to the length of the channel connecting adjacent compartments. **(b)** Simulated steady-state pattern from the example network when cross-repressive interactions are used between the two cell types. Color represents concentrations of repressor in cell types A and B, normalized to the maximum concentration across all colonies. The parameters are as given in *Table* 4.2 where corresponding biochemical parameter values are equal between strains. **(c)** Examples of reducible spatial configurations (see text).

4.4.2 Dimensionality reduction enables steady-state analysis of multi-compartmental cross-repressive networks

To apply our mathematical model to a canonical example, we focus subsequent analyses on cross-repression in multi-compartmental networks. Lateral inhibition, or the mutual inhibition between adjacent units, is a common mechanism to generate contrasting patterns. Here, $f_1(Y)$ and $f_2(X)$ are Hill functions describing repression on the production of the diffusible molecules by the opponent molecules.

Using dimensionality reduction, we can identify the existence of stable contrasting steady-state patterns in systems with a particular class of spatial configurations for which all channels are the same length, compartments of one cell type are connected only to compartments of the other cell type ("neighbors"), and the number of neighbors is the same for all compartments of the same type. Arbitrarily large multi-compartmental systems with this structure can be reduced in dimension to two-compartmental systems with multiplicative factor adjustments to the diffusion (Fig. 4.6**c**, *Supplementary Theory* Section 3). In particular, if each compartment of type A has q_1 neighbors and each compartment of type B has q_2 neighbors, then the concentrations of repressors in each of the representative compartments A, B evolve as Eq. 4.2 with $D_{X/Y}$ replaced by $q_1 D_{X/Y}$ in A (for X_A and Y_A) and by $q_2 D_{X/Y}$ in B (for X_B and Y_B). These adjustments corroborate the intuition that a compartment with more neighbors experiences higher diffusive in- and out-flux of the signaling molecules. We use "diffusion-mediated lateral inhibition" (DLI) to refer to multi-compartmental cross-repressive networks that satisfy the geometric constraint permitting reducibility. The DLI system is analogous to contact-mediated lateral inhibition mechanisms such as Notch-Delta, with the crucial distinction that DLI acts through diffusion rather than direct contact. Contact-mediated systems can be modeled by replacing the Laplacian matrix with an adjacency matrix weighted by contact area (Arcak, 2013) as described in *Supplementary Theory* Section 3D.

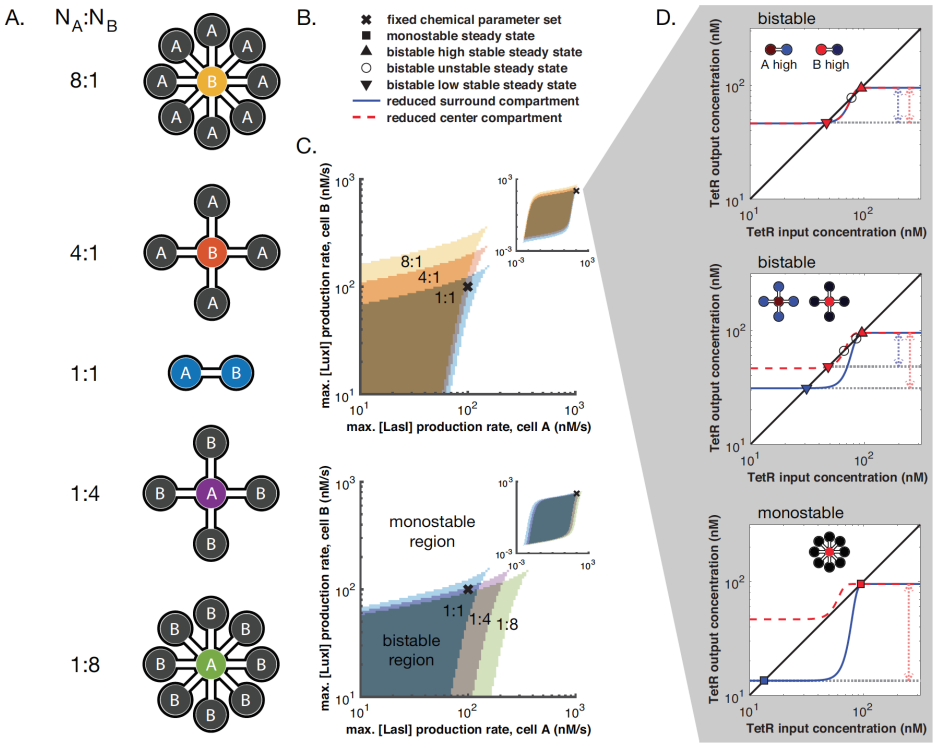

Figure 4.7: Patterning mechanism and contrast level determined by biochemical parameters and geometries in simulation. (a) Star geometries with varying numbers of surround compartments. Throughout the figure, ratios of colony numbers are given as $N_A : N_B$. **(b)** Character legend for plots in **c** and **d**. **(c)** Overlaid stability plots show biochemical parameter ranges for which the system is monostable (white) and bistable (shaded, colors corresponding to center compartment of appropriate geometry in **a**). Parameters on the axes are maximum steady-state production rates for LasI (x-axis) and LuxI (y-axis). Remaining parameters are as given in *Table* S2 where corresponding biochemical parameter values are equal between strains. As the number of points in the star changes, the shape of the bistable region remains the same (relative to log-scale axes) but shifts relative to the exact biochemical parameter values (insets show the full shape of the bistable region). $\times$ indicates an arbitrary set of fixed biochemical parameters that is bistable in the 1:1 and 4:1 cases but monostable for the 8:1 case. **(d)** For the biochemical parameters indicated by $\times$ in **c**, a graphical test reveals that contrast may arise from a bistable system (1:1 and 4:1) or from a monostable system with imbalance (8:1) between the input/output characteristics of strains in the reduced systems. Steady states are indicated by ▲ (high expression) and ▼ (low expression) for the bistable case or ■ for the monostable case. In the bistable case with imbalance (4:1), the contrast level ($\leftrightarrow$) is greater when expression in the center compartment (dashed red) is high than when expression in the surrounding compartments (solid blue) is high. Small insets show corresponding configurations and possible steady-state solutions.

The DLI system can be experimentally implemented using two strains of bacteria that communicate via orthogonal QS systems (S. R. Scott et al., 2016; Grant et al., 2016) and internal inverters (repression circuits). One strain of bacteria is seeded in each compartment with connected compartments alternating between strain types. In this implementation, x is the diffusible AHL produced by Strain A and y is the orthogonal AHL produced by Strain B.

To facilitate the choice of genetic circuit components, we expand Eqs. 4.1, 4.2, 4.3 to model the dynamics of mRNA transcription/degradation, protein translation/degradation, and AHL synthesis/degradation (*Supplementary Theory* Section 1). The equilibrium solutions of the augmented model are equivalent to the equilibrium solutions of the original model in Eq. 4.3 when f_1 is given by

$$f_1(u) = \frac{\epsilon_{I_A} N_{I_A}}{\gamma_{I_A}} a_1 h_1 \left(\frac{\epsilon_{R_A} N_{R_A}}{\gamma_{R_A}} a_2 h_2(u) \right). \tag{4.4}$$

Here, h_1 is a Hill function for repressor inhibiting the transcription of AHL synthase and h_2 is a Hill function for AHL activating the transcription of the repressor in the other strain. The Hill coefficients, dissociation constants (K_d), and leakiness of h_1 and h_2 are set by the choice of QS and repressor molecules. Independently, we can change the promoter copy numbers $N_{I_{A/B}}$, $N_{R_{A/B}}$ as well as the translation rates $\epsilon_{I_{A/B}}$, $\epsilon_{R_{A/B}}$ and degradation rates $\gamma_{I_{A/B}}$, $\gamma_{R_{A/B}}$ of AHL synthases and repressors. The composite parameters α_1 and α_2 incorporate the remaining parameters that we cannot easily vary, including maximal transcription rates, mRNA decay rates, and the synthesis rates of AHL by the synthases. The function f_2 is structured similarly except that the order of h_1 and h_2 is reversed (*Supplementary Theory* Section 1).

4.4.3 Graphical analysis of a DLI circuit predicts two mechanisms of contrasting pattern formation

Steady states for the two-compartment diffusion system in Eq. 4.2—and hence for the reduced system—can be found graphically by plotting the steady-state output X_A (or Y_B) for constant input X_A (or Y_B) and locating intersections with a line of slope one, since at steady state the output equals the input for the closed-loop system. The graphical test also reveals the stability of the equilibria: one intersection implies the system is monostable

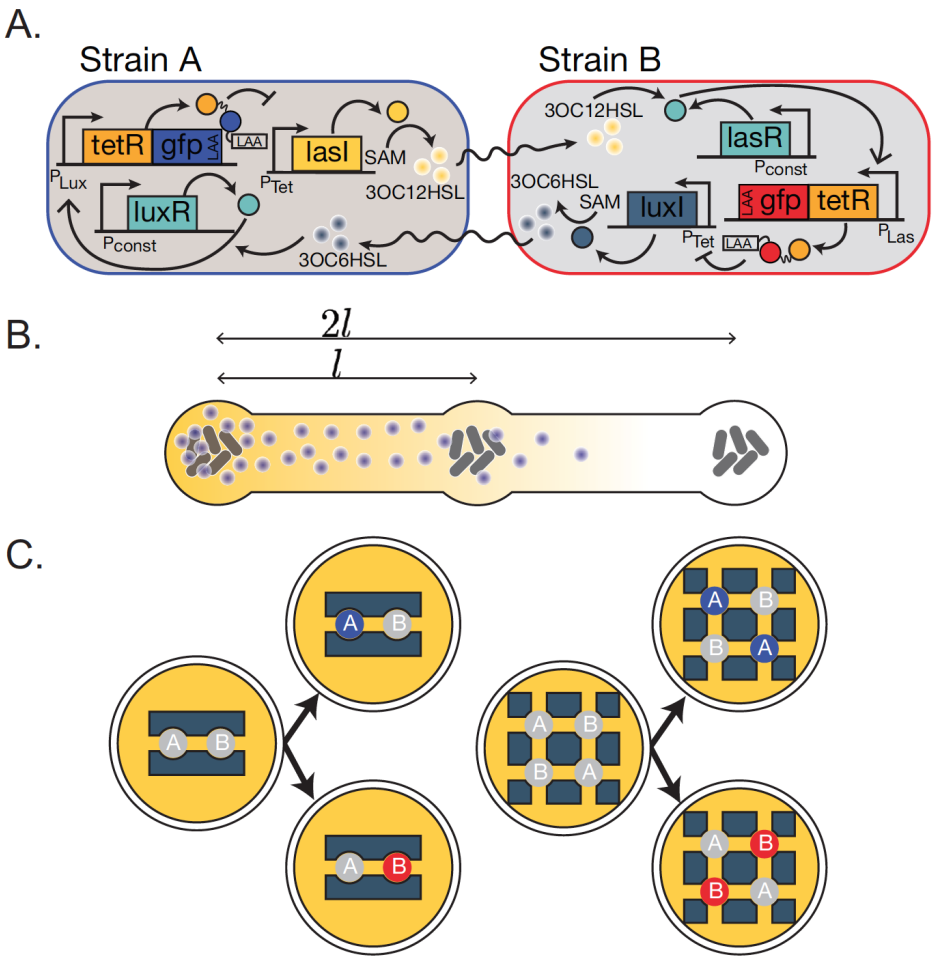

Figure 4.8: Schematic designs of the DLI system. Arrow-headed lines indicate activation and bar-headed lines indicate inhibition. **(a)** Genetic circuit diagram of cross-repressive strains. **(b)** Channel length l is chosen such that AHL diffusion establishes communication between adjacent compartments, but not between non-adjacent compartments with distance $\geq 2l$. **(c)** Each compartment of the DLI device is inoculated with one strain type. PDMS mold (indigo) is placed on a tissue culture plate to shape solid medium (yellow) into compartments and channels. Contrasting patterns emerge when two strains have different sfGFP-tagged TetR levels, either high (represented by green or magenta colored colonies) or low (represented by gray colonies).

while three intersections imply the system is bistable with an unstable homogeneous equilibrium (see *Supplementary Theory* Section 2).

Contrasting patterns result from disparity in the steady-state target gene

expression between cross-repressive strains in a DLI system. Using graphical analysis to assess system equilibria, we construct two-dimensional bifurcation diagrams to reveal two different mechanisms for contrasting pattern generation (Fig. 4.7). One mechanism originates from bistability, in which the system parameters allow two alternative stable states of x and y production based on initial conditions and possible intrinsic or extrinsic noise/perturbations. The second mechanism occurs in a monostable system when one strain always expresses higher levels than the other.

Ultrasensitivity (cooperativity) in f_1 and f_2 is necessary for bistable contrast. Proper kinetic rate matching ensures that ultrasensitivity is preserved in the feedback loop (Fig. S3) (Hooshangi et al., 2005). In addition to ultrasensitivity, DLI systems must have sufficiently similar inhibition strength between strains to be bistable. When the system loses bistability due to unbalanced inhibition strengths, monostable contrast emerges, with the extent of the contrast depending on the degree of imbalance (*Supplementary Theory* Section 5B). Changing spatial configuration triggers a bifurcation by modifying the effective inhibition strength between strains (Fig. S4).

4.4.4 DLI network design and implementation require diffusible cross-repression and a geometric culturing platform

Two *Escherichia coli* strains, A and B, were constructed using a pair of orthogonal QS systems and a highly cooperative repressor, $tetR$ (Fig. 4.8**a**). In both strains, $tetR$ is translationally fused to the green fluorescent protein reporter, $sfGFP$, with LAA $ssrA$ degradation tag in the C-terminus to allow dynamic tracking of the cell state (Flynn et al., 2001).

The length of the channel (l) between compartments determines the AHL concentration in the neighboring compartments as well as the intercompartmental communication lag time (Fig. 4.8**b**). A partial differential equation (PDE) model of AHL production, degradation, and diffusion is used to optimize l for sufficient diffusion of AHL to the immediate neighbors while preventing communication between nonadjacent compartments (Fig. 4.9**a**, *Supplementary Theory* Section 6). Since AHLs can be stable with a half-life of 6 hours up to days (Kaufmann et al., 2005), an efflux channel is added to each compartment to match the dilution rate of AHL to the degradation rates of other proteins in the DLI circuit (Fig. 4.9**b**, *Table* 4.3). We use polydimethylsiloxane (PDMS) as the mold to shape solid medium into compartments and channels in specific geometries (Fig. 4.8**c**, Fig. 4.3).

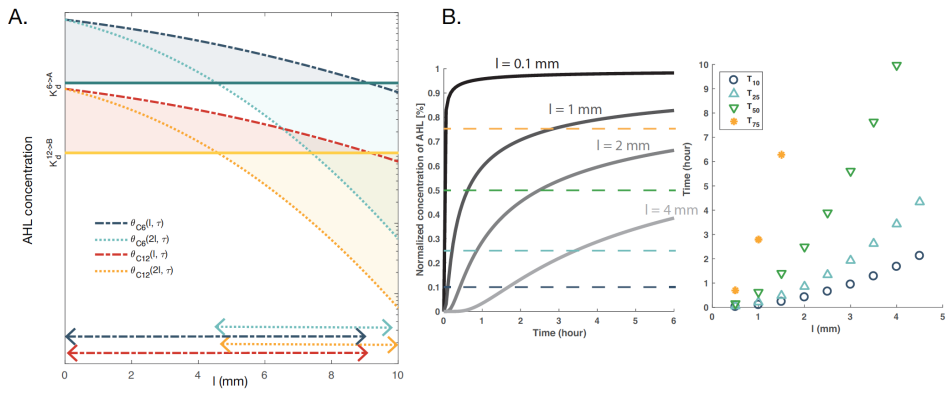

Figure 4.9: PDE simulation for DLI device design to optimize the inter-compartmental channel length (l) and the channel length connecting a compartment and reservoir (l_{res}). **(a)** Graphical representation of constraints on l. The inter-compartmental channel length should allow sufficient AHL diffusion in directly adjacent compartments while keeping the AHL concentration in the second closest compartments (distance $2l$) to be below the threshold (K_d of AHL-inducible promoters). A constraint can be represented as an area under or above a function, and the l values where all areas overlap indicates appropriate channel length range. The smallest l value satisfying the criteria is optimized for shortest communication time, while largest l value is optimized for longest non-neighbor communication time. $4.5\text{mm} < l < 9$ mm satisfies the criteria with the experimentally evaluated parameters given in Table 4.3. **(b)** Characterization of l_{res} length and decay time. The channel length connecting a compartment and a reservoir adds extra decaying mechanism for AHL by allowing AHL efflux from every compartment to the bulk solid medium of $AHL \simeq 0$. The efflux diffusion rate from the compartment to medium depends on the length of the channel. The left plot shows the portion of AHL diffused out with respect to time, normalized to the AHL concentration inside the compartment at $t = 0$. The right plot shows the time constant it takes for a portion of the AHL concentration inside the compartment to diffuse outside with respect to the channel length.

4.4.5 Bistable and contrasting gene expression was observed for DLI circuit in liquid coculture

Before testing pattern formation in compartmental structures, we verified the predicted bistability of the constructed circuits in liquid coculture. Numerical parameters required for the biochemical model Eq. 4.4 were determined by experimental measurements of individual modules of AHL reception/activation, transcriptional repression, and AHL synthesis and diffusion in the DLI circuit (Figs. 4.1, 4.2, 4.10).

Single-cell reporter gene expression was measured using flow cytometry. To examine the existence of two stable steady states, the cocultures were

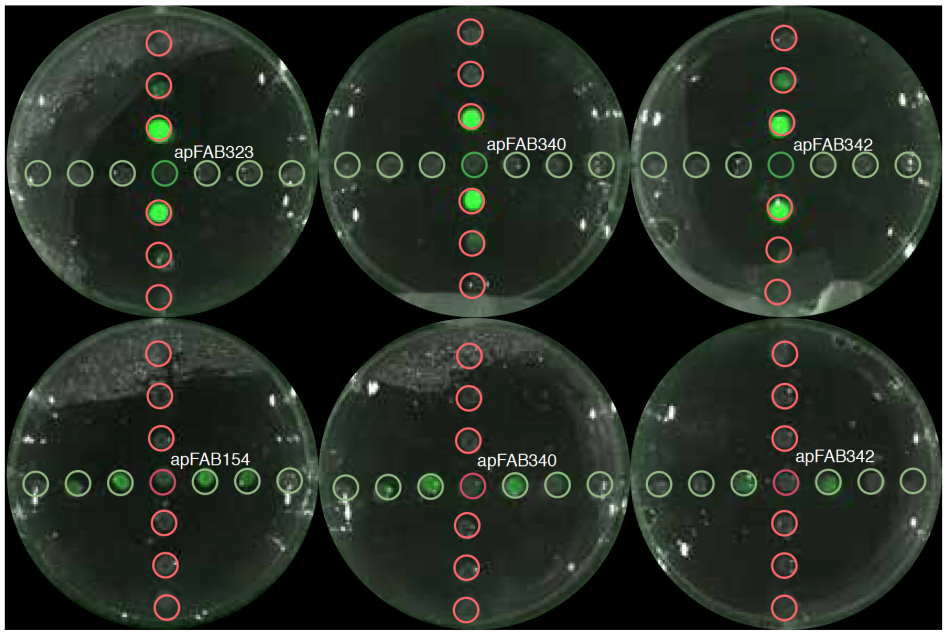

Figure 4.10: Plate fluorimeter assay of AHL production and reception by Strains A and B. On a well of $3.4\ mL$ of MOPS EZ Rich solid medium with $5\ ng/mL$ aTc, the complete strains with different $ptet$ promoter strength (Mutalik et al., 2013) were seeded at the center, surrounded by receiver strains (Strain A receiver in horizontal direction and Strain B receiver in vertical direction). The top row is seeded with Strain A variants at the center and the bottom row is seeded with Strain B variants at the center. The promoter strength of $ptet$ seems to have little effect on AHL production, as receiver strains showed similar levels of activation regardless of promoter variants within the same strain type.

biased with varying external concentrations of 3OC6HSL or 3OC12HSL. While the external AHL inductions in monocultures of Strains A (Fig. 4.11**a** diamonds and the solid line for the Hill equation fit) and B (Fig. 4.11**b** diamonds and the solid line for the Hill equation fit) resulted in gently sloped sigmoid responses, the coculture showed a sharp transition in steady state at 10 nM 3OC6HSL (Fig. 4.11 squares), which is a characteristic for a bistable feedback loop (Gardner et al., 2000).

Another characteristic of bistable systems is hysteresis. To investigate whether our system can reach two heterogeneous steady states in the same culture condition depending on initial conditions, we pre-conditioned

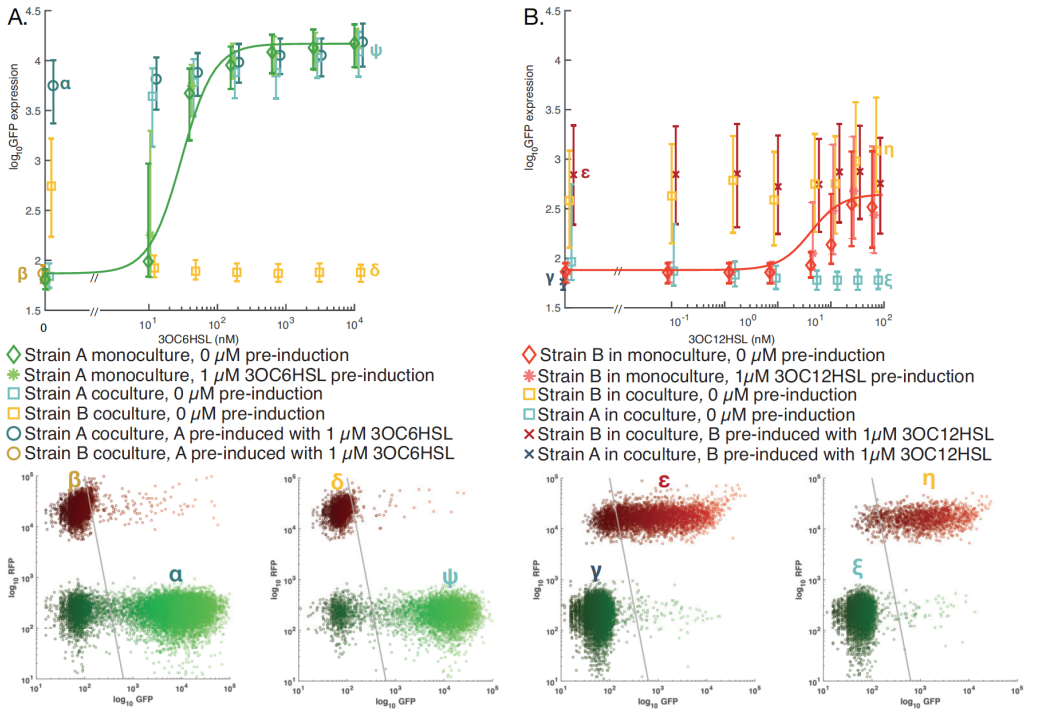

Figure 4.11: Steady states of the cross-repressive circuit characterized using flow cytometry measurements in liquid cultures. Strain B was identified using constitutively expressed *mRFP1*. Varying concentrations of **(a)** 3OC6HSL or **(b)** 3OC12HSL were externally added to the liquid medium (x-axis) and the medians of sfGFP fluorescence after 8 hours of growth were recorded (y-axis). Error bars represent 1st and 3rd quartiles of sfGFP fluorescence. All of the multi-strain cocultures, each indicated by $\square, \circ, \times$, exhibited contrasting expression profiles between Strains A and B. While the monocultures of strains A and B showed gently sloped responses to external AHL with Hill function fits of $K_d \simeq 50\ nM$ and $K_d \simeq 20\ nM$ (solid lines), the two-strain coculture showed a switch-like response at threshold [3OC6HSL] $= 10\ nM$. Hysteresis was tested by pre-inducing one of the strains with appropriate AHL prior to washing and mixing the strains into a coculture with fresh medium. The two-strain cocultures maintained pre-induced states after 8 hours of growth even without external AHL whereas the monocultures lost their pre-induction states. Bottom scatter plots show similarity in gene expression patterns between Strain-A-biased coculture (ψ, δ) and Strain-A-pre-induced coculture at 8 hours of growth after removing external AHL (α, β), and Strain-B-biased coculture (ζ, η) and Strain-B-pre-induced coculture at 8 hours of growth after removing external AHL (γ, ϵ).

Strains A and B monocultures with the saturated concentration (1 μM) of either 3OC6HSL or 3OC12HSL prior to mixing them into a coculture. The cocultures maintained the distinct gene expression states determined by preconditions over time, while the similarly pre-induced monocultures lost the

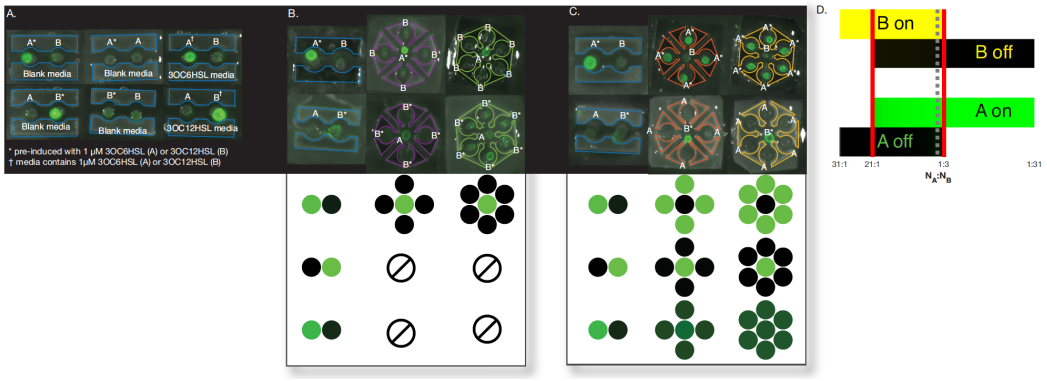

Figure 4.12: Contrasting pattern formation in various DLI devices. The fluorimeter images were taken after 12 hours of growth in room temperature. $*$indicates pre-induced strains with 1 μM AHL and † indicates strains that were biased to be fluorescent by externally added AHL in medium. **(a)** 1:1 spatial configuration seeded with cells that had different initial conditions and strain combinations. Devices were seeded with a pair of complementary strains **(left)**, negative controls consisting a single strain **(middle)**, and positive controls of complementary strains where either 1 μM 3OC6HSL or 1 μM 3OC12HSL was mixed in solid medium **(right)**. **(b,c)** 1:1, 1:4, and 1:6 spatial configurations seeded with Strain A at the center surrounded by Strain B **(b)** or Strain B surrounded by Strain A **(c)**. Top panel shows the fluorimeter images and the bottom panel shows predicted steady-state pattern from computational simulations with the parameter values given in *Table* S3. When multiple equilibria exist, the predicted patterns are plotted in the order of "A high", "B high", and "unstable". **(d)** Simulated one-dimensional bifurcation diagram in which the ratio of compartments of Strain A:B is used as the bifurcation parameter. The remaining parameters are given in *Table* S3. Brighter color indicates higher steady-state [sfGFP].

pre-conditioned state and exhibited the sigmoid induction curves when the cultures were transferred in the fresh media (Fig. 4.11 asterisks).

The appearance of both the sharp transition and hysteresis confirms that cross-repression between Strains A and B produces an effective intercellular bistable switch.

4.4.6 Bifurcation is observed for cells grown on the geometric culturing platform

Theoretically, the DLI system in a linear geometry (one-to-one compartments) with proper channel length should behave similarly to the liquid co-culture. To investigate bistability and hysteresis in the geometric platform, we plated Strains A and B pre-conditioned with either 1 μM 3OC6HSL or 1 μM 3OC12HSL on solid medium after washing of the pre-conditioning

media and observed their gene expression over time using the plate fluorimeter. The linear geometry plated with two strains maintained the preconditioned states over 12 hours whereas the single-strain systems on the same setup quickly lost the pre-induced gene expression (Fig. 4.12**a**). When 3OC6HSL pre-induced Strain A was seeded adjacent to Strain B, Strain A showed high reporter expression while Strain B showed basal expression comparable to the single-strain control, and this persisted in time. Similarly, 3OC12HSL pre-induced Strain B seeded adjacent to Strain A showed high reporter expression in Strain B and low reporter expression in Strain A, indicating that contrasting patterns depended on initial conditions.

Next, we evaluated DLI systems in star geometries (Fig. 4.12**b&c**). We experimentally observed that the system that was bistable in the linear geometry became monostable as the number of outer compartments of Strain B surrounding Strain A increased to four or more (Fig. 4.12**b**). The monostable contrasting pattern exhibited high reporter fluorescence in the center Strain A regardless of the initial cell states. Strain B, on the other hand, did not shift from monostable to bistable in our experimentally tested geometries, but rather augmented the bistable expression as the number of surrounding A increased. Although the experimental setup cannot physically accommodate > 6 compartments, the mathematical simulation suggests that further increasing the number of surrounding compartments would shift the system from bistable to monostable contrast with Strain B expressing high reporter (Fig. 4.12**d**).

Our results show that we can easily control the geometry of the DLI system to affect the circuit behavior and trigger a bifurcation. Geometry may also offset imbalance in biochemical parameters (such as difference in f_1 and f_2 in Eq. 4.3, Fig. S11) and improves the stability of bistable steady states (Fig. 4.12**c**).

4.5 Discussion

Recent advancement in high-throughput sequencing has revealed that an astonishing range of microbial biodiversity may exist in a single ecosystem (Thompson et al., 2017). While this paper focuses on spatial intercellular interactions between two strains of bacteria with two signaling molecules, the theory may be generalized to handle an arbitrary number of diffusible biochemical species. A system with $\{m_1, m_2, \ldots, m_n\}$

diffusible molecules produced respectively in $\{N_{m_1}, N_{m_2}, \ldots, N_{m_n}\}$ compartments, where the numbers of communication channels connected to each of the compartments are $\{d_{m_1}, d_{m_2}, \ldots, d_{m_n}\}$, can be modeled with $\sum_{i=1}^{n} N_{m_i} + d_{m_i}$ equations when assuming communications among only immediate neighbors. The number of equations increases to $m \times N$ when diffusion among non-immediate neighbors is accounted for and becomes even more if non-diffusible species internal to the compartments must be included in the model. Hence, arbitrary compartmental systems are generally high in dimension and difficult to analyze in full, although symmetries in spatial configuration—such as the alternating-neighbor pattern—may enable mathematical reductions to simplify the search for particular solutions (Rufino Ferreira et al., 2013).

For the analyzed DLI system, we explored the role of geometry in system behavior. Although the implementation was done in multi-strain bacterial colonies, our theory can be applied to isogenic populations as well. Here, the units of interest are individual cells rather than colonies and communication must be contact-mediated since diffusion-based signaling would form self-loops. Dimensionality reduction still applies when cells can be categorized into two separate "classes" by virtue of spatial configuration. Replacing the Laplacian matrix with the adjacency matrix (*Supplementary Theory* Section 3D) simulates cell-to-cell contact rather than diffusion. The remainder of the analysis then proceeds as before. Extensive and detailed research has been performed to accurately model the developmental processes in metazoans (Amonlirdviman et al., 2005; Samoilov et al., 2006; Arcak, 2013), and a handful of recent studies have highlighted that spatially relevant parameters such as the number of neighbors or the contact area between them can influence patterning activity even in genetically isogenic cell populations (S. S. Lee, 2016; Guisoni et al., 2017; Matsuda et al., 2015). Our work offers a unified interpretation of these results with respect to the imbalance in transfer functions between pairs of representative cells. With sufficient imbalance (*Supplementary Theory* Section 5B), the system becomes monostable, essentially guaranteeing the fate of the involved cells, and in fact only spatial control knobs can introduce monostable contrast in isogenic populations, since changes to biochemical parameters affect all cells equally. Furthering our understanding of micro-scale pattern formation would require experimental implementation of controllable contact-based systems. In bacteria, several contact-dependent inhibition systems have been discovered (E. C. Garcia et al., 2016; Morse et al., 2012) in which potential harnessing strategies have been discussed (D. Chen, 2014).

Genetic circuit design and implementation are hampered by context-dependent gene expression (Cardinale et al., 2013). Spatial control has advantages over biochemical parameter modification in that it can linearly modulate the effective interaction strength (Fig. 4.12*D*) via the number of connected channels, and the modulation is robust to intracellular conditions. Furthermore, physical separation of the composite strains reduces resource competition among different strains (H. J. Kim et al., 2008) to stabilize the intercellular network. Spatial control is constrained by structural limitations, such as the maximum number of compartments that fit on the mold or whether the layout of the desired communication network can be laid out without channels intersecting each other. Thus, synthetic biologists should exploit both biochemical and spatial control knobs for precise design of microbial consortia engineering.

4.6 Conclusions

In this work, we present a computational method for simulation of physically separated microbial colonies each implementing different gene regulatory networks. We validate our theory by experimentally demonstrating control over gene expression patterns in a diffusion-mediated lateral inhibition circuit. We highlight the importance of spatial arrangement as a control knob for modulating system behavior. Our systematic approach provides a foundation for future applications that require understanding and engineering of multi-strain microbial communities for sophisticated, synergistic functions.

4.7 Acknowledgments

The authors thank NIH National Institute of General Medical Sciences grant 1R01GM109460-01 and Air Force Office of Scientific Research grant FA9550-14-1-0089 for support. The DH10B variant TR117 was provided by Dr. Thomas L. Ruegg and the backbone image processing scripts were provided by Lynn D. Kong. We thank Prof. Michel Maharbiz for comments and discussions, and Prof. Dave Savage for kindly letting us use their plate fluorimeter.

4.8 Supplementary Theory

4.8.1 Modeling dynamics of two-compartment system

Consider two cell types A and B. Each cell produces a product along with an associated signaling molecule that diffuses to the opposite cell and inhibits that cell's production of its product and signaling molecule. Depending on factors such as the strength of inhibition and the proximity of the two cells, over time one cell may be "dominated" by the other, which produces vastly more product. If either cell could theoretically become the dominant cell, then such a system is considered bistable. There are many biological examples of such bistable systems, perhaps most famously the Notch/Delta signaling pathway (Greenwald et al., 1992).

Since quantitative factors determine whether a system of cross-repression is monostable or bistable, it can be difficult to experimentally implement a bistable system without theory that numerically describes and predicts the system's behavior. We present here an adaptation of the theory developed by Ferreira et al. (Rufino Ferreira et al., 2015) for a general compartmental system of lateral inhibition to the specific case of quorum sensing between colonies of synthetically engineered bacterial cells.

4.8.2 Setup

We consider a two-compartment system of cross-repression comprising two colonies of bacterial cells of type A and B (Fig 4.8**a** in main text). The full system employs orthogonal quorum-sensing (QS) systems. Ultrasensitivity is necessary for bistability, but most QS systems alone are not ultrasensitive, therefore we introduce a repressor such as TetR that binds a promoter with an ultrasensitive response. A channel between the two compartments allows diffusion of AHL between the colonies. To prevent AHL buildup due to its slow decay rate, we allow the possibility of adding an extra "drain" channel of the same length as the other channels that connects each compartment to a reservoir with zero concentration AHL. We incorporate this drainage into the overall decay rate for AHL.

In exponentially growing populations of bacteria, the volume at which a cell divides increases with the growth rate such that the increase in volume outpaces the increase in protein, leading to a net decrease in protein concentration with growth (Klumpp et al., 2009; Shahrezaei et al., 2015). Such dilution contributes to the effective decay rate of products within the

cell. In our model we assume colonies are diluting biochemical species concentration similarly to during exponential phase of growth and incorporate dilution directly into the decay rate of mRNA and protein within cells.

Ordinary differential equations (ODEs)

The cross-repression system is divided into three modules per cell type, each describing a particular function of signal reception, repression, and signal production & diffusion (Fig.4.4). The state space variables are listed in Table 4.1. The test parameters are defined and values provided in Table 4.2.

The repression module $\boldsymbol{H_A}$ describes LasI production by A for a given input concentration of TetR and is modeled by a repressive Hill function:

$$\boldsymbol{H_A}\ :\ \begin{cases} \dot{m}_{I_A} = V_{I_A} N_{I_A} C \left(\frac{1}{1+\left(\frac{R_A}{K_{R_A}}\right)^{n_{R_A}}} + l_{I_A} \right) - \gamma_{m_{I_A}} m_{I_A} \\ \dot{I}_A = \epsilon_{I_A} m_{I_A} - \gamma_{I_A} I_A \end{cases} . \tag{4.5}$$

The steady-state concentration of synthase for a constant input R_A^* is then

$$I_A^* = \frac{\epsilon_{I_A} V_{I_A} N_{I_A} C}{\gamma_{I_A} \gamma_{m_{I_A}}} \left(\frac{1}{1+\left(\frac{R_A^*}{K_{R_A}}\right)^{n_{R_A}}} + l_{I_A} \right) . \tag{4.6}$$

The transmission module $\boldsymbol{tx_{A\to B}}$ encapsulates the catalysis of 3OC12HSL by LasI and the diffusion of 3OC12HSL to B. It accepts an input I_A, the concentration of LasI produced by cell A, and outputs a vector $X = [X_A,\ X_B]^T$, where the first entry is the concentration of 3OC12HSL in A and the second is the concentration of 3OC12HSL in B after diffusion. The equations governing the behavior of this module are

$$\boldsymbol{tx_{A\to B}}\ :\ \begin{cases} \dot{X}_A = d(X_B - X_A) - \gamma_X X_A + \nu_X I_A \\ \dot{X}_B = d(X_A - X_B) - \gamma_X X_B \end{cases} \tag{4.7}$$

where $d(X_B - X_A)$ represents the diffusion balancing of 3OC12HSL between A and B (the term $d := \frac{D}{l^2}$ includes the diffusion coefficient D and the channel length l). The steady-state concentration of X_B in response to constant

input I_A^* is then

$$X_B^* = \frac{d\nu_Y I_A^*}{\gamma_X(\gamma_X + 2d)}. \tag{4.8}$$

The reception module $\boldsymbol{rx_B}$ captures TetR production in B as a function of X_A. The activation of repressor mRNA transcription by 3OC12HSL binding to LasR is modeled by a Hill function:

$$\boldsymbol{rx_B} \; : \; \begin{cases} \dot{m}_{R_B} = V_{R_B} N_{R_B} C \left(\frac{\left(\frac{X_B}{K_X}\right)^{n_X}}{1+\left(\frac{X_B}{K_X}\right)^{n_X}} + l_{R_B} \right) - \gamma_{m_{R_B}} m_{R_B} \\ \dot{R}_B = \epsilon_{R_B} m_{R_B} - \gamma_{R_B} R_B \end{cases} . \tag{4.9}$$

The complementary modules are identical in structure to those above, and are included here only for completeness. The repression module $\boldsymbol{H_B}$ describes LuxI production by B:

$$\boldsymbol{H_B} \; : \; \begin{cases} \dot{m}_{I_B} = V_{I_B} N_{I_B} C \left(\frac{1}{1+\left(\frac{R_B}{K_{R_B}}\right)^{n_{R_B}}} + l_{I_B} \right) - \gamma_{m_{I_B}} m_{I_B} \\ \dot{I}_B = \epsilon_{I_B} m_{I_B} - \gamma_{I_B} I_B \end{cases} . \tag{4.10}$$

The transmission module $\boldsymbol{tx_{B\to A}}$ covers catalysis of 3OC6HSL by LuxI and the diffusion of 3OC6HSL to A. The output vector $Y = [Y_A,\ Y_B]^T$ stores the concentrations of 3OC6HSL in A and B respectively:

$$\boldsymbol{tx_{B\to A}} \; : \; \begin{cases} \dot{Y}_A = d(Y_B - Y_A) - \gamma_Y Y_A \\ \dot{Y}_B = d(Y_A - Y_B) - \gamma_Y Y_B + \nu_Y I_B \end{cases} . \tag{4.11}$$

The reception module $\boldsymbol{rx_A}$ describes repressor production in A as a function of Y_A:

$$\boldsymbol{rx_A} \; : \; \begin{cases} \dot{m}_{R_A} = V_{R_A} N_{R_A} C \left(\frac{\left(\frac{Y_A}{K_Y}\right)^{n_Y}}{1+\left(\frac{Y_A}{K_Y}\right)^{n_Y}} + l_{R_A} \right) - \gamma_{m_{R_A}} m_{R_A} \\ \dot{R}_A = \epsilon_{R_A} m_{R_A} - \gamma_{R_A} R_A \end{cases} . \tag{4.12}$$

The steady-state concentration of repressor for a constant input Y_A^* is given by

$$R_A^* = \frac{\epsilon_{R_A} V_{R_A} N_{R_A} C}{\gamma_{R_A}\gamma_{m_{R_A}}} \left(\frac{\left(\frac{Y_A^*}{K_Y}\right)^{n_Y}}{1+\left(\frac{Y_A^*}{K_Y}\right)^{n_Y}} + l_{R_A} \right). \tag{4.13}$$

The steady-state solutions differ from their complements only in parameter values.

4.8.3 Evaluating bistability

Depending on the exact parameter values, the same system of ODEs may describe a monostable or bistable system. For the cross-repression circuit that we have described, one way to determine whether a certain parameter set yields a bistable system is to count the number of steady-state points: If there is one point and it is stable, then the system is monostable; if there are three points total, one unstable and two stable (one each corresponding to the dominance of A or B), then the system is bistable. While steady-state points may be located using a convenient graphical method, it can be more difficult to analytically determine whether they are stable. Fortunately, if the system satisfies a set of technical conditions that classify it as *monotone*, then the graphical method will also reveal whether the equilibria are stable. Here, we show that our system is monotone and therefore we may use the graphical method to evaluate bistability.

Monotonicity

Consider the complete system as shown in Fig.4.4. The entire unit comprises twelve differential equations, two per each of the six modules. In vector form, we represent the entire system with state space

$$Z := \begin{bmatrix} m_{I_A} \\ I_A \\ X^T \\ m_{R_B} \\ R_B \\ m_{I_B} \\ I_B \\ Y^T \\ m_{R_A} \\ R_A \end{bmatrix}$$

as $\dot{Z} = f(Z)$. Let $J(\cdot)$ be the Jacobian of $f(\cdot)$.

For the system to be monotone, evolution of the system with time must preserve ordering in the state space, i.e., for solutions $\phi(\cdot, \cdot)$ to the

differential equation, $x_1(0) \preceq_K x_2(0) \implies \phi(t, x_1(0)) \preceq_K \phi(t, x_2(0))$ for all $t \geq 0$. The ordering is defined with respect to some positivity cone K in a Euclidean space. Although physical concentrations cannot be negative, due to the presence of inhibition our system is not monotone with respect to the positive orthant, but rather to a combination of positive and negative orthants. If we redefined the state space as $\tilde{Z} := [-m_{I_A},\ -I_A,\ -X^T,\ -m_{R_B},\ -R_B,\ m_{I_B},\ I_B,\ Y^T,\ m_{R_A},\ R_A]^T$, then $\tilde{Z}$ would be monotone with respect to $\mathbb{R}^{12}_+$. Hence Z is monotone with respect to the cone K spanned by all vectors $\tilde{Z}$ corresponding to feasible Z (i.e., $Z \in \mathbb{R}^{12}_+$).

To see that the system is monotone, we apply a graphical method described in (Angeli et al., 2004b) for systems with state, input, and output spaces defined by orthants. We proceed by constructing an incidence graph (signed digraph) where each node is a species and each edge describes the relationship between two distinct species: no edge if no direct interaction; $+$ if one promotes the other; and $-$ if one inhibits the other (Fig.4.5). By Z_j promotes Z_i we mean $J_{ij} := \frac{\partial f_i}{\partial Z_j} = \frac{\partial^2 Z_i}{\partial t \partial Z_j} \geq 0$ for all $Z_i, Z_j \in \mathbb{R}_+$ whereas by Z_j inhibits Z_i we mean $J_{ij} \leq 0$ for all $Z_i, Z_j \in \mathbb{R}_+$.

Because each biochemical species in the system directly affects only one other species in the system in a sequential fashion, the Jacobian for this system is sparse. We consider the entries corresponding to $\boldsymbol{H_A}$, $\boldsymbol{tx_{B\to A}}$, and $\boldsymbol{rx_A}$; because of the symmetry of the system, we know the entries corresponding to $\boldsymbol{H_B}$, $\boldsymbol{tx_{A\to B}}$, and $\boldsymbol{rx_B}$ will differ only in exact parameter values. Then the nonzero entries in the Jacobian (except those corresponding to one species' influence on its own concentration, which do not appear in the incidence graph) are

$$
\begin{aligned}
\boldsymbol{H_A}\ :\ & \begin{cases} \frac{\partial \dot{m}_{I_A}}{\partial R_A} = V_{I_A} N_{I_A} C \frac{-n_{R_A}\left(\frac{R_A}{K_{R_A}}\right)^{n_{R_A}-1}}{K_{R_A}\left(1+\left(\frac{R_A}{K_{R_A}}\right)^{n_{R_A}}\right)^2} & \leq 0 \\ \frac{\partial \dot{I}_A}{\partial m_{I_A}} = \epsilon_{I_A} & \geq 0 \end{cases} \\
\boldsymbol{tx_{B\to A}}\ :\ & \begin{cases} \frac{\partial \dot{Y}_B}{\partial I_B} = \nu_Y & \geq 0 \\ \frac{\partial \dot{Y}_A}{\partial Y_B} = d & \geq 0 \\ \frac{\partial \dot{Y}_B}{\partial Y_A} = d & \geq 0 \end{cases} \\
\boldsymbol{rx_A}\ :\ & \begin{cases} \frac{\partial \dot{m}_{R_A}}{\partial Y_A} = V_{R_A} N_{R_A} C \frac{n_Y\left(\frac{Y_A}{K_Y}\right)^{n_Y-1}}{K_Y\left(1+\left(\frac{Y_A}{K_Y}\right)^{n_Y}\right)^2} & \geq 0 \\ \frac{\partial \dot{R}_A}{\partial m_{R_A}} = \epsilon_{R_A} & \geq 0 \end{cases}
\end{aligned}
\qquad (4.14)
$$

along with the corresponding entries for the complementary modules.

From a conceptual standpoint, monotonicity means that regardless of context, an element always has the same qualitative effect on itself after its influence is propagated through the network; i.e., the influence of the element on itself is "consistent". From the graph for this system (shown in Fig.4.5) we see that there is only one cycle and it is positive in parity; that is, the product of the signs of each edge traversed to complete one cycle is positive, regardless of the direction of travel around the cycle. This implies that the graph is consistent, and hence the system it describes is closed-loop monotone.

Finding steady-state solutions

By Theorem 3 in (Angeli et al., 2004b) and (Angeli et al., 2004a), the equilibria of a closed-loop monotone system (system with feedback) can be found by examining an open-loop (input-output, or I/O system) monotone system formed by "breaking" the feedback of the original system. For fixed points to exist, the I/O system must have a static input-output characteristic.

In our case, we can define $\boldsymbol{T_A}$, or $R_A \rightarrow R_B$, as the cascade of the three modules $\boldsymbol{H_A}$, $\boldsymbol{tx_{A\rightarrow B}}$, and $\boldsymbol{rx_B}$ having input R_A and output R_B. With $\boldsymbol{T_B}$ defined similarly for the other three modules, then the entire system "broken" at R_A is the cascade of $\boldsymbol{T_A}$ and $\boldsymbol{T_B}$. This new I/O system accepts an input $u \in \mathbb{R}_+$ and produces an output $y \in \mathbb{R}_+$. If the I/O system admits a static input-output characteristic, then the points where $u = y$ are the steady-state solutions to the closed-loop system.

Since $\boldsymbol{H_A}$ and the combined cascade of $\boldsymbol{tx_{A\rightarrow B}}$ and $\boldsymbol{rx_B}$ each has a unique steady-state solution that is a global and asymptotically stable hyperbolic equilibrium (see Section 4.8.6), then the cascade of the three systems, i.e., $\boldsymbol{T_A}$, also has a static input-output characteristic. $\boldsymbol{T_B}$ is similarly endowed. Therefore the cascade of $\boldsymbol{T_A}$ and $\boldsymbol{T_B}$ also has a static input-output characteristic, and because the I/O system is also monotone, then its fixed points and the equilibrium points of the closed-loop system correspond.

Define $\boldsymbol{T_A}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ to be the static I/O characteristic of $\boldsymbol{T_A}$, i.e., for constant input $R^*_{A_i}$, $\boldsymbol{T_A}$ produces constant output $R^*_{B_o} = \boldsymbol{T_A}(R^*_{A_i})$, and define $\boldsymbol{T_B}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ to be the static I/O characteristic of $\boldsymbol{T_B}$ for constant input $R^*_{B_i}$ and output $R^*_{A_o}$. Then the static I/O characteristic for the cascade of $\boldsymbol{T_A}$ and $\boldsymbol{T_B}$ is $\boldsymbol{T_B}(\boldsymbol{T_A}(\cdot))$, which maps constant input $R^*_{A_i}$ to output $R^*_{A_o}$. The function $\boldsymbol{T_B}(\boldsymbol{T_A}(\cdot))$ is nonnegative and sigmoidal, meaning that

there must be exactly one or three intersection points between $y = u$ and $y = \boldsymbol{T_B}(\boldsymbol{T_A}(u))$ (the fixed-point solutions to the I/O system where $u = y$ or $R^*_{A_i} = R^*_{A_o}$). If there is only one intersection then the I/O characteristic at the intersection must have a slope less than unity ($\boldsymbol{T'_B}(\boldsymbol{T_A}(y^*))\boldsymbol{T'_A}(y^*) < 1$ where y^* is the intersection point), implying that the corresponding equilibrium is stable. If there are three intersections, then the middle intersection must have a slope greater than one while the higher and lower intersections must have slopes less than one, implying that the middle equilibrium is unstable and the other two equilibria are stable (Angeli et al., 2004b). Hence determining bistability amounts to graphically counting the intersections between $y = u$ and $y = \boldsymbol{T_B}(\boldsymbol{T_A}(u))$. We could also carry out the above analysis with equivalent results for $\boldsymbol{T_A}(\boldsymbol{T_B}(\cdot))$ mapping $R^*_{B_i}$ to $R^*_{B_o}$.

4.8.4 Multicompartmental systems with symmetry

Up to this point we have considered a system of only two compartments, one each of types A and B. The formation of an interesting pattern, however, requires more than two elements. To that end we will now consider a class of systems with multiple compartments of each type, where each compartment of type A is connected to the same number of compartments of type B, and vice versa. In other words, each compartment of type A or B is essentially indistinguishable from any other. The symmetry present in such a system will allow us to apply the graphical method of analyzing bistability to sets of compartments arranged in particular geometries.

The discussion presented here is not intended as a complete overview of multicompartmental systems of lateral inhibition, but rather as a supplement to our specific experiments. We refer the interested reader to (Rufino Ferreira et al., 2015) for a more generalized treatment.

Edge weight matrix

The interior mechanics of all compartments of the same type are the same, so the ODEs governing the behavior of the reception modules $\boldsymbol{rx}$ and repression modules $\boldsymbol{H}$, as defined in Section 4.8.14.8.2, are the same for compartments of the same type. Adding more compartments does, however, change the concentration of diffusible signaling molecules that reach the compartments, and therefore changes the behavior of the transmission modules $\boldsymbol{tx}$.

We begin by noting that if we define the 2×2 matrix

$$L_2 := d \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

then we can reformulate (4.11) as

$$\dot{Y} = L_2 Y - \gamma_Y Y + \begin{bmatrix} 0 \\ \nu_Y I_B \end{bmatrix} \tag{4.15}$$

with steady-state solution

$$Y^* = (-L_2 + \gamma_Y I)^{-1} \begin{bmatrix} 0 \\ \nu_Y I_B^* \end{bmatrix}. \tag{4.16}$$

The matrix L_2 contains information on the diffusion of AHL between compartments. We now generalize to systems with N compartments and the corresponding $N \times N$ matrix L_N (which we will henceforth designate as simply L). If all compartments in the system are numbered from 1 to N, then the element $[L]_{ij}$ represents the connection strength between compartments i and j. Conceptually, the connection strength is the diffusion into compartment j from compartment i (or vice versa) if $i \neq j$, and the total diffusion out of a given compartment to all other compartments if $i = j$. Each element $[L]_{ij}$ is directly proportional to the diffusivity D of AHL and inversely proportional to the square of the distance between i and j. In the special case where there are only two compartments, the connection strength between them is identical, hence the elements in L_2 are all of magnitude d.

Mathematically, we can represent the multicompartmental system as an undirected graph where each vertex is a compartment and each edge is a channel. Let d_{ij} be the edge weight between vertices i and j. L is the Laplacian of this graph:

$$[L]_{ij} = \begin{cases} -\sum_{j=1}^{N} d_{ij} & i = j \\ d_{ij} & i \neq j \end{cases}.$$

Let N_A be the number of compartments of type A and N_B the number of compartments of type B such that $N_A + N_B = N$. Assume the diffusivity of AHL is constant, all channels have the same length, and compartments of the same type are not connected to each other. Then $d_{ij} = 0$ if i, j are of

the same type and $d_{ij} = d$ between connected compartments i, j of opposite types. Assume each compartment of type A is connected to q_B compartments of type B and each compartment of type B is connected to q_A compartments of type A.

Let the first N_A entries of a row or column of $L \in \mathbb{R}^{N\times N}$ designate compartments of type A and the last N_B entries designate compartments of type B. Then L has the form

$$L = d\begin{bmatrix} -q_B I_{N_A\times N_A} & F \\ F^T & -q_A I_{N_B\times N_B} \end{bmatrix}$$

where F is an $N_A \times N_B$ matrix for which $[F]_{ij} = 0$ indicates that the ith compartment of type A and the jth compartment of type B are not connected by a channel, and $[F]_{ij} = 1$ indicates that they are.

As shown in the following sections, the assumed structure of the system allows us to reduce our N-dimensional system to a two-dimensional one, which greatly simplifies the calculations for a steady-state contrasting pattern and enables us to use the graphical method detailed in Section 4.8.34.8.3 to determine when the overall system is bistable.

Model reduction

We use the assumptions and notations from the previous subsection. Define $M \in \mathbb{R}^{N\times 2}$ as

$$M := \begin{bmatrix} \mathbb{1}_{N_A} & 0_{N_A} \\ 0_{N_B} & \mathbb{1}_{N_B} \end{bmatrix} \tag{4.17}$$

where $\mathbb{1}_n$ designates a length-n vector of all ones and 0_n designates a length-n vector of all zeros. Then because all compartments of the same type have the same number of connections to compartments of the opposite type, there exists some $\bar{L} \in \mathbb{R}^{2\times 2}$ such that

$$LM = M\bar{L}.$$

Let $d_A := dq_B$ be the (nonnegative) total outgoing edge weight for a compartment of type A and $d_B := dq_A$ be the (nonnegative) total outgoing edge weight for a compartment of type B. Because we have assumed no connections between compartments of the same type, $\bar{L}$ has the form

$$\bar{L} = \begin{bmatrix} -d_A & d_A \\ d_B & -d_B \end{bmatrix}. \tag{4.18}$$

We would like to solve for the steady-state value of $X = [X_A,\ X_B]^T \in \mathbb{R}^N$ in $\boldsymbol{tx_{A\to B}}$:

$$(-L + \gamma_X I_{N\times N})\, X = \begin{bmatrix} \nu_X I_A \mathbb{1}_{N_A} \\ 0_{N_B} \end{bmatrix}. \tag{4.19}$$

We restrict our search to a subset of solutions for which the variables of interest are identical among compartments of the same type, i.e.,

$$X = M \begin{bmatrix} x_A \\ x_B \end{bmatrix} \tag{4.20}$$

for $x_A, x_B \in \mathbb{R}$. Let $x := [x_A,\ x_B]^T$. We can then rewrite (4.19) as

$$\begin{aligned} (-L + \gamma_X I_{N\times N})\, Mx &= M \begin{bmatrix} \nu_X I_A \\ 0 \end{bmatrix} \\ \implies M\left(-\bar{L} + \gamma_X I_{2\times 2}\right) x &= M \begin{bmatrix} \nu_X I_A \\ 0 \end{bmatrix} \end{aligned}$$

which implies that solutions x^* to

$$\begin{aligned} \left(-\bar{L} + \gamma_X I\right) x^* &= \begin{bmatrix} \nu_X I_A^* \\ 0 \end{bmatrix} \\ \implies x^* &= \left(-\bar{L} + \gamma_X I\right)^{-1} \begin{bmatrix} \nu_X I_A^* \\ 0 \end{bmatrix} \end{aligned} \tag{4.21}$$

provide solutions to (4.19) by way of (4.20). In other words, x_A^* is the steady-state concentration of X in any compartment of type A and x_B^* is the steady-state concentration of X in any compartment of type B for constant input I_A^*. The derivation for $\boldsymbol{tx_{B\to A}}$ proceeds similarly.

Now recall from (4.18) that the matrix $\bar{L}$ has form

$$\bar{L} = \begin{bmatrix} -d_A & d_A \\ d_B & -d_B \end{bmatrix}.$$

Then the inverse matrix in (4.21) can be directly evaluated, yielding

$$\begin{aligned} \left(-\bar{L} + \gamma_X I\right)^{-1} &= \frac{1}{(d_A + \gamma_X)(d_B + \gamma_X) - d_A d_B} \begin{bmatrix} d_B + \gamma_X & d_A \\ d_B & d_A + \gamma_X \end{bmatrix} \\ &= \frac{1}{\gamma_X(\gamma_X + d_A + d_B)} \begin{bmatrix} d_B + \gamma_X & d_A \\ d_B & d_A + \gamma_X \end{bmatrix}. \end{aligned}$$

Since d_A, $d_B > 0$, the matrix is always invertible provided that $\gamma_X \neq 0$.

In essence, the new multicompartmental system is identical to the two-compartmental system with a revision to the transmission modules:

$$\boldsymbol{tx_{B\to A}} \; : \; \begin{cases} \dot{Y}_A = d_A(Y_B - Y_A) - \gamma_Y Y_A \\ \dot{Y}_B = d_B(Y_A - Y_B) - \gamma_Y Y_B + \nu_Y I_B \end{cases}$$

and similarly for $\boldsymbol{tx_{A\to B}}$. Since $d_A, d_B > 0$ the Jacobian equations from (4.14) maintain their parity and the analysis developed in Section 4.8.34.8.3 holds. In Section 54.8.6 we show that the local stability or instability of a steady state in the reduced system implies local stability or instability of the state in the full system, and therefore we can therefore continue to use the graphical intersection method to determine when the system admits a solution where all cells of the same type are identically (and reversibly) high or low.

Finite differences within the channel

Thus far, implicit in our definition of L is that we approximate diffusion between compartments using the method of finite differences for a step size of l, the channel length. Specifically, we have taken Fick's diffusion equation in one dimension

$$\frac{dX}{dt} = D\frac{d^2X}{dr^2}$$

where X is the concentration of some species and r is distance. Under the finite differences approximation, we discretize space along r and approximate the change in concentration at each point as

$$\frac{dX}{dt} \approx D\frac{(X(r+\Delta) - X(r)) + (X(r-\Delta) - X(r))}{\Delta^2}.$$

In the process of discretizing r we pick two boundary points. To model a single channel connecting a cell of type A to one of type B we simply pick one end, say $r = 0$, to correspond to A and the other end, $r = l$, to correspond to B. We assume no diffusion outside the channel. Since $X(0)$ has no neighbors $r < 0$ and $X(l)$ has no neighboring points $r > l$, the approximation at the boundaries is performed using only one difference. At the $r = 0$ boundary we have

$$\frac{dX(0)}{dt} \approx D\frac{X(\Delta) - X(0)}{\Delta^2}$$

and at the $r = l$ boundary

$$\frac{dX(l)}{dt} \approx D\frac{X(l-\Delta) - X(l)}{\Delta^2}.$$

If we let $\Delta = l$, $X(0) = X_A$, and $X(l) = X_B$, we recover the familiar

$$\begin{cases} \frac{dX_A}{dt} = \frac{D}{l^2}(X_B - X_A) \\ \frac{dX_B}{dt} = \frac{D}{l^2}(X_A - X_B) \end{cases}.$$

If l is sufficiently small, this approximation is appropriate. If the channel is too long, however, then setting $\Delta = l$ produces an extremely coarse approximation that may not be an accurate description of the physical process. Suppose now we discretize r such that there are N_L points between $r = 0$ and $r = l$, with a step size of $\Delta = \frac{l}{N_L+1}$. The equation to describe diffusion is written in matrix form as

$$\begin{bmatrix} \dot{X}(0) \\ \dot{X}(\Delta) \\ \dot{X}(2\Delta) \\ \vdots \\ \dot{X}(l-\Delta) \\ \dot{X}(l) \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & ... & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} X(0) \\ X(\Delta) \\ X(2\Delta) \\ \vdots \\ X(l-\Delta) \\ X(l) \end{bmatrix}.$$

Note that this describes only the diffusion, not the reactions that occur within $X(0)$ and $X(l)$.

We can accommodate multiple cells with multiple channels by adding more terms and associated concentration variables. Imagine that we number the discretized points within a channel from 1 to N_L where 1 is the point nearest to a cell of type A and N_L is the point nearest to a cell of type B. Then we assume all channels with the same number form a similar class; i.e., if we have N_C channels we define $M \in \mathbb{R}^{N+N_C N_L \times N_C+2}$ as

$$M := \begin{bmatrix} \nVdash_{N_A} & 0_{N_L} & 0_{N_L} & ... & 0_{N_L} \\ 0_{N_L} & \nVdash_{N_L} & 0_{N_L} & ... & 0_{N_L} \\ 0_{N_L} & 0_{N_L} & \nVdash_{N_L} & ... & 0_{N_L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{N_L} & 0_{N_L} & 0_{N_L} & ... & \nVdash_{N_B} \end{bmatrix}$$

such that

$$\bar{L} := \frac{D}{\Delta^2}\begin{bmatrix} -q_B & q_B & 0 & 0 & ... & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & ... & 0 & -q_A & q_A \end{bmatrix}$$

with q_A and q_B as defined in Section 4.8.44.8.4. We then rewrite (4.19) as

$$\left(-\bar{L} + \gamma_X I\right) X = \begin{bmatrix} \nu_X I_A \\ 0_{N_L N_C + 1} \end{bmatrix}.$$

The equation takes a similar form for Y.

In the above discussion we have assumed that production takes place only in single compartments, which shrink in size with increasing N_L. Let w be the width of a cell colony. To accurately reckon with arbitrary l and N_C we would have to either (a) dilute the concentrations of intercellular components ($l > w$); or (b) designate multiple compartments as production compartments, which would rapidly increase the size of the ODE system ($l < w$). To avoid these situations we set $\Delta = w$ such that $l = kw$ where $k \geq 0$ is an integer. This limitation in the considered range of l is justified by the small range of l that we implement expermientally as well as the minimal effect of l on the theoretically evaluated bistability of the system (for systems with and without multiple diffusive compartments).

Extension to contact-mediated communication

To represent forms of communication other than diffusion, the matrix L representing the Laplacian can be replaced by any matrix with entires appropriately weighted to reflect the strength of the connections between compartments. For example, to model communication for which compartments do not lose concentration of signaling molecule to communicate with neighbors, the diagonal entries can be set to 0. Using this particular form of L instead of the Laplacian in the remainder of our theory for lateral inhibition allows us to model traditional cell-to-cell contact-based mechanisms such as the Notch/Delta system. Note that if L is an asymmetric matrix, dimensionality reduction as described here for particular spatial configurations will in general no longer be mathematically valid.

4.8.5 Results: bistable regimes

Parameter ranges

We assessed the effect of changing the parameters in Table 4.3 on the bistability of the system for a given connectivity of compartments—hereafter referred to as a geometry—satisfying the assumptions in Section 4.8.44.8.4. We used a custom MATLAB script to generate "bistability maps" as a function of two parameters at a time to determine in which parameter regimes the system is theoretically bistable. Such maps are useful to elucidate not only whether the system falls in the bistable region, but also its "location" on the bistability map. The location is important for two reasons: (a) the closer a system is to the edge of the bistable region, the smaller a change to the parameters is necessary to push the system into or out of the region; and (b) where the system is in the region determines the system's exact steady-state values, which in turn determine the observed contrast between the colonies' expression levels of reporter. These maps are useful in designing bacterial strains to achieve desired behaviors (i.e., bistability, steady-state values) and also for picking geometric configurations that produce those behaviors. For example, a system farther from the edge of the bistable region will be more robust in the sense that perturbations to parameter values (as through experimental error or random variation) are less likely to cause the system to become bistable if it is nominally monostable or monostable if it is nominally bistable.

Dynamic range matching

In essence, a system of mutual inhibition is equivalent to a pair of switches, each of which turns on for low input and off for high input. The switches cascade to form an open-loop system, which is closed by connecting the output of the second switch to the input of the first. In order for the closed-loop system to be bistable, the input to each switch must span some significant fraction of the dynamic range of the switch; i.e., the minimum output from the first switch should be low enough to turn on the second switch while the maximum output should be high enough to turn it off, and vice versa. The relative dynamic ranges of the switches are determined by the parameters and geometry. The curves in Fig.4.4 show input/output curves (nullclines) for the first switch with input X_B and output Y_A (modules $\mathbf{rx_B} \rightarrow \mathbf{H_B} \rightarrow \mathbf{tx_{B\rightarrow A}}$) and the second switch with input Y_A and output X_B (modules $\mathbf{rx_A} \rightarrow \mathbf{H_A} \rightarrow \mathbf{tx_{A\rightarrow B}}$).

4.8.6 Additional proofs

Contrasting steady states emerge if and only if the system can be reduced to the two-compartment case

In Section 4.8.44.8.4 we showed that if compartments of the same type have the same number of neighbors, then the system can be reduced to a two-compartment case for the purposes of identifying contrasting steady states, i.e., those in which all compartments of the same type have the same steady-state concentrations. We now show that if a steady state exists for which (a) all compartments of the same type have the same concentrations and (b) every compartment of type A is connected to at least one compartment of type B and vice versa, then the underlying edge weight matrix L can be reduced to a two-compartment system. We do not need to assume that compartments of the same type must have neighbors *only* of the opposite type, although a "checkerboard" of alternating high/low immediately adjacent neighbors will require this arrangement. We do not even require that all channels be of the same length. We only require that for a fixed compartment type, the total incoming edge weight from other compartments of the same type as well as the incoming edge weight from other compartments of the opposite type is the same regardless of the choice of compartment.

First, we assume that there exists a steady-state solution of the form

$$X^* = \begin{bmatrix} X_A^* \\ X_B^* \end{bmatrix} = \begin{bmatrix} \mathbb{1}_{N_A} & 0_{N_A} \\ 0_{N_B} & \mathbb{1}_{N_B} \end{bmatrix} \begin{bmatrix} x_A^* \\ x_B^* \end{bmatrix} =: Mx^*$$

where $x_A, x_B \in \mathbb{R}$ as in Section 4.8.44.8.4. X^* a steady-state solution implies that

$$\begin{aligned} (-L + \gamma_X I_N)\, X^* &= \begin{bmatrix} \nu_X I_A^* \mathbb{1}_{N_A} \\ 0_{N_B} \end{bmatrix} \\ \implies (-L + \gamma_X I_N)\, Mx^* &= M \begin{bmatrix} \nu_X I_A^* \\ 0 \end{bmatrix} \end{aligned} \tag{4.22}$$

since there cannot be multiple values of I_A^* that result in the same X_A^*. Now define

$$L = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}$$

where $L_{11} \in \mathbb{R}^{N_A \times N_A}$, $L_{12} = L_{21}^T \in \mathbb{R}^{N_A \times N_B}$, and $L_{22} \in \mathbb{R}^{N_B \times N_B}$, and observe that

$$(-L + \gamma_X I_N) M = \begin{bmatrix} -\sum_{j=1}^{N_A} [L_{11}]_j + \gamma_X \nVdash_{N_A} & -\sum_{j=1}^{N_B} [L_{12}]_j \\ -\sum_{j=1}^{N_A} [L_{21}]_j & -\sum_{j=1}^{N_B} [L_{22}]_j + \gamma_X \nVdash_{N_B} \end{bmatrix} \tag{4.23}$$

where $[L_{nm}]_j$ denotes the jth column of L_{nm}, $n, m = 1, 2$. Now define

$$Q_{nm}^i = \begin{cases} -\sum_{j=1}^{N_A} [L_{nm}]_{ij}, & m = 1 \\ -\sum_{j=1}^{N_B} [L_{nm}]_{ij}, & m = 2 \end{cases}.$$

Combining (4.23) with (4.22) yields

$$\begin{cases} (Q_{11}^a + \gamma_X) x_A^* + (Q_{12}^a) x_B^* = \nu_X I_A^*, & a = 1, 2, ..., N_A \\ (Q_{21}^b) x_A^* + (Q_{22}^b + \gamma_X) x_B^* = 0, & b = N_A + 1, ..., N_A + N_B \end{cases} \tag{4.24}$$

and similarly for Y^*,

$$\begin{cases} (Q_{11}^a + \gamma_Y) y_A^* + (Q_{12}^a) y_B^* = 0, & a = 1, 2, ..., N_A \\ (Q_{21}^b) y_A^* + (Q_{22}^b + \gamma_Y) y_B^* = \nu_Y I_B^*, & b = N_A + 1, ..., N_A + N_B \end{cases}. \tag{4.25}$$

Suppose we fix a and b and construct a system of four equations from (4.24) and (4.25). In these four equations there are four "unknowns" $(Q_{11}^a, Q_{12}^a, Q_{21}^b, Q_{22}^b)$, suggesting that there is a unique solution to

$$\begin{bmatrix} x_A^* & x_B^* & 0 & 0 \\ 0 & 0 & x_A^* & x_B^* \\ y_A^* & y_B^* & 0 & 0 \\ 0 & 0 & y_A^* & y_B^* \end{bmatrix} \begin{bmatrix} Q_{11}^a \\ Q_{12}^a \\ Q_{21}^b \\ Q_{21}^b \end{bmatrix} = \begin{bmatrix} \nu_X I_A^* - \gamma_X x_A^* \\ -\gamma_X x_B^* \\ -\gamma_Y y_A^* \\ \nu_Y I_B^* - \gamma_Y y_B^* \end{bmatrix} \tag{4.26}$$

provided that the LHS matrix is full rank, i.e., $\frac{x_A^*}{x_B^*} \neq \frac{y_A^*}{y_B^*}$. This unique solution is the same regardless of the choice of a and b, which implies that Q_{11}^a, Q_{12}^a must be the same for all a and Q_{21}^b, Q_{22}^b must be the same for all b (since the solution is unique). This, in turn, implies that we can write

$$LM = M \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} =: M\bar{L}.$$

The only fact left to show, then, is that our system cannot admit a solution

for which $\frac{x_A^*}{x_B^*} = \frac{y_A^*}{y_B^*}$, which will guarantee a unique solution to (4.26). Our proof leverages the lateral inhibition structure of the system.

Suppose $\frac{x_A^*}{x_B^*} = \frac{y_A^*}{y_B^*}$ and without loss of generality, define $r := \frac{x_A^*}{y_A^*} = \frac{x_B^*}{y_B^*}$. Substituting $x_A^* = ry_A^*$ and $x_B^* = ry_B^*$ into (4.24) and (4.25) for fixed a, b (equivalently (4.26)), we obtain

$$\begin{cases} \left(Q_{11}^a + \gamma_X\right) x_A^* + Q_{12}^a x_B^* = \nu_X I_A^* \\ \left(Q_{11}^a + \gamma_Y\right) x_A^* + Q_{12}^a x_B^* = 0 \\ Q_{21}^b x_A^* + \left(Q_{22}^b + \gamma_X\right) x_B^* = 0 \\ Q_{21}^b x_A^* + \left(Q_{22}^b + \gamma_Y\right) x_B^* = r\nu_Y I_B^* \end{cases} \implies \begin{cases} x_A^* = \frac{-\left(Q_{22}^b+\gamma_X\right)x_B^*}{Q_{21}^b} = \frac{r\nu_Y I_B^* - \left(Q_{22}^b+\gamma_Y\right)x_B^*}{Q_{21}^b} \\ x_B^* = \frac{\nu_X I_A^* - \left(Q_{11}^a+\gamma_X\right)x_A^*}{Q_{12}^a} = \frac{-\left(Q_{11}^a+\gamma_Y\right)x_A^*}{Q_{12}^a} \end{cases}.$$

The assumption that every compartment of type A is connected to at least one cell of type B and vice versa guarantees that $Q_{12}^a \neq 0, Q_{21}^b \neq 0$. Rearranging terms, we find

$$\begin{cases} \left(\gamma_Y - \gamma_X\right) x_B^* = r\nu_Y I_B^* \\ \left(\gamma_X - \gamma_Y\right) x_A^* = \nu_Y I_A^* \end{cases}. \tag{4.27}$$

Now we know $\gamma_X, \gamma_Y, \nu_X, \nu_Y > 0$ and since $x_A^*, x_B^*, I_A^*, I_B^* \geq 0$ (we cannot have negative concentrations) we also know $r \geq 0$. Because of the mutual inhibition relationship we further know that $I_A^* =: h_1(rx_A^*)$ is a bounded, nonnegative, nonincreasing function of x_A^* and $I_B^* =: h_2(x_B^*)$ is a bounded, nonnegative, nonincreasing function of x_B^*. If we are concerned with a nontrivial system we must have $r \neq 0$ because otherwise

$$r = 0 \implies x_A^* = 0 \implies I_A^* = 0 \implies h_1(0) = 0 \implies h_1(z \geq 0) = 0 \; \forall z,$$

i.e., cells of type A are insensitive to inputs. Similarly, we must have r finite because we could as easily have set up the system in terms of y_A^* and y_B^* with $\frac{1}{r}$ as the ratio of interest, and following the same logic we would have (with slight abuse of notation) $\frac{1}{r} \neq 0$.

Therefore, since $r > 0$, then from (4.27), nonnegative x_B^*, I_B^* requires $\gamma_Y \geq \gamma_X$ while nonnegative x_A^*, I_A^* requires $\gamma_X \geq \gamma_Y$. These two conditions can only be satisfied if $\gamma_X = \gamma_Y$, which implies that $I_A^* = I_B^* = 0$. But this is impossible from the definition of h_1, h_2 in (4.6) (except in the limit as $R_A^*, R_B^* \to \infty$, which is anyway unattainable because R_A^*, R_B^* are bounded). Hence we cannot have a steady-state contrasting solution for

which $\frac{x_A^*}{x_B^*} = \frac{y_A^*}{y_B^*}$, implying that the matrix in (4.26) will have a unique solution. This completes our claim.

Local (in)stability of full system about reduced system steady states

Let $Z \in \mathbb{R}_+^{6N}$ be the vector of states for the full system where $\dot{Z} = f(Z)$ governed by the equations in Section 4.8.14.8.2. Let Z^* be a steady state where all cells of the same type have identical states, i.e., a steady state identified in the reduced system. Assume the states are ordered such that

$$Z^* := \begin{bmatrix} Y^* \\ m_{R_A}^* \mathbb{1}_{N_A} \\ R_A^* \mathbb{1}_{N_A} \\ m_{I_A}^* \mathbb{1}_{N_A} \\ I_A^* \mathbb{1}_{N_A} \\ X^* \\ m_{R_B}^* \mathbb{1}_{N_B} \\ R_B^* \mathbb{1}_{N_B} \\ m_{I_B}^* \mathbb{1}_{N_B} \\ I_B^* \mathbb{1}_{N_B} \end{bmatrix}.$$

Define

$$K_{A1} := \left.\frac{\partial \dot{m}_{R_A}}{\partial Y_A}\right|_{Y_A^*} = V_{R_A} N_{R_A} C \frac{n_Y \left(\frac{Y_A^*}{K_Y}\right)^{n_Y - 1}}{K_Y \left(1 + \left(\frac{Y_A^*}{K_Y}\right)^{n_Y}\right)^2}$$

$$K_{A2} := -\left.\frac{\partial \dot{m}_{I_A}}{\partial R_A}\right|_{R_A^*} = V_{I_A} N_{I_A} C \frac{n_{R_A} \left(\frac{R_A^*}{K_{R_A}}\right)^{n_{R_A} - 1}}{K_{R_A} \left(1 + \left(\frac{R_A^*}{K_{R_A}}\right)^{n_{R_A}}\right)^2}.$$

K_{B1} and K_{B2} are defined analogously with the appropriate subscripts for cell B. Then the Jacobian of the full system evaluated at the steady state Z^* is given by

$$J = \begin{bmatrix} S_A & P_Y \\ P_X & S_B \end{bmatrix} \tag{4.28}$$

where $J \in \mathbb{R}^{6N\times 6N}$, $S_A \in \mathbb{R}^{N+4N_A\times N+4N_A}$, $S_B \in \mathbb{R}^{N+4N_B\times N+4N_B}$, $P_Y \in \mathbb{R}^{N+4N_A\times N+4N_B}$, and $P_X \in \mathbb{R}^{N+4N_B\times N+4N_A}$, defined as

$$S_A := \begin{bmatrix} L-\gamma_Y I_N & 0 & 0 & 0 & 0 \\ \begin{bmatrix} K_{A1}I_{N_A} & 0 \end{bmatrix} & -\gamma_{m_{R_A}} I_{N_A} & 0 & 0 & 0 \\ 0 & \epsilon_{R_A} I_{N_A} & -\gamma_{R_A} I_{N_A} & 0 & 0 \\ 0 & 0 & -K_{A2} I_{N_A} & -\gamma_{m_{I_A}} I_{N_A} & 0 \\ 0 & 0 & 0 & \epsilon_{I_A} I_{N_A} & -\gamma_{I_A} I_{N_A} \end{bmatrix}$$

$$P_Y := \begin{bmatrix} 0_{N\times N+3N_B} & \begin{bmatrix} 0 \\ \nu_Y I_{N_B} \end{bmatrix} \\ 0_{4N_A\times N+3N_B} & 0_{4N_A\times N_B} \end{bmatrix}$$

$$P_X := \begin{bmatrix} 0_{N\times N+3N_A} & \begin{bmatrix} \nu_X I_{N_A} \\ 0 \end{bmatrix} \\ 0_{4N_B\times N+3N_A} & 0_{3N_A\times N_A} \end{bmatrix}$$

$$S_B := \begin{bmatrix} L-\gamma_X I_N & 0 & 0 & 0 & 0 \\ \begin{bmatrix} 0 & K_{B1}I_{N_B} \end{bmatrix} & -\gamma_{m_{R_B}} I_{N_B} & 0 & 0 & 0 \\ 0 & \epsilon_{R_B} I_{N_B} & -\gamma_{R_B} I_{N_B} & 0 & 0 \\ 0 & 0 & -K_{B2} I_{N_B} & -\gamma_{m_{I_B}} I_{N_B} & 0 \\ 0 & 0 & 0 & \epsilon_{I_B} I_{N_B} & -\gamma_{I_B} I_{N_B} \end{bmatrix}$$

The matrix measure of a matrix M with respect to the one-norm is defined as

$$\mu(M) := \max_j \left\{ M_{jj} + \sum_{i\neq j} |M_{ij}| \right\}.$$

If there exists an invertible diagonal matrix D such that

$$\mu(DJD^{-1}) < 0, \tag{4.29}$$

then the mapping described by J is contractive. This implies that the eigenvalues are negative, which for our nonlinear system means that the steady state around which J is linearized is locally stable.

Consider the reduced version of the full system, which has a Jacobian $\bar{J}$ identical in form to (4.28) for $N_A = 1$, $N_B = 1$, and $L = \bar{L}$ as defined in Section 4.8.44.8.4:

$$\bar{L} := \begin{bmatrix} -d_A & d_A \\ d_B & -d_B \end{bmatrix}$$

where d_A is the total (nonnegative) outgoing edge weight for a cell of type A and d_B is the total (nonnegative) outgoing edge weight for a cell of type

B. Let $\bar{D}$ take the form

$$\bar{D} = diag\,(l_{1A}, l_{1B}, a_1, a_2, a_3, a_4, l_{2A}, l_{2B}, b_1, b_2, b_3, b_4) \tag{4.30}$$

where $l_{1A}, l_{1B}, a_1, a_2, a_3, a_4, l_{2A}, l_{2B}, b_1, b_2, b_3, b_4 \in \mathbb{R}$ are arbitrary constants. For $\bar{D}$ to satisfy (4.29), we require

$$\begin{aligned}
-\gamma_Y + \frac{a_1}{l_{1A}} K_{A1} - d_A + \frac{l_{1B}}{l_{1A}} d_A &< 0 \\
-\gamma_Y - d_B + \frac{l_{1A}}{l_{1B}} d_B &< 0 \\
-\gamma_{m_{R_A}} + \frac{a_2}{a_1} \epsilon_{R_A} &< 0 \\
-\gamma_{R_A} + \frac{a_3}{a_2} K_{A2} &< 0 \\
-\gamma_{m_{I_A}} + \frac{a_4}{a_3} \epsilon_{I_A} &< 0 \\
-\gamma_{I_A} + \frac{l_{2A}}{a_4} \nu_X &< 0
\end{aligned}$$

and analogously for columns corresponding to compartments B.

We can combine the inequalities to obtain a single expression if we rearrange them to reflect the relative sizes of certain constants. Then we obtain

$$\frac{a_1 K_{A1} + l_{1B} d_A}{\gamma_Y + d_A} < l_{1A} \tag{4.31}$$

$$\begin{aligned}
l_{1A} &< \frac{\gamma_Y + d_B}{d_B} l_{1B} \qquad (4.32) \\
a_2 &< \frac{\gamma_{m_{R_A}}}{\epsilon_{R_A}} a_1 \\
a_3 &< \frac{\gamma_{R_A}}{K_{A2}} a_2 \\
a_4 &< \frac{\gamma_{m_{I_A}}}{\epsilon_{I_A}} a_3 \\
l_{2A} &< \frac{\gamma_{I_A}}{\nu_X} a_4.
\end{aligned}$$

We can rearrange (4.31) and (4.32) as

$$a_1 < \frac{\gamma_Y(\gamma_Y + d_B + d_A)}{K_{A1} d_B}$$

whereupon the combination of inequalities yields

$$l_{2A} < \frac{\gamma_{I_A}\gamma_{m_{I_A}}\gamma_{R_A}\gamma_{m_{R_A}}\gamma_Y(\gamma_Y + d_B + d_A)}{\nu_X\epsilon_{I_A}K_{A2}\epsilon_{R_A}K_{A1}d_B} l_{1B} =: \frac{1}{C_1} l_{1B}.$$

Similarly, for compartments B,

$$l_{1B} < \frac{\gamma_{I_B}\gamma_{m_{I_B}}\gamma_{R_B}\gamma_{m_{R_B}}\gamma_X(\gamma_X + d_B + d_A)}{\nu_Y\epsilon_{I_B}K_{B2}\epsilon_{R_B}K_{B1}d_A} l_{1A} =: \frac{1}{C_2} l_{2A}$$

such that

$$l_{2A} < \frac{1}{C_1 C_2} l_{2A} \implies C_1 C_2 < 1.$$

Incidentally,

$$C_1 C_2 = \left(\frac{\epsilon_{I_A}K_{A2}}{\gamma_{I_A}\gamma_{m_{I_A}}}\right)\left(\frac{\epsilon_{R_B}K_{B1}}{\gamma_{R_B}\gamma_{m_{R_B}}}\right)\left(\frac{d_B\nu_X}{\gamma_X(\gamma_X + d_A + d_B)}\right) \times \left(\frac{\epsilon_{I_B}K_{B2}}{\gamma_{I_B}\gamma_{m_{I_B}}}\right)\left(\frac{\epsilon_{R_A}K_{A1}}{\gamma_{R_A}\gamma_{m_{R_A}}}\right)\left(\frac{d_A\nu_Y}{\gamma_Y(\gamma_Y + d_A + d_B)}\right)$$

can be written as

$$\left(\frac{dI_A^*}{dR_{A_i}^*}\bigg|_{R_A^*}\frac{dX_B^*}{dI_A^*}\bigg|_{I_A^*}\frac{dR_B^*}{dX_B^*}\bigg|_{X_B^*}\right)\left(\frac{dI_B^*}{dR_B^*}\bigg|_{R_B^*}\frac{dY_A^*}{dI_B^*}\bigg|_{I_B^*}\frac{dR_{A_o}^*}{dY_A^*}\bigg|_{Y_A^*}\right) = \frac{dR_{A_o}^*}{dR_{A_i}^*}\bigg|_{R_A^*}, \tag{4.33}$$

where $R_{A_i}^*$ is a constant input and $R_{A_o}^*$ is the associated output. Hence (4.33) is the slope of the I/O system evaluated at the steady state $R_{A_i}^* = R_A^*$, which matches the graphical stability test developed in Section 4.8.34.8.3.

Now consider the full system with Jacobian (4.28) corresponding to the reduced system for which $\bar{D}$ satisfies (4.29). We construct a D for the full system that has diagonal entries

$$[l_{1A}I_{N_A}, l_{1B}I_{N_A}, a_1 I_{N_A}, a_2 I_{N_A}, a_3 I_{N_A}, a_4 I_{N_A}, l_{2A}I_{N_B}, l_{2B}I_{N_B}, b_1 I_{N_B}, b_2 I_{N_B}, b_3 I_{N_B}, b_4 I_{N_B}] \tag{4.34}$$

where the constants are the same as for $\bar{D}$. DJD^{-1} is then effectively organized into "blocks" that correspond to columns of $\bar{D}\bar{J}\bar{D}^{-1}$. Let

$$m_j(M) := M_{jj} + \sum_{i \neq j} |M_{ij}|,$$

i.e., $\mu(M) = \max_j m_j(M)$. Because J is linearized about a reduced-system steady state, it is straightforward to see that

$$m_j(J) = \begin{cases} m_1(\bar{J}), & j = 1, 2, ..., N_A \\ m_2(\bar{J}), & j = N_A + 1, N_A + 2, ..., N \\ m_3(\bar{J}), & j = N + 1, N + 2, ..., N + N_A \\ m_4(\bar{J}), & j = N + N_A + 1, ..., N + 2N_A \\ m_5(\bar{J}), & j = N + 2N_A + 1, ..., N + 3N_A \\ m_6(\bar{J}), & j = N + 3N_A + 1, ..., N + 4N_A \\ m_7(\bar{J}), & j = N + 4N_A + 1, ..., N + 5N_A \\ m_8(\bar{J}), & j = N + 5N_A + 1, ..., 2N + 4N_A \\ m_9(\bar{J}), & j = 2N + 4N_A + 1, ..., 3N + 3N_A \\ m_{10}(\bar{J}), & j = 3N + 3N_A + 1, ..., 4N + 2N_A \\ m_{11}(\bar{J}), & j = 4N + 2N_A + 1, ..., 5N + N_A \\ m_{12}(\bar{J}), & j = 5N + N_A, ..., 6N \end{cases}.$$

(Note that the equivalence of $m_j(J)$ to $m_1(\bar{J})$, $m_2(\bar{J})$, $m_7(\bar{J})$, and $m_8(\bar{J})$ for appropriate j arises out of the form of L, which has diagonal entries $-d_A$ and $-d_B$ repeated N_A and N_B times respectively, with non-diagonal entries summing to d_A for the first N_A columns and d_B for the last N_B columns.)

Because $m_j(\bar{J}) < 0$ for $j = 1, 2, ..., 12$, then $m_j(J) < 0$ for $j = 1, 2, ..., 6N$. Therefore for J linearized about a given reduced-system steady state, the conditions required for J to be contractive—and therefore for the full system to be locally stable at the steady state—is the same as the graphical condition for the reduced system to be stable or unstable at that steady state, namely, that the slope of the I/O system must be less than 1. Conversely, if (4.33) is greater than 1, the steady state in the reduced-system subspace is unstable and so the full system (which contains the reduced-system subspace) cannot be stable. Therefore the graphical test developed in Section 4.8.44.8.4 for the stability or instability of steady states in the reduced system is sufficient to determine the local stability or instability of corresponding steady states in the full system.

Note that when J is a contractive map, the steady state is locally exponentially stable because the real parts of all eigenvalues are negative, therefore analysis as in chapter 9 of (Khalil, 2002) ensures that sufficiently small perturbations to system behavior will not destroy the stability of the steady state.

Lastly, this proof only suffices to show local stability of the steady state in the full system. Other solutions representing different spatial patterns may exist and be stable in the full system even though the system reduction as performed herein would not identify them. Equivalently, the fact that the bistable steady state is global in the reduced system but only local in the full system exemplifies how introducing more compartments enlarges the space of possible solutions beyond those that exist in lower-dimensional regimes. Exploring the rich, complex range of behaviors for high-dimensional systems is a challenging and fascinating avenue for future research.

4.8.7 Channel length determination

To mimic close-range interaction of lateral inhibition with the diffusion-based system, a PDE model of AHL production, degradation, and diffusion is used to optimize l that allows sufficient diffusion of AHL to the immediate neighbors ([AHL] $\geq K_d$ at $\Delta x = l$) while not sufficient diffusion between nonadjacent compartments ([AHL] $< K_d$ at $\Delta x \geq 2l$). Let $\theta(t)$ to be AHL concentration over time at the center of an AHL-producing colony ($x = 0$). During the exponential growth ($t \leq 10h$), the concentration of AHL can be expressed using parameters defined in Table 4.3 as:

$$\begin{aligned}
\dot{\theta}(t) =& \nu I^* P_0 \exp(\mu t) - (\gamma + \frac{D}{l_{res}^2})\theta(t) \\
\xrightarrow{\mathcal{L}} s\Theta(s) =& \nu I^* P_0 \frac{1}{s-\mu} - (\gamma + \frac{D}{l_{res}^2})\Theta(s) \\
\Leftrightarrow \Theta(s) =& \nu I^* P_0 \frac{1}{(s-\mu)(s+\gamma+\frac{D}{l_{res}^2})} \\
\xrightarrow{\mathcal{L}^{-1}} \theta(t) = \nu I^* P_0 & \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} \left[\exp(\mu t) - \exp(-(\gamma + \frac{D}{l_{res}^2})t)\right]
\end{aligned} \tag{4.35}$$

where P_0 is the initial population size of the colony and l_{res} is the length of channels that connect compartments to reservoir (Section 4.8.14.8.2).

Now, consider diffusion of AHL across a channel. Define [AHL]$(x, t) =$

$\theta(x,t)$ in one-directional, infinite length of channel, with following boundary conditions:

$$\begin{cases} \theta(x,0) = 0, \forall x, \\ \theta(0,t) = \nu I^* P_0 \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} \left[\exp(\mu t) - \exp(-(\gamma + \frac{D}{l_{res}^2})t)\right], 0 \leq t \leq 10h, \\ \theta(\infty,t) = 0, \forall t. \end{cases} \tag{4.36}$$

Assume the degradation of AHL in the channel is negligible. Then, from Fick's second law,

$$\begin{aligned} \frac{\partial \theta}{\partial t} &= D\frac{\partial^2 \theta}{\partial x^2} \\ \xrightarrow{\mathcal{L}} s\Theta - 0 &= D\frac{d^2\Theta}{dx^2} \\ \Leftrightarrow \Theta(x,s) &= A\exp(\sqrt{\frac{s}{D}}x) + B\exp(-\sqrt{\frac{s}{D}}x) \end{aligned}$$

Using boundary conditions Eq. 4.36,

$$\Theta(x,s) = \nu I^* P_0 \frac{1}{(s-\mu)(s+\gamma+\frac{D}{l_{res}^2})} \exp(-\sqrt{\frac{s}{D}}x).$$

$$\Leftrightarrow \Theta(x,s) = \nu I^* P_0 \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} (\frac{\exp(-\sqrt{\frac{s}{D}}x)}{s-\mu} - \frac{\exp(-\sqrt{\frac{s}{D}}x)}{s+\gamma+\frac{D}{l_{res}^2}}). \tag{4.37}$$

Since there is no simple solution to inversely transform Eq. 4.37, instead, we will find an upper bound of AHL concentration and use that constant to be the boundary conditions.
Let $t \leq \tau_f$,

$$\begin{aligned} \theta(0,t) \leq & \theta(0,\tau_f) = \nu I^* P_0 \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} \left[\exp(\mu\tau_f) - \exp(-(\gamma+\frac{D}{l_{res}^2})\tau_f)\right] \\ \xrightarrow{\mathcal{L}} \Theta(0) \leq & \frac{\nu I^* P_0 \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} \left[\exp(\mu\tau_f) - \exp(-(\gamma+\frac{D}{l_{res}^2})\tau_f)\right]}{s} \end{aligned}$$

then we get

$$\xrightarrow{\mathcal{L}^{-1}} \theta(x,t) \leq \nu I^* P_0 \frac{1}{\mu+\gamma+\frac{D}{l_{res}^2}} \left[\exp(\mu\tau_f) - \exp(-(\gamma+\frac{D}{l_{res}^2})\tau_f)\right] erfc(\frac{x}{2\sqrt{Dt}}). \tag{4.38}$$

Using the upper-bound-concentration, we will determine the optimum channel length for short communication time ($T_{1/2}$ defined as time for a compartment to reach half concentration of the center) between the adjacent colonies while long the communication time between the non-adjacent compartments. $T_{1/2}$ at location x can be calculated as

$$\frac{x}{2\sqrt{DT_{1/2}}} = erfc^{-1}(0.5)$$

thus,

$$T_{1/2} = \frac{1}{4D}(\frac{x}{erfc^{-1}(0.5)})^2. \tag{4.39}$$

The objective function and the constraints can be set as follows:

$$\min_L T_{1/2}(x=l) + (\tau_f - T_{1/2}(x=2l)),$$

such that

$$\begin{aligned}
&\theta_{C_6}(L,\tau_f) \geq K_d^{C_6\to A},\\
&\theta_{C_6}(L,\tau_f) < K_d^{C_6\to B},\\
&\theta_{C_6}(2L,\tau_f) < K_d^{C_6\to A},\\
&\theta_{C_{12}}(L,\tau_f) \geq K_d^{C_{12}\to B},\\
&\theta_{C_{12}}(L,\tau_f) < K_d^{C_{12}\to A},\\
&\theta_{C_{12}}(2L,\tau_f) < K_d^{C_{12}\to B}
\end{aligned}$$

where the functions and parameters represent

[3OC6HSL]$(t) = \theta_{C_6}(x,t)$,
[3OC12HSL]$(t) = \theta_{C_{12}}(x,t)$,
$t = [0, \tau_f]$ is the time period when cells are in the exponential growth,
$K_d^{C_6 \to A}$ is the dissociation constant of 3OC6HSL on $plux$,
$K_d^{C_{12} \to A}$ is the cross-talk dissociation constant of 3OC12HSL on $plux$,
$K_d^{C_{12} \to B}$ is the dissociation constant of 3OC12HSL on $plas$,
$K_d^{C_6 \to B}$ is the cross-talk dissociation constant of 3OC6HSL on $plas$.

Using parameter values in Table 4.3, the optimum channel length is determined to be 4.5 mm $\leq l \leq 9$mm (Fig.4.13**a**).

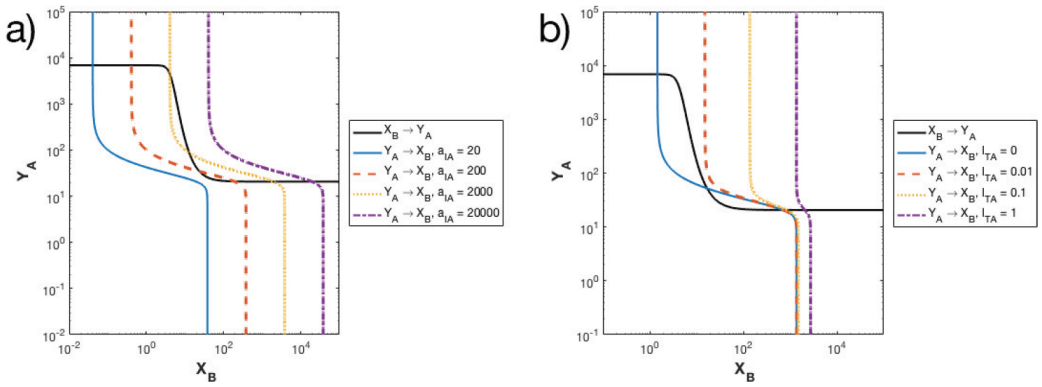

Figure 4.13: Kinetic rates determine whether the cross-repressive switches operate in the appropriate regime to turn each other on and off. Changes in relative output ranges between the two on/off switches for different values of **(a)** $a_{I_A} := \frac{V_{R_A} N C \epsilon_{I_A}}{\gamma_I \gamma_m}$ and **(b)** the leakiness of $plux$ (l_{RA}). Remaining parameters are as given in Table 4.2. The intersections between two transfer functions $X_B \to Y_A$ and $Y_A \to X_B$ indicate the steady states of the full system. As a_{IA} scales, so does the maximum and minimum output of LasI, which geometrically translates the composite transfer function $Y_A \to X_B$. Increasing the leakiness increases the minimum output of LasI, which decreases the dynamic range of the composite transfer function $Y_A \to X_B$. Leakiness reduces ultrasensitivity of $Y_A \to X_B$, which leads to loss of bistability. The translational movement of $Y_A \to X_B$ causes mismatched tuning between $X_B \to Y_A$ and $Y_A \to X_B$ also leads to loss of bistability.

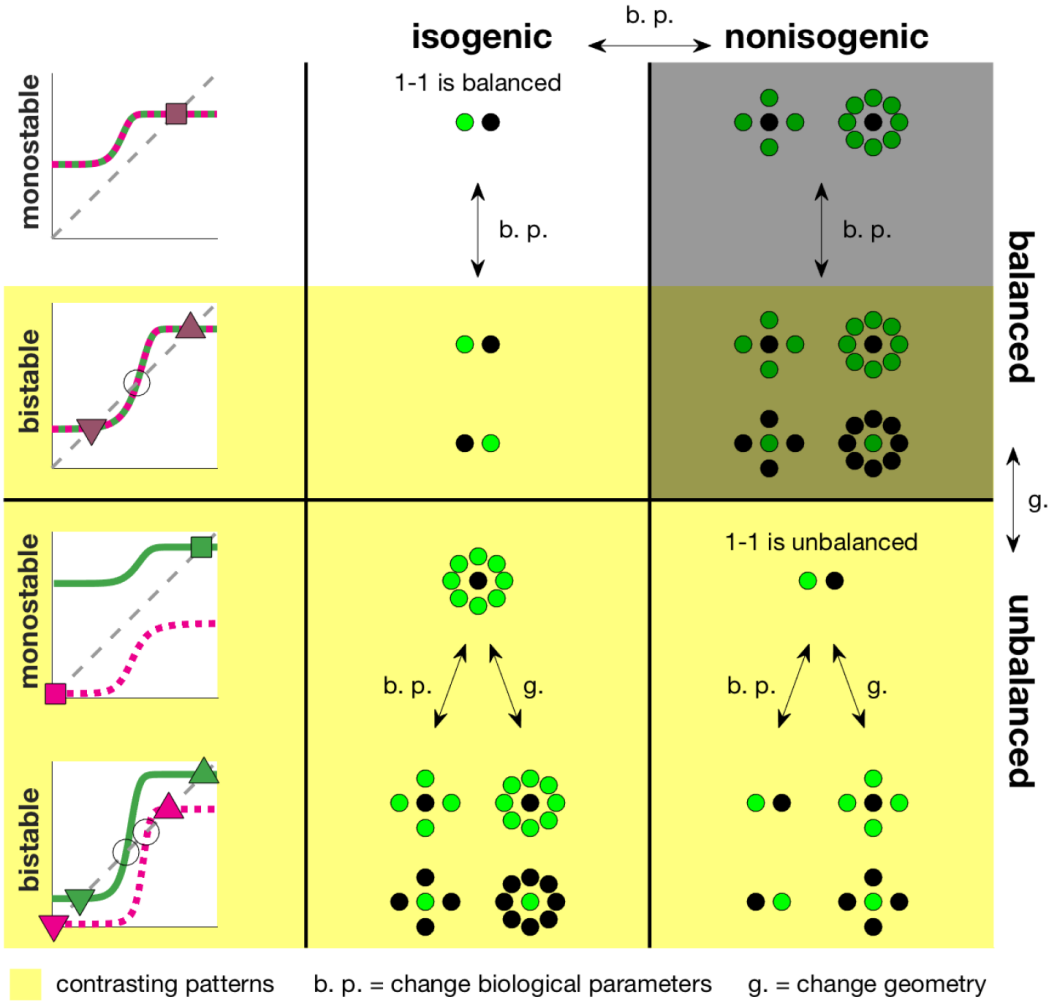

Figure 4.14: Schematic qualitatively identifying parameter changes that lead to bifurcation in two-strain circuits in isogenic (parameters identical between strains) and nonisogenic (parameters differ between strains) cases. Yellow background identifies systems that form contrasting patterns. In a balanced system, the cross-representative compartments exhibit identical transfer functions ($\boldsymbol{T_A}(\cdot) = \boldsymbol{T_B}(\cdot)$), while in an unbalanced system, the transfer functions differ. Monostable isogenic populations are homogeneous while nonisogenic populations exhibit contrast in the one-to-one geometry. Geometry may introduce imbalance or offset the biochemical difference, depending on the arrangement. Contrast in balanced systems is only attainable through bistability.

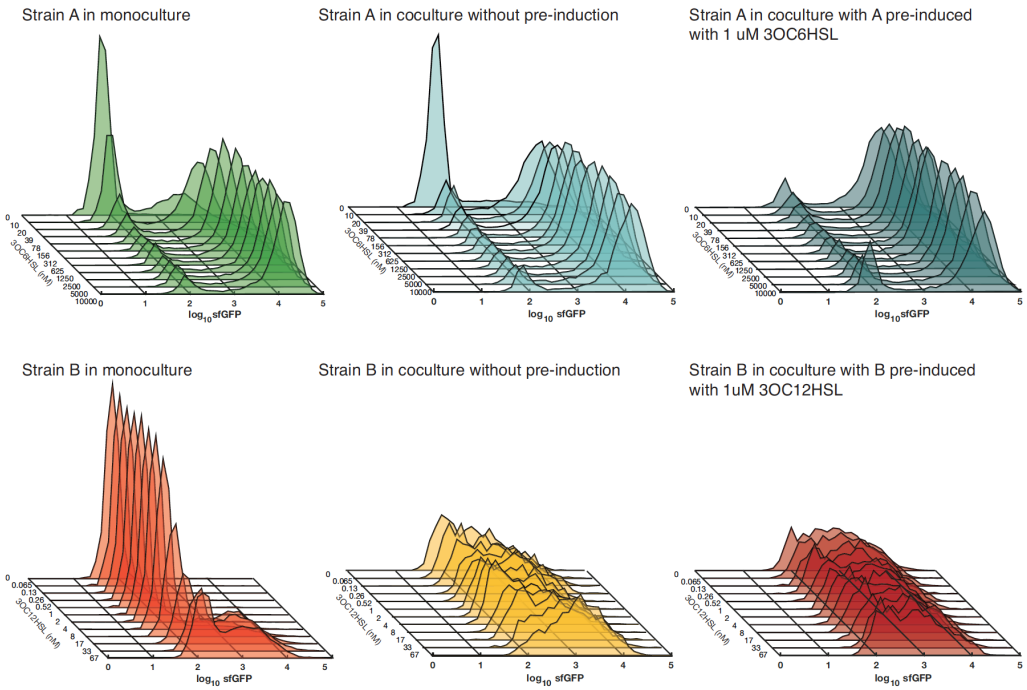

Figure 4.15: Single-cell measurements reveals variability in QS promoter activity at various subsaturating AHL concentrations. Each histogram in the stacked plot represents steady-state sfGFP (t = 8 hour) at indicated AHL concentration. The reporter fluorescence showed bimodal distribution in monocultures of Strains A and B at medium levels of AHL. As more external AHL was added, the population fraction with the high fluorescent mode increased. In coculture, most cells belonged to the high fluorescent mode fraction, implying higher AHL concentration in medium compared to the monocultures. This further supports the conclusion that cross-repression in coculture amplified the initial bias in the AHL concentration at steady state. Additionally, Strain B showed highly heterogeneous reporter expression even in a population fraction with the high fluorescence mode.

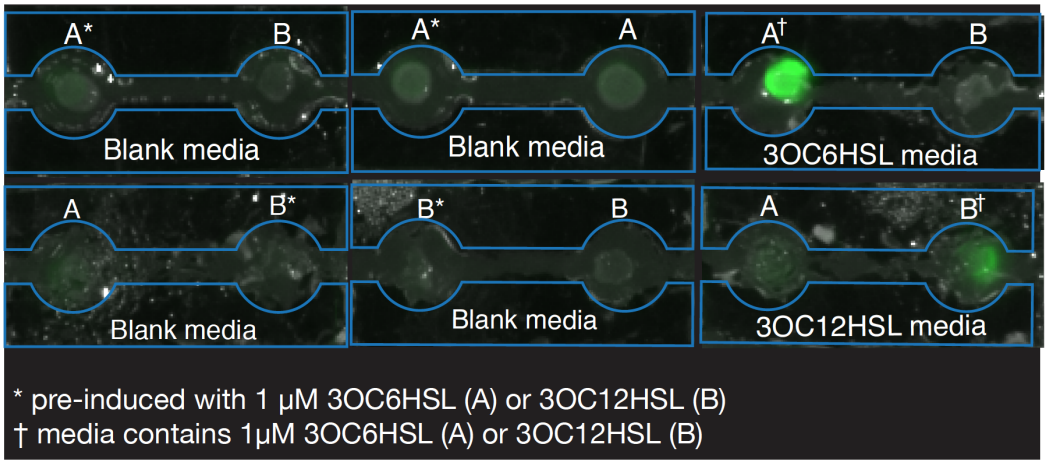

Figure 4.16: A plate fluorimeter image of the 1:1 spatial configuration devices with channel length of $9\ mm$**.** The image was taken after 12 hours of growth in room temperature. $*$indicates pre-induced strains with $1\ \mu M$ AHL and † indicates strains that were biased to be fluorescent by externally added AHL in medium. $9\ mm$ channel length was too long to establish communication between the adjacent compartments, which deviated from the expectation in Section 4.8.7. Positive controls were prepared by mixing either $1\ \mu M$ 3OC6HSL or 3OC12HSL in solid medium on the right-most devices in order to bias the gene expression pattern to be either Strain A fluorescent or Strain B fluorescent. The deviation is possibly caused by approximation of PDE solution using the upper bound AHL concentration.

Part III

Community Engineering

Chapter 5

Tuning gene expression in alpha-, beta-, and gamma-proteobacteria

5.1 Introduction

Billions of microbes inhabit nearly every microenvironment on earth ranging from geothermal vents deep in the ocean to human intestines. Through focused sequencing and functional studies from microbiologists and ecologists, these microbes appear to not only inhabit these spaces, but also exert effects on its local environment. Of particular focus are the significant positive and negative impact on human or agricultural health: for example, probiotic Bacillus have been found to suppress Staphylococcus aureus infection in the human gut (Piewngam et al., 2018), and diazotrophic microbiota have been found to support nutrient acquisition in maize roots (Van Deynze et al., 2018).

Programming synthetic biological circuits into the genome or plasmid of such microbes allows further control over the performance of these beneficial communities. However, intensive studies of genetic parts characterization and genetic engineering applications are, to-date, mostly limited to a handful of model organisms. There is increasing interest in engineering diverse microbes toward becoming optimized chasses that can persist and exert the desired activity in the target environment (Shepherd et al., 2018; Brophy et al., 2018). Genetic expression tools such as promoter libraries in diverse microbes should expand our engineering boundaries and lead to optimization of the synthetic circuit performance.

One key challenge in moving toward this goal is the tuning of gene expression levels. Biological circuit performance depends critically on optimal gene expression, as there is a trade-off between yield and stability of the heterologous gene expression inside a cell. Host processes and synthetic pathways inevitably compete for the same intracellular resource pool

(Carbonell-Ballestero et al., 2016). Moreover, the chemical (by)products from the synthetic pathways can be toxic and diminish host fitness. Hence, there is a need to use promoter libraries or inducible promoters to precisely control expression levels. Such gene expression tools have already been developed for specific species (S. Chen et al., 2007; Alagesan et al., 2018), classes of species (Kaczmarczyk et al., 2013), and in broad-hosts (Ruegg et al., 2018).

Here, we perform high-throughput characterization of promoter sequences in broad host species, to develop a registry of precise genetic expression tools as well as to evaluate the versatility of those tools using statistical modeling of functional promoters in each species.

5.2 Materials and Methods

5.2.1 Strains and growth

Agrobacterium tumefaciens C58 (At) was cultured in TSB with 20 $\mu g\ mL^{-1}$ of chloramphenicol, *Sinorhizobium meliloti* 1021 (Sm) was cultured in PYE, *Cupriavidus necator* H16 (Cn) was cultured in LB with 10 $\mu g\ mL^{-1}$ of gentamycin, *Paraburkholderia phytofirmans* PsJN (Bp) was cultured in LB with 15 $\mu g\ mL^{-1}$ of polymyxin B, *Escherichia coli* DH10β (Ec) was cultured in LB, *E. coli* conjugation strain WM3064 was cultured in LB supplemented with 300 μM diaminopimelic acid (DAP), and *Pseudomonas putida* KT2440 (Pp) was cultured in LB with 5 $\mu g\ mL^{-1}$ chloramphenicol. For selection, 50 $\mu g\ mL^{-1}$ kanamycin was added to the media for most bacteria except 1021, where 25 $\mu g\ mL^{-1}$ kanamycin was added to PYE. All bacteria were cultured in 30 ^{o}C.

5.2.2 Cloning

PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (Thermo) and oligonucleotides (IDT). Phosphorylated primers FW2915 *TATAATgtgtggaacaattcattaaagaggagaaaggtacc* and RV2902 *NNNNNNNNNNNNNNNNTGTCAAgttatgcaagagacctcgtggacatca* were used for $L1$, FW2916 *NNNNNNgtgtggaacaattcattaaagaggagaaaggtacc* and RV2902 were used for $L2$, and FW2915 and RV2904 *NNNNNNNNNNNNNNNNNNNNNNgttatgcaagagacctcgtggacatca* were used for $L3$ to amplify *pFAB5532* at annealing temperature of 63 ^{o}C. The resulting linear DNA were ligated in 16 ^{o}C overnight using T4 DNA Ligase (NEB). 50 ng of DNA products were electroporated to 0.5x DH10β (NEB)

in 1 mm gap cuvette at 1.8 kV, 200 Ω, and 25 μF. The typical time constant was 4 ms. For electroporation and conjugation of strains other than DH10β, extracted plasmids from DH10β were used for higher efficiency.

5.2.3 Electroporation and conjugation

At overnight culture was washed twice with 10 % glycerol and electroporation in a 1 mm gap cuvette was performed at 1.25 kV, 200 Ω, 25 μF. WM3064 was processed similarly except 1.8 kV was used for electroporation. Pp was washed with 300 mM sucrose and electroporation was performed at 1.25 kV, 200 Ω, and 25 μF. For Sm, Bp, and Rn, double-washed saturated cultures of donor cells were mixed with double-washed WM3064 cells at 4:1, 1:1, 1:4 ratios, respectively. The cocultures were grown on LB plates supplemented with DAP overnight in room temperature.

5.2.4 Fluorescence activated cell sorting

Saturated cultures were sorted according to gates in forward-scatter-area (FSC-A) and back-scatter-area (BSC-A) for size selection, forward-scatter-width (FSC-W) for singlet selection, back-scatter-width (BSC-W) for further singlet selection, and fluorescence-area (FL1-A) for "High", "Medium", "Low", and "Off" activity selections.

5.2.5 Barcode sequencing

We adapted the protocol of BarSeq in (Wetmore et al., 2015) for 20 μL PCR reactions with Q5 High-Fidelity DNA Polymerase and GC enhancer. PAGE-purified BarSeq_GFP_Rev *aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctnnnnngtgaaaagttcttctcctttactcat* was ordered through IDT and replaced P1 primer in the protocol. We multiplexed 96 samples in a lane of Illumina MiSeq.

5.3 Results

5.3.1 Promoter library design and construction

To characterize gene expression level in high throughput screening, Sort-Seq was performed for six ATCC strains of bacteria representing three different bacterial classes. Random promoter sequences were generated and

replaced *pTac* promoter in *pFAB5532* plasmid that drives reporter *sfGFP*. The plasmid has *pBBR1* origin (Kovach et al., 1995), *E. coli* RBS (Lutz et al., 1997), plasmid mobilizing gene *mob* (Simon et al., 1983), and Kanamycin resistance marker.

We aimed to create randomized $\sigma 70$ promoters. As the gene *rpoD* is highly conserved in regions 2 and 4 across all bacteria, where it binds to -35 and -10 consensus sequences of promoters, respectively (Lonetto et al., 1992), postulated that the recognition sequences in -35 and -10 of $\sigma 70$ promoters are also conserved to be $GGTACA$ and $TATAAT$. To dissect the effects of spacer, -35, and -10 sequences, we constructed three libraries with variability at the consensus regions. All of libraries contain random 17-base-pair spacers while the first library *L1* has conserved -35 and -10, the second library *L2* has conserved -35 and randomized –10, and the third library *L3* has randomized -35 and conserved -10 regions (Figure 5.1**A**).

Three libraries were either transformed by electroporation or conjugated through *WM3064* $\Delta dapA$::[erm pir] strain of *E. coli* to six species of bacteria, two each representing Alpha-, Beta-, and Gammaproteobacteria classes as shown in Figure 5.1**B**. Table 5.1 shows the transformation or conjugation efficiency of hosts. The transformation or conjugation processes were scaled until we obtain > 1 millions of colonies to ensure enough sequences can be sorted and analyzed.

5.3.2 Species-dependent variability in GFP expression

The single-cell analysis on 18 libraries showed species-dependent fluorescence profile (Figure 5.2). For most bacteria, cells harboring library with conserved -35 and -10 consensus sequence exhibited highest mean and median GFP expression, except *Cupriavidus necator* where library with random -35 sequence showed highest mean and median GFP expression. Further sequencing on these sorted libraries would clarify the variability in reporter gene expression for different bacteria.

5.3.3 Discussion

High-throughput screening of reporter fluorescence using flow cytometry enables characterization of species-specific expression patterns from three libraries that were constructed with combination of random and consensus promoter sequences. Subsequent high-throughput sequencing and statistical analysis would allow us to correlate promoter sequences and gene

A

Original pTac	T	T	G	A	C	A	a	t	t	a	a	t	c	a	t	c	c	g	g	c	t	c	g	T	A	T	A	A	T
Lib1	T	T	G	A	C	A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	A	T	A	A	T
Lib2	T	T	G	A	C	A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Lib3	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	A	T	A	A	T

B

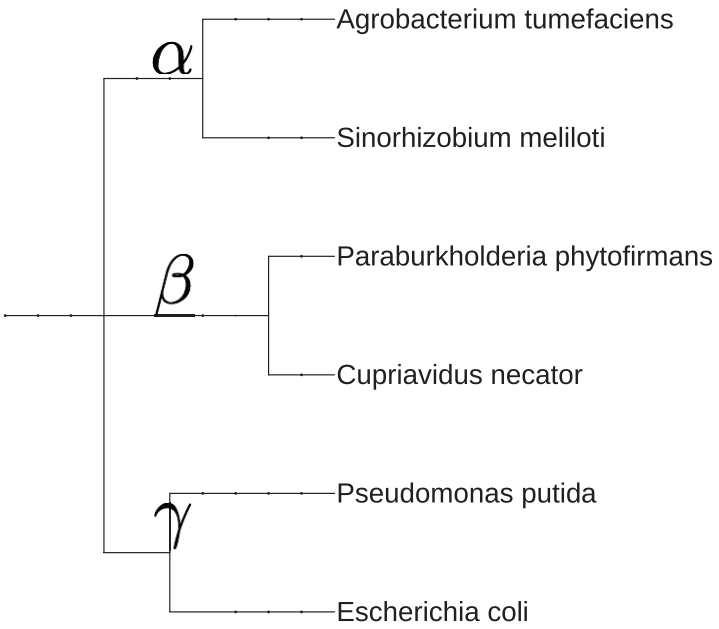

FIGURE 5.1: **Full scope of promoter library construction in six different bacterial species.** **A** Three types of DNA sequences are used to characterize σ70 promoters. The spacer, -35, and -10 of original *pTac* promoter of *E. coli* are replaced by random sequences to dissect the contribution of each region. **B** Phylogenetic tree of six host bacteria based on 16S ribosomal RNA sequence distance.

expression rates once the project is completed. The registry of σ70 constitutive promoters with variable expression rates will help optimizing the genetic circuit implementation in the six bacteria that are characterized in this study. The search for a universal promoter sequence across three classes of proteobacteria will be useful for implementing genetic circuits that work in broad host species.

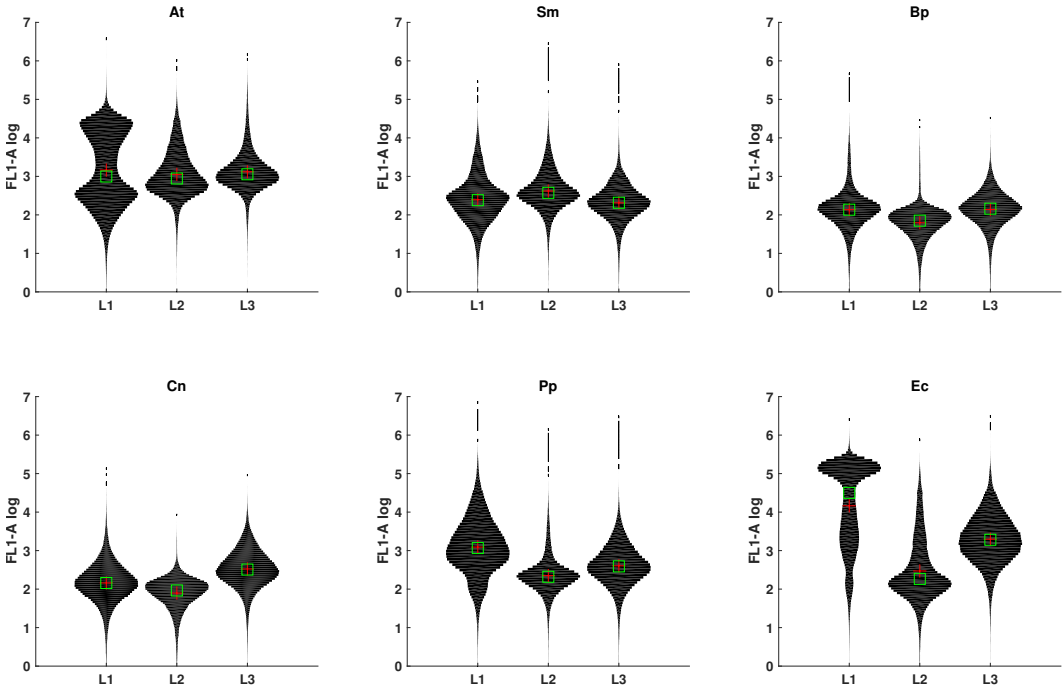

Figure 5.2: Density plots of flow cytometry measurements of library *L1, L2, L3* in six bacterial strains. Red indicates mean and green square indicates median of the distribution.

TABLE 5.1: Transformation and conjugation efficiency

Strain	Method	Efficiency
Agrobacterium tumefaciens C58	Electroporation	1e5 μg^{-1}
Sinorhizobium meliloti 1021	Conjugation	3.5e-4 cell^{-1}
Burkholderia phytofirmans PsJN	Conjugation	4e-2 cell^{-1}
Cupriavidus necator H16	Conjugation	2e-3 cell^{-1}
Pseudomonas putida KT2440	Electroporation	1.1e-5 μg^{-1}
Escherichia coli DH10β	Electroporation	1e7 μg^{-1}

5.4 Acknowledgments

I would like to thank Drs. Vivek Mutalik and Thomas L. Ruegg for their initial testing and advice on this work. I appreciate Eugene Pang for her diligent work on strain construction, FACS, and plasmid extraction. She contributed significantly to the experimental side of this project.

Bibliography

Afgan, Enis et al. (2018). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update". In: *Nucleic acids research* 46.W1, W537–W544.

Alagesan, Swathi, Erik K R Hanko, Naglis Malys, Muhammad Ehsaan, Klaus Winzer, and Nigel P Minton (2018). "Functional Genetic Elements for Controlling Gene Expression in Cupriavidus necator H16." In: *Applied and environmental microbiology* 84.19, e00878–18.

Amitai, Shahar, Yussuf Yassin, and Hanna Engelberg-Kulka (2004). "MazF-mediated cell death in Escherichia coli: a point of no return". In: *Journal of bacteriology* 186.24, pp. 8295–8300.

Amonlirdviman, Keith, Narmada A. Khare, David R. P. Tree, Wei-Shen Chen, Jeffrey D. Axelrod, and Claire J. Tomlin (2005). "Mathematical Modeling of Planar Cell Polarity to Understand Domineering Nonautonomy". In: *Science* 307.5708, pp. 423–426.

Andersen, Jens Bo, Claus Sternberg, Lars Kongsbak Poulsen, Sara Petersen Bjørn, Michael Givskov, and Søren Molin (1998). "New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria". In: *Applied and environmental microbiology (1998)* 64.6, pp. 2240–2246. arXiv: 98/{\$}04.00îĂś0 [0099-2240].

Anderson, Christopher J., Christopher A. Voigt, and Adam P. Arkin (2007). "Environmental signal integration by a modular and gate". In: *Molecular Systems Biology* 3.1, p. 133.

Anderson, J. Christopher, Elizabeth J. Clarke, Adam P. Arkin, and Christopher A. Voigt (2006). "Environmentally controlled invasion of cancer cells by engineered bacteria". In: *Journal of Molecular Biology* 355.4, pp. 619–627.

Andrews, Lauren B, Alec A.K. Nielsen, and Christopher A Voigt (2018). "Cellular checkpoint control using programmable sequential logic". In: *Science* 361.6408, eaap8987.

Angeli, David, James E Ferrell, and Eduardo D Sontag (2004a). "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems." In: *Proceedings of the National Academy of Sciences of the United States of America* 101.7, pp. 1822–7.

Angeli, David and Eduardo D. Sontag (2004b). "Multi-stability in monotone input/output systems". In: *Systems & Control Letters* 51.3-4, pp. 185–202.

Arcak, Murat (2013). "Pattern Formation by Lateral Inhibition in Large-Scale Networks of Cells". In: *IEEE TRANSACTIONS ON AUTOMATIC CONTROL* 58.5, pp. 1250–1262.

Arkin, Adam P et al. (2018). *KBase: The United States department of energy systems biology knowledgebase*.

Banerjee, Samiran, Klaus Schlaeppi, and Marcel G. A. van der Heijden (2018). "Keystone taxa as drivers of microbiome structure and functioning". In: *Nature Reviews Microbiology* 16, pp. 567–576.

Bassalo, Marcelo C, Andrew D Garst, Andrea L Halweg-Edwards, William C Grau, Dylan W Domaille, Vivek K Mutalik, Adam P Arkin, and Ryan T Gill (2016). "Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli". en. In: *ACS Synth. Biology* 5.7, pp. 561–568.

Basu, Subhayu, Yoram Gerchman, Cynthia H. Collins, Frances H. Arnold, and Ron Weiss (2005). "A synthetic multicellular system for programmed pattern formation". In: *Nature* 434.7037, pp. 1130–1134.

Bintu, Lacramioara, Nicolas E. Buchler, Hernan G. Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips (2005). "Transcriptional regulation by the numbers: Models". In: *Current Opinion in Genetics and Development* 15.2, pp. 116–124. arXiv: 0412011 [q-bio].

Boetius, Antje et al. (2000). "A marine microbial consortium apparently mediating anaerobic oxidation of methane". In: *Nature* 407.6804, pp. 623–626.

Bokinsky, Gregory et al. (2011). "Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.50, pp. 19949–54.

Brenner, K., D. K. Karig, R. Weiss, and F. H. Arnold (2007). "Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium". In: *Proceedings of the National Academy of Sciences* 104.44, pp. 17300–17304.

Brockman, Irene M and Kristala LJ Prather (2015). "Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites". In: *Metabolic engineering* 28, pp. 104–113.

Brophy, Jennifer A. N., Alexander J. Triassi, Bryn L. Adams, Rebecca L. Renberg, Dimitra N. Stratis-Cullum, Alan D. Grossman, and Christopher A. Voigt (2018). "Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria". In: *Nature Microbiology* 3.9, pp. 1043–1053.

Buchler, Nicolas E and Matthieu Louis (2008). "Molecular titration and ultrasensitivity in regulatory networks". In: *Journal of molecular biology* 384.5, pp. 1106–1119.

Burgard, Anthony P, Priti Pharkya, and Costas D Maranas (2003). "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization". en. In: *Biotechnol. Bioeng.* 84.6, pp. 647–657.

Cameron, D Ewen and James J Collins (2014). "Tunable protein degradation in bacteria". en. In: *Nature Biotechnology* 32.12, pp. 1276–1281.

Carbonell-Ballestero, M., E. Garcia-Ramallo, R. Montañez, C. Rodriguez-Caso, and J. Macía (2016). "Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the Ohm's law". In: *Nucleic Acids Research* 44.1, pp. 496–507.

Cardinale, Stefano, Marcin Pawel Joachimiak, and Adam Paul Arkin (2013). "Effects of genetic variation on the e. coli host-circuit interface". In: *Cell Reports* 4.2, pp. 231–237.

Ceroni, Francesca et al. (2018). "Burden-driven feedback control of gene expression". In: *Nature Methods* 15.5, pp. 387–393.

Chan, Clement T Y, Jeong Wook Lee, D Ewen Cameron, Caleb J Bashor, and James J Collins (2016). "'Deadman' and 'Passcode' microbial kill switches for bacterial containment". In: *Nature Chemical Biology* 12.2, pp. 82–86.

Chang, C. Y., K. Nam, and R. Young (1995). "S gene expression and the timing of lysis by bacteriophage". In: *Journal of Bacteriology* 177.11, pp. 3283–3294.

Chappell, James, Alexandra Westbrook, Matthew Verosloff, and Julius B. Lucks (2017). "Computational design of small transcription activating RNAs for versatile and dynamic gene regulation". In: *Nature Communications* 8.1, p. 1051.

Chen, David (2014). "Designing Genetic Circuits for Memory and Communication". PhD thesis. University of California, Berkeley.

Chen, Huiyi, Katsuyuki Shiroguchi, Hao Ge, and Xiaoliang Sunney Xie (2015). "Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli." In: *Molecular systems biology* 11.5, p. 808.

Chen, Shicheng, Michael Bagdasarian, Michael G Kaufman, Adam K Bates, and Edward D Walker (2007). "Mutational analysis of the ompA promoter from Flavobacterium johnsoniae." In: *Journal of bacteriology* 189.14, pp. 5108–18.

Chen, Po-Wei, Matthew K Theisen, and James C Liao (2017). "Metabolic systems modeling for cell factories improvement". In: *Current Opinion in Biotechnology* 46, pp. 114–119.

Chen, Ye, Jae Kyoung Kim, Andrew J. Hirning, Krešimir Josić, and Matthew R. Bennett (2015). "Emergent genetic oscillations in a synthetic microbial consortium". In: *Science* 349.6251, pp. 986–989. arXiv: arXiv:1011.1669v3.

Chevalier, Michael, Mariana Gomez-Schiavon, Andrew Ng, and Hana El-Samad (2018). "Design and analysis of a Proportional-Integral-Derivative controller with biological molecules". In: *bioRxiv*, p. 303545.

Christensen-Dalsgaard, Mikkel and Kenn Gerdes (2008). "Translation affects YoeB and MazF messenger RNA interferase activities by different mechanisms". In: *Nucleic acids research* 36.20, pp. 6472–6481.

Collier, Joanne R., Nicholas A. M. Monk, Phillip K. Maini, and Julian H. Lewis (1996). "Pattern Formation by Lateral Inhibition with Feedback : a Mathematical Model of Delta – Notch Intercellular Signalling". In: *Journal of Theoretical Biology* 183.4, pp. 429–446.

Cunningham, Alfred B., John E. Lennox, and Rockford J. Ross (2011). "Biofilms: The Hypertextbook". In: *Biofilms: The Hypertextbook*. Version 4. Chap. Section 10.

Dahl, Robert H. et al. (2013). "Engineering dynamic pathway regulation using stress-response promoters". In: *Nature Biotechnology* 31.11, pp. 1039–1046.

Dalbow, D G and R Young (1975). "Synthesis time of beta-galactosidase in Escherichia coli B/r as a function of growth rate." In: *The Biochemical journal* 150.1, pp. 13–20.

De Vos, Dirk, Frank J Bruggeman, Hans V Westerhoff, and Barbara M Bakker (2011). "How molecular competition influences fluxes in gene expression networks". In: *PloS one* 6.12, e28494.

Del Vecchio, Domitilla (2015). "Modularity, context-dependence, and insulation in engineered biological circuits". In: *Trends in Biotechnology* 33.2, pp. 111–119.

Deutscher, Murray P (2006). "Degradation of RNA in bacteria: comparison of mRNA and stable RNA". In: *Nucleic acids research* 34.2, pp. 659–666.

Dong, Hengjiang, Lars Nilsson, and Charles G Kurland (1995). "Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction." In: *Journal of bacteriology* 177.6, pp. 1497–1504.

Ellis, Tom, Xiao Wang, and James J. Collins (2009). "Diversity-based, model-guided construction of synthetic gene networks with predicted functions". In: *Nature Biotechnology* 27.5, pp. 465–471.

Elowitz, Michael B. and Stanislas Leibler (2000). "A synthetic oscillatory network of transcriptional regulators". In: *Nature* 403.6767, pp. 335–338.

Eraso, Jesus M, Lye M Markillie, Hugh D Mitchell, Ronald C Taylor, Galya Orr, and William Margolin (2014). "The highly conserved MraZ protein is a transcriptional regulator in Escherichia coli". In: *Journal of bacteriology*, JB–01370.

Federle, Michael J. and Bonnie L. Bassler (2003). "Interspecies communication in bacteria". In: *Journal of Clinical Investigation* 112.9, pp. 1291–1299. arXiv: Federle,MichaelJ,2003,Interspecies.

Fernandez-Rodriguez, Jesus, Felix Moser, Miryoung Song, and Christopher A. Voigt (2017). "Engineering RGB color vision into Escherichia coli". In: *Nature Chemical Biology* 13.7, pp. 706–708.

Ferrell Jr, James E and Sang Hoon Ha (2014). "Ultrasensitivity part III: cascades, bistable switches, and oscillators". In: *Trends in biochemical sciences* 39.12, pp. 612–618.

Fischbach, Michael A. (2018). *Microbiome: Focus on Causation and Mechanism*.

Flynn, Julia M, Igor Levchenko, Meredith Seidel, Sue H Wickner, Robert T Sauer, and Tania A Baker (2001). "Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis". In: *Proceedings of the National Academy of Sciences* 98.19, pp. 10584–10589.

Garcia, Erin C., Andrew I. Perault, Sara A. Marlatt, and Peggy A. Cotter (2016). "Interbacterial signaling via Burkholderia contact-dependent growth inhibition system proteins". In: *Proceedings of the National Academy of Sciences* 113.29, pp. 8296–8301.

Garcia, Hernan G and Rob Phillips (2011). "Quantitative dissection of the simple repression input-output function". In: *Proceedings of the National Academy of Sciences* 108.29, pp. 12173–12178.

Gardner, Timothy S., Charles R. Cantor, and James J. Collins (2000). "Construction of a genetic toggle switch in Escherichia coli". In: *Nature* 403.6767, pp. 339–342. arXiv: NIHMS150003.

Germain, E, M Roghanian, K Gerdes, et al. (2015). "Stochastic induction of persister cells by HipA through (p) ppGpp-mediated activation of mRNA endonucleases". In: *Proceedings of the*.

Gertz, Jason and Barak A Cohen (2009). "Environment-specific combinatorial cis-regulation in synthetic promoters". In: *Molecular systems biology* 5.1, p. 244.

Gil, Rosario, Francisco J Silva, Juli Peretó, and Andrés Moya (2004). "Determination of the core of a minimal bacterial gene set". In: *Microbiology and Molecular Biology Reviews* 68.3, pp. 518–537.

Giuliodori, Anna Maria, Anna Brandi, Mara Giangrossi, Claudio O Gualerzi, and Cynthia L Pon (2007). "Cold-stress-induced de novo expression of infC and role of IF3 in cold-shock translational bias". In: *Rna*.

González Fernández-Niño, Susana M, A Michelle Smith-Moritz, Leanne Jade G Chan, Paul D Adams, Joshua L Heazlewood, and Christopher J Petzold (2015). "Standard flow liquid chromatography for shotgun proteomics in bioenergy research". In: *Frontiers in bioengineering and biotechnology* 3, p. 44.

Gorochowski, Thomas E, Amin Espah Borujeni, Yongjin Park, Alec Ak Nielsen, Jing Zhang, Bryan S Der, D Benjamin Gordon, and Christopher A Voigt (2017). "Genetic circuit characterization and debugging using RNA-seq." In: *Molecular systems biology* 13.11, p. 952.

Gotta, S L, O L Miller, S L French, and S L French (1991). "rRNA transcription rate in Escherichia coli." In: *Journal of Bacteriology* 173.20, pp. 6647–9.

Grant, P. K., N. Dalchau, J. R. Brown, F. Federici, T. J. Rudge, B. Yordanov, O. Patange, A. Phillips, and J. Haseloff (2016). "Orthogonal intercellular signaling for programmed spatial behavior". In: *Molecular Systems Biology* 12.1, pp. 849–849.

Greenwald, Iva and Gerald M. Rubin (1992). "Making a difference: The role of cell-cell interactions in establishing separate identities for equivalent cells". In: *Cell* 68.2, pp. 271–281.

Gualerzi, Claudio O (2011). *Ribosomes: Structure, Function, Dynamics*. Spring Science and Business Media.

Gualerzi, Claudio O, Anna Maria Giuliodori, and Cynthia L Pon (2003). "Transcriptional and post-transcriptional control of cold-shock genes". In: *Journal of molecular biology* 331.3, pp. 527–539.

Guisoni, Nara, Rosa Martinez-Corral, Jordi Garcia-Ojalvo, and Joaquín de Navascués (2017). "Diversity of fate outcomes in cell pairs under lateral

inhibition". In: *Development* 144.7, pp. 1177–1186. arXiv: arXiv:1011.1669v3.

Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather (2017). "Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit". In: *Nature Biotechnology* 35.3, pp. 273–279.

Gyorgy, Andras, José I Jiménez, John Yazbek, Hsin-Ho Huang, Hattie Chung, Ron Weiss, and Domitilla Del Vecchio (2015). "Isocost lines describe the cellular economy of genetic circuits". In: *Biophysical journal* 109.3, pp. 639–646.

Hale, Cynthia A, Hans Meinhardt, and Piet AJ De Boer (2001). "Dynamic localization cycle of the cell division regulator MinE in Escherichia coli". In: *The EMBO journal* 20.7, pp. 1563–1572.

Halleran, Andrew D and Richard M Murray (2018). "Cell-Free and in Vivo Characterization of Lux, Las, and Rpa Quorum Activation Systems in E. coli". In: *ACS Synthetic Biology* 7.2, pp. 752–755. arXiv: /doi.org/10.1101/159988 [https:].

Hays, Stephanie G, Leo L.W. Yan, Pamela A Silver, and Daniel C Ducat (2017). "Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction". In: *Journal of Biological Engineering* 11.1, p. 4.

Herrera Paredes, Sur et al. (2018). "Design of synthetic bacterial communities for predictable plant phenotypes". In: *PLoS Biology* 16.2, e2003962.

Hooshangi, S., S. Thiberge, and R. Weiss (2005). "Ultrasensitivity and noise propagation in a synthetic transcriptional cascade". In: *Proceedings of the National Academy of Sciences* 102.10, pp. 3581–3586.

Hu, Chelsea Y, Melissa K Takahashi, Yan Zhang, and Julius B Lucks (2018). "Engineering a Functional Small RNA Negative Autoregulation Network with Model-Guided Design". In: *ACS Synthetic Biology* 7.6, pp. 1507–1518.

Hwang, In Young, Hui Ling Lee, James Guoxian Huang, Yvonne Yijuan Lim, Wen Shan Yew, Yung Seng Lee, and Matthew Wook Chang (2018). "Engineering microbes for targeted strikes against human pathogens". In: *Cellular and Molecular Life Sciences* 75.15, pp. 2719–2733.

Ishida, Yojiro, Jung-Ho Park, Lili Mao, Yoshihiro Yamaguchi, and Masayori Inouye (2013). "Replacement of all arginine residues with canavanine in MazF-bs mRNA interferase changes its specificity". en. In: *J. Biological Chemistry.* 288.11, pp. 7564–7571.

Jaroentomeechai, Thapakorn, Jessica C. Stark, Aravind Natarajan, Cameron J. Glasscock, Laura E. Yates, Karen J. Hsu, Milan Mrksich, Michael C. Jewett, and Matthew P. DeLisa (2018). "Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery". In: *Nature Communications* 9.1, p. 2686.

Kaczmarczyk, Andreas, Julia A. Vorholt, and Anne Francez-Charlot (2013). "Cumate-Inducible Gene Expression System for Sphingomonads and Other Alphaproteobacteria". In: *Appl. Environ. Microbiol.* 79.21, pp. 6795–6802.

Kamada, Katsuhiko, Fumio Hanaoka, and Stephen K Burley (2003). "Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition". In: *Molecular cell* 11.4, pp. 875–884.

Kaufmann, G. F. et al. (2005). "Revisiting quorum sensing: Discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones". In: *Proceedings of the National Academy of Sciences* 102.2, pp. 309–314.

Keiler, Kenneth C (2015). "Mechanisms of ribosome rescue in bacteria". In: *Nature Reviews Microbiology* 13.5, p. 285.

Khalil, Hassan K. (2002). *Nonlinear systems*. Prentice Hall, p. 750. ISBN: 0130673897.

Kim, H. J., J. Q. Boedicker, J. W. Choi, and R. F. Ismagilov (2008). "Defined spatial structure stabilizes a synthetic multispecies bacterial community". In: *Proceedings of the National Academy of Sciences* 105.47, pp. 18188–18193. arXiv: arXiv:1408.1149.

Kim, Jun-Seob et al. (2016). "Fumarate-mediated persistence of Escherichia coli against antibiotics". In: *Antimicrobial agents and chemotherapy*, AAC–01794.

Klavins, Eric (2014). "Lightening the load in synthetic biology". In: *Nature Biotechnology* 32.12, pp. 1198–1200.

Klumpp, Stefan, Zhongge Zhang, and Terence Hwa (2009). "Growth Rate-Dependent Global Effects on Gene Expression in Bacteria". In: *Cell* 139.7, pp. 1366–1375.

Kong, Wentao, Andrew E. Blanchard, Chen Liao, and Ting Lu (2017). "Engineering robust and tunable spatial structures with synthetic gene circuits". In: *Nucleic Acids Research* 45.2, pp. 1005–1014. arXiv: 1611.06654.

Kong, Wentao, David R. Meldgin, James J. Collins, and Ting Lu (2018). "Designing microbial consortia with defined social interactions". In: *Nature Chemical Biology*, p. 1.

Kovach, Michael E., Philip H. Elzer, D. Steven Hill, Gregory T. Robertson, Michael A. Farris, R.Martin Roop, and Kenneth M. Peterson (1995). "Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes". In: *Gene* 166.1, pp. 175–176.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". In: *Genome biology* 10.3, R25.

Lee, Sungrim Seirin (2016). "Lateral inhibition-induced pattern formation controlled by the size and geometry of the cell". In: *Journal of Theoretical Biology* 404, pp. 51–65.

Lee, Taek, Rachel A Krupa, Fuzhong Zhang, Meghdad Hajimorad, William J Holtz, Nilu Prasad, Sung Lee, and Jay D Keasling (2011). "BglBrick vectors and datasheets: A synthetic biology platform for gene expression". In: *Journal of Biological Engineering* 5.1, p. 12.

Levine, Joe H. and Michael B. Elowitz (2014). "Polyphasic feedback enables tunable cellular timers". In: *Current Biology* 24.20, R994–R995.

Li, Guang-Yao, Yonglong Zhang, Mitchell CY Chan, Tapas K Mal, Klaus P Hoeflich, Masayori Inouye, and Mitsuhiko Ikura (2006). "Characterization of dual substrate binding sites in the homodimeric structure of Escherichia coli mRNA interferase MazF". In: *Journal of molecular biology* 357.1, pp. 139–150.

Li, Yanran, Sijin Li, Kate Thodey, Isis Trenchard, Aaron Cravens, and Christina D Smolke (2018). "Complete biosynthesis of noscapine and halogenated alkaloids in yeast." In: *Proceedings of the National Academy of Sciences of the United States of America* 115.17, E3922–E3931.

Lim, Wendell A, Connie M Lee, and Chao Tang (2013). "Design principles of regulatory networks: searching for the molecular algorithms of the cell." In: *Molecular cell* 49.2, pp. 202–12.

Lim, Yan Ping, Maybelle K Go, Manfred Raida, Takao Inoue, Markus R Wenk, Jay D Keasling, Matthew W Chang, and Wen Shan Yew (2018). "Synthetic Enzymology and the Fountain of Youth: Repurposing Biology for Longevity". In: *ACS Omega* 3.9, pp. 11050–11061.

Lonetto, M, M Gribskov, and C A Gross (1992). "The σ70 family: Sequence conservation and evolutionary relationships". In: *Journal of Bacteriology* 174.12, pp. 3843–3849.

Lutz, Rolf and Hermann Bujard (1997). "Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O

and AraC/I1-I2regulatory elements". In: *Nucleic Acids Research* 25.6, pp. 1203–1210.

Ma, Wenzhe, Ala Trusina, Hana El-Samad, Wendell A. Lim, and Chao Tang (2009). "Defining Network Topologies that Can Achieve Biochemical Adaptation". In: *Cell* 138.4, pp. 760–773.

Mangan, S. and U. Alon (2003). "Structure and function of the feed-forward loop network motif". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.21, pp. 11980–11985. arXiv: 0711.0685.

Marcos, Enrique et al. (2018). "De novo design of a non-local β-sheet protein with high stability and accuracy". In: *Nature Structural & Molecular Biology* 25.11, pp. 1028–1034.

Matsuda, Mitsuhiro, Makito Koga, Knut Woltjen, Eisuke Nishida, and Miki Ebisuya (2015). "Synthetic lateral inhibition governs cell-type bifurcation with robust ratios". In: *Nature Communications* 6, pp. 1–12.

Mengin-Lecreulx, Dominique, Juan Ayala, Ahmed Bouhss, Jean van Heijenoort, Claudine Parquet, and Hiroshi Hara (1998). "Contribution of the P mra promoter to expression of genes in the Escherichia coli mra cluster of cell envelope biosynthesis and cell division genes". In: *Journal of bacteriology* 180.17, pp. 4406–4412.

Mishra, Deepak, Phillip M Rivera, Allen Lin, Domitilla Del Vecchio, and Ron Weiss (2014). "A load driver device for engineering modularity in biological networks". In: *Nature Biotechnology* 32.12, pp. 1268–1275.

Mok, Wendy WK, Junyoung O Park, Joshua D Rabinowitz, and Mark P Brynildsen (2015). "RNA futile cycling in model persisters derived from MazF accumulation". In: *MBio* 6.6, e01588–15.

Momose, Yoshika, Kazuhiro Hirayama, and Kikuji Itoh (2008). "Competition for proline between indigenous Escherichia coli and E. coli O157:H7 in gnotobiotic mice associated with infant intestinal microbiota and its contribution to the colonization resistance against E. coli O157:H7". In: *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* 94.2, pp. 165–171.

Morse, R. P., K. C. Nikolakakis, J. L. E. Willett, E. Gerrick, D. A. Low, C. S. Hayes, and C. W. Goulding (2012). "Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems". In: *Proceedings of the National Academy of Sciences* 109.52, pp. 21480–21485.

Mousa, Walaa K., Charles Shearer, Victor Limay-Rios, Cassie L. Ettinger, Jonathan A. Eisen, and Manish N. Raizada (2016). "Root-hair endophyte stacking in finger millet creates a physicochemical barrier to trap the

fungal pathogen Fusarium graminearum". In: *Nature Microbiology* 1.12, pp. 1–12.

Mutalik, Vivek K. et al. (2013). "Precise and reliable gene expression via standard transcription and translation initiation elements". In: *Nature Methods* 10.4, pp. 354–360.

Nevozhay, Dmitry, Rhys M Adams, Kevin F Murphy, Krešimir Josić, and Gábor Balázsi (2009). "Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression". In: *Proceedings of the National Academy of Sciences* 106.13, pp. 5123–5128.

Nielsen, Alec A K, Bryan S Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A Strychalski, David Ross, Douglas Densmore, and Christopher A Voigt (2016). "Genetic circuit design automation." In: *Science* 352.6281, p. 53.

Nielsen, Jens and Jay D. Keasling (2016). "Engineering Cellular Metabolism". In: *Cell* 164.6, pp. 1185–1197.

Nuñez, Isaac N., Tamara F Matute, Ilenne D Del Valle, Anton Kan, Atri Choksi, Drew Endy, Jim Haseloff, Timothy J Rudge, and Fernan Federici (2017). "Artificial Symmetry-Breaking for Morphogenetic Engineering Bacterial Colonies". In: *ACS Synthetic Biology* 6.2, pp. 256–265.

Olson, Evan J, Constantine N Tzouanas, and Jeffrey J Tabor (2017). "A photoconversion model for full spectral programming and multiplexing of optogenetic systems." In: *Molecular systems biology* 13.4, p. 926.

Oppenheim, Amos B., Oren Kobiler, Joel Stavans, Donald L. Court, and Sankar Adhya (2005). "Switches in Bacteriophage Lambda Development". In: *Annual Review of Genetics* 39.1, pp. 409–429.

Ostrov, Nili, Miguel Jimenez, Sonja Billerbeck, James Brisbois, Joseph Matragrano, Alastair Ager, and Virginia W. Cornish (2017). "A modular yeast biosensor for low-cost point-of-care pathogen detection". In: *Science Advances* 3.6, e1603221.

Pai, Anand and Lingchong You (2009). "Optimal tuning of bacterial sensing potential". In: *Mol. Systems Biology* 5.286.

Park, J H, Y Yamaguchi, and M Inouye (2011). "Bacillus subtilis MazF?bs (EndoA) is a UACAU?specific mRNA interferase". In: *FEBS Lett.*

Peters, Jason M et al. (2016). "A comprehensive, CRISPR-based functional analysis of essential genes in bacteria". In: *Cell* 165.6, pp. 1493–1506.

Pharkya, Priti, Anthony P Burgard, and Costas D Maranas (2004). "OptStrain: a computational framework for redesign of microbial production systems". en. In: *Genome Res.* 14.11, pp. 2367–2376.

Phillips, Rob (2017). "Musings on mechanism: Quest for a quark theory of proteins?" In: *FASEB Journal* 31.10, pp. 4207–4215.

Piewngam, Pipat et al. (2018). "Pathogen elimination by probiotic Bacillus via signalling interference". In: *Nature* 562, pp. 532–537.

Poszytek, Krzysztof, Martyna Ciżkowska, Aleksandra Skłodowska, and Łukasz Drewniak (2016). "Microbial Consortium with High Cellulolytic Activity (MCHCA) for Enhanced Biogas Production". In: *Frontiers in Microbiology* 7, p. 324.

Prindle, Arthur, Phillip Samayoa, Ivan Razinkov, Tal Danino, Lev S Tsimring, and Jeff Hasty (2012). "A sensing array of radically coupled genetic 'biopixels'". In: *Nature* 481, pp. 39–44.

Qi, Lei S, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim (2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". In: *Cell* 152.5, pp. 1173–1183.

Rebbapragada, A, M S Johnson, et al. (1997). "The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for Escherichia ?" In: *Proceedings of the*.

Rhodius, Virgil A. et al. (2013). "Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters". In: *Molecular Systems Biology* 9.702, pp. 1–13.

Richards, Jamie, Preeti Mehta, and A Wali Karzai (2006). "RNase R degrades non-stop mRNAs selectively in an SmpB-tmRNA-dependent manner". In: *Molecular microbiology* 62.6, pp. 1700–1712.

Ro, Dae Kyun et al. (2006). "Production of the antimalarial drug precursor artemisinic acid in engineered yeast". In: *Nature* 440.7086, pp. 940–943.

Ruegg, Thomas L. et al. (2018). "Author Correction: Jungle Express is a versatile repressor system for tight transcriptional control". In: *Nature Communications* 9.1, p. 4596.

Rufino Ferreira, Ana S. and Murat Arcak (2013). "A Graph Partitioning Approach to Predicting Patterns in Lateral Inhibition Systems". In: *SIAM Journal on Applied Dynamical Systems* 12.4, pp. 2012–2031.

Rufino Ferreira, Ana S., Justin Hsia, and Murat Arcak (2015). "A Compartmental Lateral Inhibition System to Generate Contrasting Patterns". In: *IEEE Life Sciences Letters* 1.1, pp. 7–10.

Salis, Howard M., Ethan A. Mirsky, and Christopher A. Voigt (2009). "Automated design of synthetic ribosome binding sites to control protein expression". In: *Nature Biotechnology* 27.10, pp. 946–950.

Samoilov, Michael S., Gavin Price, and Adam P. Arkin (2006). "From fluctuations to phenotypes: the physiology of noise." In: *Science's STKE : signal transduction knowledge environment* 2006.366, pp. 1–10.

Sauert, Martina, Michael T Wolfinger, Oliver Vesper, Christian Müller, Konstantin Byrgazov, and Isabella Moll (2016). "The MazF-regulon: a toolbox for the post-transcriptional stress response in Escherichia coli". en. In: *Nucleic Acids Research* 44.14, pp. 6660–6675.

Scholz, Oliver, Peter Schubert, Martin Kintrup, and Wolfgang Hillen (2000). "Tet Repressor Induction without Mg2+". In: *Biochemistry* 39.35, pp. 10914–10920.

Scott, M., C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa (2010). "Interdependence of Cell Growth and Gene Expression: Origins and Consequences". In: *Science* 330.6007, pp. 1099–1102.

Scott, Spencer R. and Jeff Hasty (2016). "Quorum Sensing Communication Modules for Microbial Consortia". In: *ACS Synthetic Biology* 5.9, pp. 969–977. arXiv: 15334406.

Shachrai, Irit, Alon Zaslaver, Uri Alon, and Erez Dekel (2010). "Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth". In: *Molecular cell* 38.5, pp. 758–767.

Shahrezaei, Vahid and Samuel Marguerat (2015). "Connecting growth with gene expression: of noise and numbers". In: *Current Opinion in Microbiology* 25, pp. 127–135.

Shepherd, Elizabeth Stanley, William C. DeLoache, Kali M. Pruss, Weston R. Whitaker, and Justin L. Sonnenburg (2018). "An exclusive metabolic niche enables strain engraftment in the gut microbiota". In: *Nature* 557.7705, pp. 434–438.

Simon, R., U. Priefer, and A. Pühler (1983). "A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria". In: *Nature Biotechnology* 1.9, pp. 784–791.

Solomon, Kevin V, Tae Seok Moon, Brian Ma, Tarielle M Sanders, and Kristala LJ Prather (2012). "Tuning primary metabolism for heterologous pathway productivity". In: *ACS synthetic biology* 2.3, pp. 126–135.

Stanton, Brynne C., Alec A K Nielsen, Alvin Tamsir, Kevin Clancy, Todd Peterson, and Christopher A. Voigt (2014). "Genomic mining of prokaryotic repressors for orthogonal logic gates." In: *Nature chemical biology* 10.2, pp. 99–105. arXiv: NIHMS150003.

Tamsir, Alvin, Jeffrey J. Tabor, and Christopher A. Voigt (2011). "Robust multicellular computing using genetically encoded NOR gates

and chemical 'wires'". In: *Nature* 469.7329, pp. 212–215. arXiv: NIHMS150003.

Tayar, Alexandra M, Eyal Karzbrun, Vincent Noireaux, and Roy H Bar-Ziv (2017). "Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 114.44, pp. 11609–11614.

Thompson, Luke R. et al. (2017). "A communal catalogue reveals Earth's multiscale microbial diversity". In: *Nature* 551.7681, pp. 457–463.

Tiwari, Abhinav and Oleg A Igoshin (2012). "Coupling between feedback loops in autoregulatory networks affects bistability range, open-loop gain and switching times". In: *Physical biology* 9.5, p. 055003.

Toda, Satoshi, Lucas R Blauch, Sindy K Y Tang, Leonardo Morsut, and Wendell A Lim (2018). "Programming self-organizing multicellular structures with synthetic cell-cell signaling." In: *Science* 361.6398, pp. 156–162.

Tripathi, Arti, Pooja C Dewan, Shahbaz A Siddique, and Raghavan Varadarajan (2014). "MazF-induced growth inhibition and persister generation in Escherichia coli". In: *Journal of Biological Chemistry* 289.7, pp. 4191–4205.

Tsoi, Ryan, Feilun Wu, Carolyn Zhang, Sharon Bewick, David Karig, and Lingchong You (2018). "Metabolic division of labor in microbial systems". In: *Proceedings of the National Academy of Sciences* 115.10, p. 201716888.

Van Deynze, Allen et al. (2018). "Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota". In: *PLoS Biology* 16.8. Ed. by Eric Kemen, e2006352.

Venayak, Naveen, Nikolaos Anesiadis, William R Cluett, and Radhakrishnan Mahadevan (2015). "Engineering metabolism through dynamic control". In: *Current opinion in biotechnology* 34, pp. 142–152.

Venturelli, Ophelia S, Alex C Carr, Garth Fisher, Ryan H Hsu, Rebecca Lau, Benjamin P Bowen, Susan Hromada, Trent Northen, and Adam P Arkin (2018). "Deciphering microbial interactions in synthetic human gut microbiome communities". In: *Molecular Systems Biology* 14.6, e8157.

Venturelli, Ophelia S, Robert G Egbert, and Adam P Arkin (2016). "Towards engineering biological systems in a broader context". In: *Journal of molecular biology* 428.5, pp. 928–944.

Venturelli, Ophelia S, Hana El-Samad, and Richard M Murray (2012). "Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response". In: *Proceedings of the National Academy of Sciences* 109.48, E3324–E3333.

Venturelli, Ophelia S., Mika Tei, Stefan Bauer, Leanne Jade G. Chan, Christopher J. Petzold, and Adam P Arkin (2017). "Programming mRNA decay to modulate synthetic circuit resource allocation". In: *Nature Communications* 8, p. 15128.

Vesper, Oliver, Shahar Amitai, Maria Belitsky, Konstantin Byrgazov, Anna Chao Kaberdina, Hanna Engelberg-Kulka, and Isabella Moll (2011). "Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli". en. In: *Cell* 147.1, pp. 147–157.

Volke, Daniel C. and Pablo I. Nikel (2018). "Getting Bacteria in Shape: Synthetic Morphology Approaches for the Design of Efficient Microbial Cell Factories". In: *Advanced Biosystems*, p. 1800111.

Weiße, Andrea Y, Diego A Oyarzún, Vincent Danos, and Peter S Swain (2015). "Mechanistic links between cellular trade-offs, gene expression, and growth". In: *Proceedings of the National Academy of Sciences*, p. 201416533.

Wetmore, Kelly M et al. (2015). "Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons." In: *mBio* 6.3, e00306–15.

Xu, Peng, Kangjian Qiao, and Gregory Stephanopoulos (2017). "Engineering oxidative stress defense pathways to build a robust lipid production platform in Yarrowia lipolytica". In: *Biotechnology and Bioengineering* 114.7, pp. 1521–1530.

Yamaguchi, Yoshihiro, Hirofumi Nariya, Jung-Ho Park, and Masayori Inouye (2012). "Inhibition of specific gene expressions by protein-mediated mRNA interference". en. In: *Nature Communications* 3, p. 607.

Yamaguchi, Yoshihiro, Jung-Ho Park, and Masayori Inouye (2011). "Toxin-antitoxin systems in bacteria and archaea". In: *Annual review of genetics* 45, pp. 61–79.

Yao, Lun, Ivana Cengic, Josefine Anfelt, and Elton P Hudson (2015). "Multiple gene repression in cyanobacteria using CRISPRi". In: *ACS synthetic biology* 5.3, pp. 207–212.

Yeung, Enoch, Aaron J Dy, Kyle B Martin, Andrew H Ng, Domitilla Del Vecchio, James L Beck, James J Collins, and Richard M Murray (2017). "Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks." In: *Cell Systems* 5.1, 11–24.e12.

Zalatan, Jesse G et al. (2015). "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds". In: *Cell* 160.1-2, pp. 339–350.

Zhang, Haoqian et al. (2014). "Programming a Pavlovian-like conditioning circuit in Escherichia coli". In: *Nature communications* 5, p. 3102.

Zhang, Yonglong, Junjie Zhang, Hiroto Hara, Ikunoshin Kato, and Masayori Inouye (2005). "Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase". In: *Journal of Biological Chemistry* 280.5, pp. 3143–3150.

Zhang, Yonglong, Junjie Zhang, Klaus P Hoeflich, Mitsuhiko Ikura, Guoliang Qing, and Masayori Inouye (2003). "MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli". In: *Molecular cell* 12.4, pp. 913–923.

Zhou, Kang, Lihan Zhou, Qing'En Lim, Ruiyang Zou, Gregory Stephanopoulos, and Heng-Phon Too (2011). "Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR". In: *BMC molecular biology* 12.1, p. 18.