Synthesis of Biological and Mathematical Methods for Gene Network Control

by

David J Menn

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2018 by the
Graduate Supervisory Committee

Xiao Wang, Chair
Samira Kiani
Karmella Haynes
David Nielsen
Pamela Marshall

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Synthetic biology is an emerging field which melds genetics, molecular biology, network theory, and mathematical systems to understand, build, and predict gene network behavior. As an engineering discipline, developing a mathematical understanding of the genetic circuits being studied is of fundamental importance. In this dissertation, mathematical concepts for understanding, predicting, and controlling gene transcriptional networks are presented and applied to two synthetic gene network contexts. First, this engineering approach is used to improve the function of the guide ribonucleic acid (gRNA)-targeted, dCas9-regulated transcriptional cascades through analysis and targeted modification of the RNA transcript. In so doing, a fluorescent guide RNA (fgRNA) is developed to more clearly observe gRNA dynamics and aid design. It is shown that through careful optimization, RNA Polymerase II (Pol II) driven gRNA transcripts can be strong enough to exhibit measurable cascading behavior, previously only shown in RNA Polymerase III (Pol III) circuits. Second, inherent gene expression noise is used to achieve precise fractional differentiation of a population. Mathematical methods are employed to predict and understand the observed behavior, and metrics for analyzing and quantifying similar differentiation kinetics are presented. Through careful mathematical analysis and simulation, coupled with experimental data, two methods for achieving ratio control are presented, with the optimal schema for any application being dependent on the noisiness of the system under study. Together, these studies push the boundaries of gene network control, with potential applications in stem cell differentiation, therapeutics, and bio-production.

DEDICATION

To my perfect wife Hao Hu.

I wouldn't have made it without you.

I'll love you for one hundred years, always starting today.

To Jonathan & Nancy, Yaping & Huiying, Liang, Jessica & Rob, John & Sadie, Rebekah & Henry, Paul & Emma, and Julia for being a great and supportive family.

ACKNOWLEDGMENTS

Thanks to the Wang Lab. To Dr. Xiao Wang for mentoring me and giving me the opportunity to learn how to do science in the first place. To Kylie Standage-Beier, Christophe Faucon, and Parithi Balachandran for making the day to day grind a little easier. To the rest of the Wang Lab for being a good group of folks to work around.

Thanks to the Kiani Lab (and Mo's lab too!). To Dr. Samira Kiani for inviting me into her lab when I was struggling with biology and giving me much needed advice. To Dr. Mo Ebrahimkhani for incisive suggestions during lab meeting. To Farzaneh Moghaddam, Swechchha Pradhan, Jeremy Velazquez, and Ryan LeGraw for keeping life interesting over in ISTB1.

Thanks to the Haynes Lab. To Dr. Karmella Haynes for helpful conversations and providing a realistic outlook on life. To René Daer, Cassandra Barrett, and the rest of the Haynes lab for being good neighbors and being supportive of me and Hao.

Thanks to the Brafman Lab. To Dr. David Brafman for collecting such a good group of people for me to be work neighbors with. To Josh Cutts, Nick Brookhouser, and the rest of the Brafman lab for good conversations at work and good times outside of it.

Thanks to Drs. Pamela Marshall and David Nielsen for helpful discussion helping me succeed in this Ph.D.

Thanks to Arizona State University and the School of Biological and Health Systems Engineering for financial support through the Dean's Fellowship.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 SYNTHETIC BIOLOGY

Biology is a complex topic, nearly infinitely so. Ignoring the complexity that comes with multicellularity, even the simplest cells are powerful computational organisms, capable of synthesizing information from myriad self and environmental inputs to make basic decisions: to move (1), to divide (2, 3), to absorb nutrients (3–5), to adsorb to the environment (6–9), and much more (10, 11). When multicellularity is considered, the variety of observed behaviors expands to even more possibilities: differentiation into one of many diverse tissue types (12–17), intricate intracellular communication (18–20), or partaking in large-scale actions – such as contracting a muscle – that utilize the combined effort of thousands or millions of cells (21–23).

Much of what we know about how biological systems make decisions is based on what has come to be known as the "Central Dogma" of biology: deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA) which is in turn translated into proteins which carry out specific functions throughout the cell (24). While this is, of course, an incredibly simplified version of what goes on in the cell which ignores some concepts that might be considered fundamental – epigenetics and DNA silencing, for example (25–27) – it has, nonetheless, been a powerful framework for conceptualizing a vast swathe of cellular behavior and holds true for all known organisms. While the Central Dogma is presented as a linear progression from stored data (DNA) to coded message (RNA) to signal effect (protein), it is in fact, quite frequently, a circular process (28–32). The primary function of a large class of proteins, known as transcription factors, is the

regulation of DNA transcription, increasing or decreasing that rate at which specific genes (segments which code for a protein) are transcribed into RNA (33, 33, 34).

These regulatory interactions are still far from well understood. Partly, this is because the genetic engineering field is relatively young, so technology, techniques, and best practices for exploring intercellular processes are still in need of refinement. Possibly more confounding, however, is the vast interconnectivity present in natural biological systems (35–37). It may come as no surprise that this is the case, as cells have evolved to made decisions by synthesizing information from a wide range of sources. For understanding the fundamentals of gene regulation; however, this interconnectivity can be a strong barrier to understanding, as induced changes to a natural system may have many and far-reaching unintended off-target effects (38, 39).

It is here that the field of synthetic biology shines (40). Due to the universality (to our knowledge) of the DNA code and the Central Dogma, it is possible to transplant genes from one organism into another to which they are non-native (41, 42). In this new host, most of the regulatory interconnectivity is stripped away, allowing researchers to study gene regulation in a less confounded form. Multiple genes can be inserted into a new host organism in this way, with engineered interactions between them based on gene and promoter selection, meaning that gene networks can be designed and built from the ground-up (43–46). This bottom-up approach to genetics serves as a strong compliment to the top-down approach which has been the foundation of most biological science until relatively recently. Synthetic biology, in taking an engineering approach to constructing gene networks, opens a new set of questions to be answered and allows biological engineers to employ a build-to-understand methodology (41).

In addressing biological phenomena from an engineering perspective, synthetic biology employs the most powerful tool in the engineering toolbox – math – with the intent of transitioning from a qualitative, descriptive understanding of biological processes to a quantitative, predictive one (47). To do so, a solid grasp of the underlying biology and of mathematical methods by which to understand the biology are required. There are unique challenges associated with biological systems that differentiate it from many other engineering fields. Primary among these are the ideas of stochasticity and decision making (48, 49). While both concepts are present in other engineering fields, they are at the forefront of biological engineering and warrant special consideration.

Stochasticity refers to the randomness inherent to biological processes (50). The intracellular environment is packed with all manner of proteins, molecules, and atoms jostling around semi-randomly, following roughly Brownian kinetics (51). Nearly every aspect of cellular action is dependent on these random interactions. While randomness can be considered in deterministic terms when viewed on a large enough scale, the scale at which cells operate often makes this mass-action assumption inappropriate. Therefore, it is up to biological engineers to understand the role that stochasticity plays in intra- and inter-cellular interactions (52–54).

Relatedly, the concept of decision making arises frequently when discussing cellular systems (49, 50). While this topic also arises in other engineering fields, such as computer or software engineering, the biological mechanisms and inherent stochasticity makes conceptualization more difficult (55). Unlike computers which work with a discrete, binary system to perform logic operations, biology is – often frustratingly – analog (56, 57). Therefore, in addition to the question of whether a cellular system can

perform an operation, this introduces the idea of the robustness of that decision-making behavior (58–61).

Together, these challenges provide a unique landscape for biological engineering and synthetic biology to occupy. In addition to attempts to understand fundamental biological interaction from a ground-up perspective, unique challenges of stochasticity and decision robustness must be simultaneously engaged. In this dissertation, I attempt to do just that.

1.2 IN THIS DISSERTATION

Chapter 2, "Control of Synthetic Gene Networks and its Applications," delves into the relationship between biology and mathematics. One of the underlying assumptions of synthetic biology is that biological processes can be engineered in a controllable way. Here we discuss this assumption as it relates to synthetic gene regulatory networks (GRNs). I first cover the theoretical basis of GRN control, then address three major areas in which control has been leveraged: engineering and analysis of network stability, temporal dynamics, and spatial aspects. These areas lay a strong foundation for further expansion of control in synthetic GRNs and pave the way for future work synthesizing these disparate concepts.

Chapter 3, "Modeling Gene Networks to Understand Multistability," covers some more specific ways that mathematics can be leveraged to understand biology. Stem cells are unique in their ability to differentiate into diverse phenotypes capable of displaying radically different, yet stable gene expression profiles. Understanding this multistable behavior is key to rationally influencing stem cell differentiation for both research and

therapeutic purposes. To this end, mathematical paradigms have been adopted to simulate and explain the dynamics of complex gene networks. In this chapter, I introduce strategies for building deterministic and stochastic mathematical models of gene expression and demonstrate how analysis of these models can benefit our understanding of complex observed behaviors. Developing a mathematical understanding of biological processes is of utmost importance in understanding and controlling cellular behavior.

Chapter 4, "Fluorescent Guide RNAs Facilitate Development of Layered Pol II Driven CRISPR Circuits," presents original research on utilizing the clustered regularly interspaced short palindromic repeat (CRISPR) system for building gene networks. Efficient CRISPR guide RNA (gRNA) expression from RNA Polymerase II (Pol II) promoters will aid in construction of complex CRISPR-based synthetic gene networks. Yet, we require tools to properly visualize gRNA directly to quantitatively study the corresponding network behavior. To address this need, I employed a fluorescent gRNA (fgRNA) to visualize synthetic CRISPR network dynamics without affecting gRNA functionality. I show that studying gRNA dynamics directly enables circuit modification and improvement of network function in Pol II-driven CRISPR circuits. This approach generates information necessary for optimizing the overall function of these networks and provides insight into the hurdles remaining in Pol II-regulated gRNA expression.

Chapter 5, "Intracellular Noise Level Determines Ratio Control Strategy, Speed, and Accuracy," presents original research on inducing cellular populations to partially differentiate into reliable fractions. Robust and precise ratio control of heterogeneous phenotypes within an isogenic population is a common phenomenon, especially in the development and differentiation of large number of cells such as bacteria, sensory

receptors, and blood cells. However, the mechanisms of such ratio control are poorly understood. Here, I employ experimental and mathematical techniques to understand the combined effects of signal induction and gene expression stochasticity on phenotypic multimodality. I identify two methods for generating phenotypic ratios from an initially homogenous population, suitable roughly to high-noise and low-noise intracellular environments, and I show that both can be used to generate precise fractional differentiation. In noisy gene expression contexts, such as those found in bacteria, induction within the circuit's bistable region is enough to cause noise-induced bimodality within a feasible timeframe. However, in less noisy contexts, such as tightly controlled eukaryotic systems, spontaneous state transitions are rare and hence bimodality needs to be induced with a controlled pulse of induction that falls outside the bistable region. Finally, I show that noise levels and system response time impose limitations on both ratio control methods, and I develop a framework for determining the best method for a given parameter set.

Chapter 6, "The Future of Synthetic Biology," explores issues that are still outstanding in the synthetic biology field and offers paths forward for further research.

2 CONTROL OF SYNTHETIC GENE NETWORKS AND ITS APPLICATIONS

This chapter was prepared in collaboration with Ri-Qi Su, Ph.D. and Xiao Wang, Ph.D. for publication in *Quantitative Biology* (47).

2.1 INTRODUCTION

Synthetic biology, since its flourishing in the early 2000s, has undergone rapid advancement. The is due in a large part to the promise that the field holds for such diverse applications as bioenergy (62–64), personalized medicine and therapeutics (65), bioremediation (66, 67), and biopharmaceuticals (68). Paired with advancements in DNA synthesis (69–71) and sequencing (72), the field has grown exponentially in the last 15 years. This growth has seen the development of new tools, such as the implementation of the Cas9 protein for gene editing and transcriptional regulation (73–75), to further plumb the depths of our biological understanding and the applications thereof. In particular, advancements in synthetic biology have allowed the study of gene regulatory networks (GRNs) in a simplified setting amenable for precise experimental controls (76, 77).

Using a build-to-understand, bottom-up approach (41, 78, 79), synthetic biologists can strip away much of the complexity of highly interconnected natural biological systems while studying gene regulation within an *in vivo* (78, 79) or *in vitro* system (80, 81). Synthetic networks which function orthogonally to natural networks give researchers more control over their behavior and avoid the confounding effects of the many unknown genetic interactions endemic to natural systems (18, 51). These functional synthetic networks have been used to demonstrate many fundamental biological processes such as

multiple stabilities (79, 82, 83), complex temporal behavior (43, 77, 78, 84), and rich spatial patterning (85–87). Many of these small functional networks, often referred to as motifs, utilize positive or negative feedback topologies. For example, bistable GRNs can be constructed either through the use of two mutually inhibitory components (79) or with self-activating components (83). From these small functional networks, great efforts have been made to build up, combining multiple motifs into larger and more complex networks (44, 45, 88, 89). In addition to realizing immediate applications for complex synthetic networks, engineered circuits also shed additional light on the underlying mechanisms of biological regulation and control (51, 82). For example, by constructing a symmetrical circuit expressing two different fluorescent proteins, Elowitz *et al* demonstrated the existence of intrinsic and extrinsic stochasticity within a cell (51). While Wu *et al* illustrated impacts of such stochasticity on cell fate determination using a synthetic toggle switch in yeast (82).

Control of GRNs has been a constant research focus and is of paramount importance to continued advancement in synthetic biology. In its most abstract sense, GRN control refers to methods by which researchers can engineer, modulate, and predict robust network behavior. In a physical sense, this entails proper selection of internal and external factors which influence network behavior. Internal factors function as a closed feedback control loop within the cell and include the selection of cell type, GRN topology and motifs, and specific components comprising the GRN: promoters driving individual gene's expression (69, 84), transcription factors modulating expression of downstream genes (77, 79), localization signals or synthetic protein domains affecting protein interactions (18, 85, 86), *etc*. Once put into the cell, many of these components do

not leave researchers with a direct means of interaction. External factors form an open control loop and are therefore easier to control throughout the course of an experiment. These include factors such as growth media composition (69, 82, 90), ambient temperature (77, 91, 92), light exposure and wavelength (93, 94), magnetic fields (91, 95), or small molecule inducers which either bind surface receptors or permeate the cell to cause changes in protein behavior (18, 69, 77, 79). A crucial third component in GRN control is the theoretical framework that describes the predictability of the system and allows synthetic biologists to compose networks toward a desired outcome, rationally select components to achieve that outcome, and predict the parameters under which that desired outcome is attainable (41, 96, 97).

In the following sections, we first examine in close detail the theoretical basis of GRN control. We then use this framework to inform discussion on three aspects of GRN behavior: multi-stability, temporal dynamics, and spatial relations. Each of these areas is further explored with discussion of how the intrinsic and extrinsic biological control factors relate to the theoretical framework, problems faced in realizing these behaviors, and examples and applications of the

behavior to broader aims.

2.2 THEORIES AND COMPUTATION OF CONTROL

The mathematical foundation of control theory has been well developed for both linear and nonlinear dynamical systems (98, 99). Its application to many biological fields, especially systems biology, has produced progress in both designing experiments and understanding results (100, 101). In synthetic biology, with its bottom-up design mindset,

the functional motifs are well isolated from the cellular environment and thus provide better test platforms for control theories in biology. However, there are several challenges remaining in mathematically modeling and predicting gene network functions: genetic networks are often highly nonlinear, cellular environments and internal kinetics are stochastic, and natural genetic networks can have high dimensionality with unknown interconnectivity between genes.

Several approaches have been taken in addressing GRN nonlinearity. Ordinary differential equations (ODEs) are frequently used to model deterministic systems with the aim of obtaining a sketch of the underlying interactions and the effect of varying parameters within these systems (79, 82). The regulation of gene expression has often been described in the form of nonlinear Hill equations (102). Many theoretical approaches for analyzing nonlinear GRNs (96, 103, 104) have borrowed from the large amount of work that has been developed around linear control theory (98, 99, 105). These approaches have been adapted into several network simplification methodologies. By linearizing nonlinear Hill functions around an equilibrium point, Shin *et al* studied the transfer function for simple GRNs and reproduced experimental results in continuous models (96). Liu *et al* proposed a linear control theory for large networked systems (103) and used it to analyze minimum control inputs in metabolic networks (104).

However, most of these theoretical approaches still cannot be directly applied to model and predict complex behaviors of GRNs, and an *ad hoc* model based on ODEs is still required for each specific system. After developing a system of ODEs to describe a GRN, bifurcation analysis is often employed to investigate how the network's

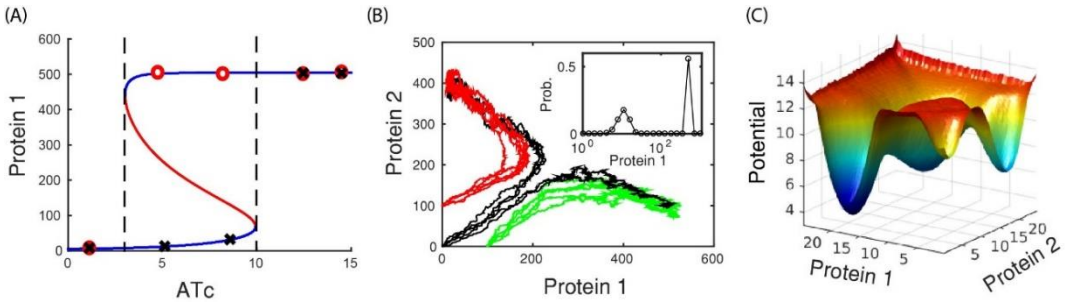

Figure 2-1 Gene network multistability analysis

(A) Bifurcation diagram of the toggle switch controlled by the concentration of ATc (43) and illustration of hysteresis analysis. The blue lines represent the SSSs under a range of ATc concentrations, while the red line represents the unstable steady states. The black crosses are the predicted SSSs for the cell which is first grown in media lacking ATc then transferred to media with variable ATc concentrations, while the red circles represent SSSs for cells initially grown in high ATc concentrations (250 ng/ml). By performing hysteresis analysis, the bistable region can be identified without knowing the system parameters, denoted as the area between the two dash lines. **(B)** Temporal trajectories simulated using Gillespie algorithm for experiments in (82). The black trajectories were initiated directly on the separatrix, while the red and green ones were not. The inserted panel shows the resultant histogram for trajectories which began on the separatrix and subsequently differentiated into two populations. **(C)** A pseudo potential landscape for systems with four SSSs, as shown in (41). SSSs are represented by local minima within the parameter space, while the stability of any given state is represented by the depth of the energy well relative to its neighbors.

deterministic behaviors change with system parameters. This can reveal parameter regions of multistability and phase transitions (18, 82, 105, 106), as it is shown in Figure 2-1A. This is important for investigating networks which can have multiple states, such as toggle switches which can switch states in response to environmental stimuli (43, 79), and it can also be applied to oscillatory systems (77). This *in silico* method can be paired with hysteresis analysis, an experimental design which is used to probe dynamical systems without knowing their detailed dynamical form or parameters (79, 107, 108). Hysteresis analysis involves performing experiments to investigate the parameter space and its effect on system stability. It is often paired iteratively with bifurcation analysis to further develop the model to more accurately describe the system in question (79). Figure 2-1A demonstrates how the bistable region can be identified by hysteresis analysis without knowing exact system parameters.

ODEs and other deterministic methods have been invaluable in many GRN analysis applications; however, the strict determinism of these equations limits their application to cellular behaviors influenced by stochasticity (52). Stochastic simulation tools, including the stochastic differential equation (SDE), cellular automata (CA), potential landscape (59), and Gillespie algorithm (GA) (109) are used to simulate and study inherently noisy processes within the cell (41, 51, 59). Deterministic models fail in situations in which there are several potential outcomes from a common set of parameters and initial conditions, such as the stochastic differentiation from an undetermined state to one of two SSSs shown in Figure 2-1B. When the trajectories are initiated from the separatrix dividing the two states' energy wells (82), they differentiate into one or the other population as a result of gene expression noise. The resulting distribution can be predicted using Gillespie algorithm. When the system's gene expression is perturbed by external stimuli, the overall expression distribution amongst various attractor basins can be quantified using SDEs or GA. Furthermore, the pseudo potential landscape can be portrayed from the stationary distribution, which provides better characterization of system stabilities and state regulation behaviors for the GRN (41, 82), as shown in Figure 2-1C. The positions together with the stabilities of all SSSs of a system can be illustrated by the pseudo landscape. PDEs and GA can describe the stochastic interactions between different cells and simulate how the cell population is distributed in space (59, 87, 110). In their study of synthetic ecosystems, Song *et al* modeled the spatio-temporal dynamics of two synthetic *Escherichia coli* populations using PDEs (111, 112). On an intracellular level, GA (or the related Monte Carlo simulation) is often used to simulate stochastic fluctuations in transcriptional regulator numbers (60, 82). These types of simulations can

be used to determine likelihood of state transitions under varying amounts of noise (106) or to thoroughly analyze the GRNs potential landscape under a single noise condition (59).

Though synthetic networks are, thus far, limited in size, understanding the regulation of cell differentiation and state transitions within a natural system requires computational tools capable of dealing with large dimensionality (35). On one hand, researchers have tried to abstract the large scale GRNs into different motifs with varying functions (113). They then tried to understand the relationship between the motifs' structure and cellular behavior (78, 79, 114). On the other hand, PDEs and CA have also been applied to model gene expression distribution over time of high dimensional systems. For example, Wang *et al* proposed a pseudo potential landscape based on the equilibrium distribution in state space of gene expression levels, and solved the equilibrium distribution using a PDE model (59). CA models have been utilized to simulate complex stochastic interactions between cells. These are used frequently in tumor modeling, where individual cell behavior within a group is highly dependent on its immediate neighbors and environment rather than relatively simple chemical gradients (115). The primary drawback of CA models is that they tend to consume large computational resources (87).

Studying control problems in GRNs will generally require the application of multiple theoretical tools at the same time. One prominent example which has important ramifications in many areas is how to control transitions between different stable steady states (SSSs) in gene expression space. This is important in an area like cancer research, since cancer is frequently characterized as cells which have fallen into an unhealthy but

stable gene expression state (116). With better control strategies, it may be possible to transition these genes' expression back to a healthy state via a novel route: a different tactic from what current therapies provide. To this end, Wang *et al* used bifurcation analysis to identify possible transitions paths between different SSSs in multistable GRNs, and then suggested to model the GRNs as a network of attractors to reduce its dimensionality. Based on hysteresis analysis, they proposed transient and sequential control signals to navigate the state transition from an arbitrary cancer attractor to a health attractor (41). Separately, work has been done on minimizing the effects of failed nodes within a larger network and on determining the best methods to limit large-scale cascading effects if single nodes display anomalous behavior (117). Finally, ongoing progress has been made in developing frameworks for understanding and controlling genetic regulation and metabolic flux in complex biological networks (104).

2.3 MULTISTABILITY OF GRNS

Multistable systems can hold two or more stable gene expression profiles, SSS, with the same set of parameters. This ubiquitous property of natural systems allows isogenic populations to express a range of behaviors in response to their needs and environment (118). In single-celled organisms, this division of labor can lead to increased population fitness. In bacteria, for example, often a sub-population can enter a competent state in which the uptake of foreign DNA is increased, allowing the bacteria to increase genetic diversity (119). Similarly, some bacteria within a population may enter a state called persistence, in which the cell becomes dormant (11). If a catastrophic event, such as contact with an antibiotic, wipes out the colony, these persistent cells can remain

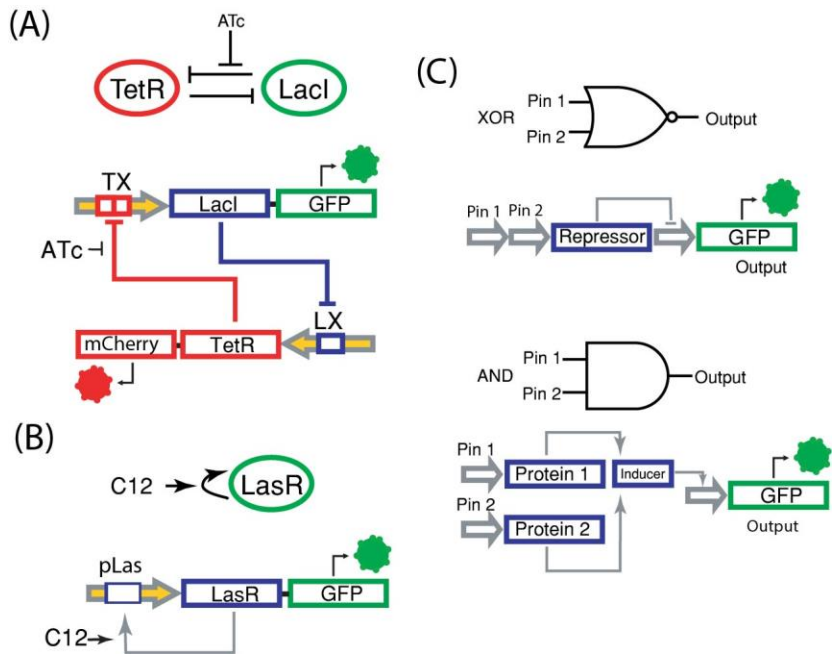

Figure 2-2 Common gene network motifs

(A) Schematic diagram and simplified schematic for the mutual inhibition toggle (79). **(B)** Another bistable circuit of autoactivation (121). **(C)** Schematic diagram for two representative logic gates, the XOR and AND gates reviewed by Singh *et al* (122).

unaffected, thereby ensuring the colony's survival (120). In multi-cellular organisms, the role of multistability is primarily to allow the development of a multitude of tissues from a single stem state (36).

Topologically, there are two general ways for a system to achieve multiple SSSs: mutual inhibition and autoactivation. In a mutually inhibitive GRN, the gene or genes associated with one state actively repress the expression of those associated with one or more competing genes and vice versa (79, 121), as seen in Figure 2-2A. Synthetic mutual inhibition circuits have been demonstrated in multiple organisms (79, 82, 122), and examples of similar topologies are rife in nature (118). A system can also express multistable behavior through autoactivation (123). As illustrated in Figure 2-2B, a single gene can keep itself activated if its expression has passed a certain threshold; below this

threshold, however, the gene remains inactive (124). Self-activating motifs like this tend to be noisy on their own, but they can also play a stabilizing role to the expression of mutual inhibition GRNs (90). From these basic underpinnings, several control problems have come to define the study of multistability.

First, researchers have sought to better understand and control the proportionality of differentiation into various SSSs. Wu *et al*, studied the effect of both internal and external factors on differentiation into one of two states in a synthetic yeast network (82). Using a novel design strategy, researchers positioned the cells' expression near the separatrix dividing the energy wells of mutually inhibitive red-expressing or green-expressing states. The stochastic process of gene expression then caused the cells to gravitate towards either the red or the green state. By changing the promoters driving the antagonizing repressors, and by changing inducer concentrations to alter the efficiency of those repressors, the percentage of cells falling into each state can be tuned. Both methods of controlling cell fate determination show how changing the underlying energy landscape of a multistable system can affect the behavior of the system itself. Ishimatsu *et al* built on this foundation, using gene overexpression to force a bistable network into temporary monostability (90). By tuning overexpression, the single steady state could be adjusted in state-space, and this adjusted location became the new initial point from which the cells would differentiate upon cessation of overexpression. By placing the cell expression near the system's separatrix, a tuning of the population fraction in each state was observed, similar to that demonstrated by Wu *et al* (82).

This leads to a second area of study: how to control the transitions between states in multistable systems. Bifurcation analysis is a commonly employed method for

investigating the parameter space in which a system can maintain multistability. On either side of the multistable region there is a bifurcation point: a parameter value at which one of the SSSs disappears or emerges. Using transient chemical or thermal induction, Gardner *et al* showed that bistable networks could be switched between states by temporarily forcing them out of the bistable region (79). Ellis *et al* further demonstrated that the transition time between states could be both predicted mathematically and tuned through selection of different promoters from a synthetic library in order to temporally control the flocculation of yeast (43). Unlike purely stochastic cell fate determination process used to tune population percentages, these experiments showed that full populations' expression could be controlled essentially deterministically with a high degree of accuracy, accounting for both expression levels and transition times.

Related to both of these areas of study is the control of a system's multistable region itself. Multistability generally occurs only within a small range of parameter and induction values. To engineer robust networks, expansion of multistable regions is crucial. This is partially determined by the network topology and relies on proper selection of network components. Using a library of synthetic promoters, it has been shown that the same topology can yield bistable regions responding to low, mid, or high levels of induction (82). Additionally, the regulatory proteins used have a profound effect on hysteresis behaviors. In an autoactivation network, Wu *et al* demonstrated that different pairs of activator and chemical inducers produced different bistable regions (18). Interestingly, it was also demonstrated that pairing poorly interacting

inducer/activator pairs – due to quorum sensing (QS) crosstalk in this instance – yielded an expanded toolbox of parts with a range of bistable regions to choose from.

As the physical construction of multistable networks has expanded, so too has the investigation of the theoretical underpinnings of stability. A large amount of *in silico* work has been dedicated to understanding the topological basis of multistability. For example, Yao *et al* identified a minimal circuit to generate bistability from a simplified Rb-E2F network which regulates the initiation of DNA replication (125). Faucon *et al* looked for instances of possible ways in which a three-gene network could exhibit multistability (106). Additional work has been done to demonstrate the role that small motifs play in enhancing network stability (126).

Studies such as this focus on both how to attain multistability as well as on quantifying how stable the discovered multistable states are. Quantification of the stability of an energy well is still an evolving field, and stochastic simulations are often applied to determine how likely a cell is to jump out of a given SSS due to inherent noise.

Finally, there are also stationary synthetic circuits that are not multistable but have multiple outputs and can be used to control and integrate environmental and cellular signals. By layering multiple feedback systems, researchers have created digital logic gates, as reviewed in (88, 89, 127). Researchers have demonstrated the ability to engineer AND, OR, NOR, XOR, NOT, and NAND gates in multiple organisms (127–130). Two examples of such logic gates are shown in Figure 2-2C. However, with increasing complexity comes increasing design constraints. Synthesis of two or more inputs can require engineering of new synthetic promoters capable of interacting with multiple proteins, and layered circuits require that upstream gene expression be clear enough that

the signal translates into downstream expression. Regulator crosstalk can become a problem, as more parts are added to a circuit. Wu *et al* designed an orthogonal AND gate in *Escherichia coli* and studied the effects of regulators' crosstalk in autoactivated quorum sensing circuits (18). The effect of integrating multiple module layers can be even more unpredictable, so additional design principles from engineering of digital control systems, such as timescale separation, have been introduced to overcome these obstacles. Mishra *et al* designed a genetic device called the 'load driver' to mitigate the interference between different genetic modules in *Saccharomyces cerevisiae* (131, 132).

2.4 TEMPORAL DYNAMICS OF GRNS

While gene network stability is important for developmental processes, cell differentiation, and population fitness, dynamic temporal behavior is equally relevant to sustained biological processes. For instance, many cellular processes are informed by the oscillatory dynamics of the cell cycle (133) or by the daily circadian rhythm (134). Additionally, certain sensory inputs are subject to the phenomenon of adaptation, in which a stimulatory signal has a reduced effect if introduced repeatedly within a short period (135). Relatedly, a large amount of intercellular signaling is due to temporal bursts of activity, as seen in neuronal spiking (136) and the subsequent release of regulatory neurotransmitters (137). To be able to engineer biological processes effectively, researchers need to be able to control the time scales together with the stabilities of these types of behaviors by utilizing internal and external control methods.

Oscillators were some of the earliest dynamic synthetic GRNs (78). Most instances of oscillation have been shown to arise from two primary topologies: a three-

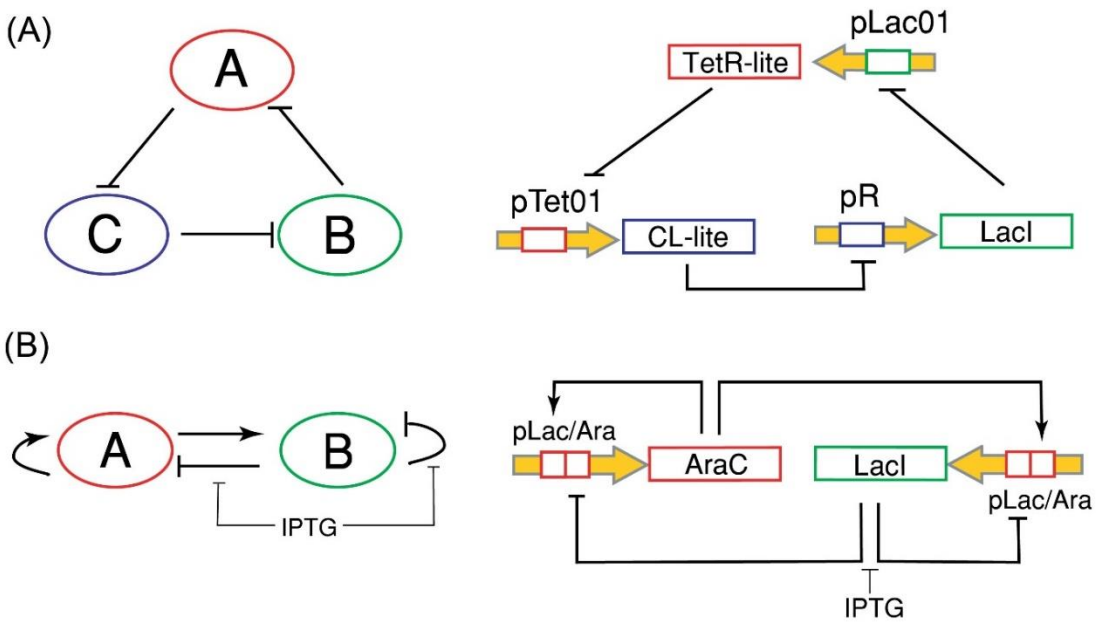

Figure 2-3 Oscillatory gene networks

(A) Simplified illustration and schematic diagram of the represillator from (78) and (84). The three nodes A, B and C jointly form a negative feedback loop. **(B)** Simplified illustration and schematic diagram of the two-node oscillatory network (77) with autoactivation on node A. The negative feedback strength can be controlled by the concentration of IPTG which regulates the repression strength from node B to A.

node negative feedback loop known as the represillator (78, 84) and a two-node network comprised of one autoactivator and one repressor (77) (see Figure 2-3A and B). In both topologies, the oscillator relies on two kinetic elements: negative feedback and a delay which grants enough time for one gene to turn on before being turned off by an antagonizing gene. To this end, control of oscillatory GRNs relies on tuning the negative feedback loop and/or the regulation duration, thus selection of GRN components and external regulators which alter degradation, production, and regulation kinetics play an important role in producing the desired oscillatory behavior. Much research has been done to modulate the frequency, stability, and synchronization of oscillations within a large population (77, 78, 84, 138).

The negative feedback is the most critical component for generating oscillations. Since the kinetics associated with feedback loops can be described by nonlinear ODE functions, possible approaches to control the oscillatory behavior can be analyzed and predicted using bifurcation analysis across a range of parameters. Changes in parameters can be caused by altering production efficiency of network genes, regulation strength between genes, concentration of inducers, and environmental conditions. As with multistability, bifurcation analysis can be utilized in conjunction with oscillatory network models to determine parameter ranges at which one is likely to observe oscillation, and it can give insight into the range and frequency of the expected fluctuations (139, 140). For examples, Stricker *et al* introduced autoactivation into the negative feedback circuit which enhanced the production activities in node **A** (See Figure 2-3B). Also, by changing the concentration of IPTG, which inhibited the repression of node **A** by node **B**, the regulation of negative feedback was be modulated, allowing further control of the oscillatory periods. It was also shown that the oscillatory period decreased when the temperature was increased due to a decrease in cell doubling time (77). Additionally, together with the frequency, the amplitude of oscillations can also be controlled with relative plasmid dosage changes in mammalian cells (138). The negative feedback loop can be extended to intercellular processes with the help of QS genes, which produce diffusible signaling molecules, and further achieve synchronization within a population. This sort of synchronization behavior has been modeled synthetically in bacteria (77), and it is a first step toward engineering large scale oscillation synchronization, an important aspect of multicellular life.

Time delay within the feedback loop is another critical element that generates oscillatory behavior in GRNs. A sufficiently long delay has been numerically demonstrated to be one of the required conditions to generate oscillations from a single autoinhibitory gene (141). The time delay in the feedback loop arises from finite interactions and production time in stochastic gene expression and can be highly noisy, so the effect of delays needs to be analyzed via stochastic simulation tools. There are many ways to extend or shorten the delay to further control the period and robustness of oscillation. In their pioneering work of synthetic oscillatory GRNs, Elowitz *et al* synthesized three cascading repressors into a represillator (78) and extended the delay duration by introduction of additional cascading processes. Genomic structure can also affect the delay; Swinburne *et al* engineered oscillatory GRNs in animal cells and found that longer introns, which require longer production time, can increase the transcriptional delays, thus generating longer expression pulses (142). The delay duration will also affect the stability of resulting oscillations. Potvin *et al* reduced the delay in the original represillator by choosing low copy plasmids and thereby generated more stable oscillations while maintaining a minimal topology (84).

Another biologically relevant temporal behavior is that of spiking. This can be seen in neural signaling or in the response to certain sensory inputs, in which an external stimulus causes a short burst of activity before the system returns to a resting state. Adaptation is a dynamic behavior in which an extracellular signal causes the temporary excitation of a GRN which eventually returns to its basal level despite the signal's continuation (114). Ma *et al* exhaustively identified all possible three node GRNs topologies that can generate adaptation signals, and they found that the precision and

sensitivity of adaptation can be independently modulated by tuning the system's parameters. Unlike oscillations, adaptation does not rely on a bifurcation for the desired behavior to appear; rather, it is a result of temporary perturbation of an otherwise stable network. Alternatively, spiking signals of neural systems exist in networks which operate close to a bifurcation point, so small environmental cues can push the cell into a region of either random or periodic spiking, depending on the network topology (143).

2.5 SPATIAL ASPECTS OF GRNS

In addition to the multistability and temporal aspects of GRNs, another active area of synthetic biology research is the exploration of the spatial properties of complex networks. Organization of individual cells into population wide patterns is a common behavior found throughout nature (85, 144) and biomedical applications (86, 145). There are many open questions in controlling GRN regulated spatial patterns. In lower organisms, there have been two primary foci in exploring GRN spatial properties: pattern formation and population density control (144). Besides their biological significance, synthetic GRNs in bacteria also serve as platforms to study how the cells communicate with each other and respond to the environmental signals. In higher organisms, the primary thrust in studying spatial patterning has been to better understand tissue and organ development (145). Understanding how this emergent behavior can be engineered and controlled can lead to a better understanding of developmental processes, cellular signaling and signal processing paradigms, and construction of complex behavior from simple components.

Typically, synthetic GRN directed pattern formation requires three basic functional modules: a mechanism to send signals, a way to receive extracellular signals, and cellular actions responding to signals. Since spatial distribution and patterning is a population-wide phenomenon, intercellular signaling is required. It has been demonstrated by using QS genetic components (144) that bacterial populations can be engineered to form patterns in response to extracellular concentrations of acyl-homoserine lactones (AHLs), a class of signaling molecules used by QS (85). Once received, the signal is then processed into a stable output phenotype or behavior, typically through a negative feedback topology. When the signal is being processed by the synthetic GRNs the cells can respond by utilizing one or more mechanisms of cell movement (86), proliferation, and/or death (146, 147) in order to form the intended patterns.

The control schemes used in cell-cell communication involve how the specific intercellular regulator molecules are added to the system (external addition or internally produced by the GRN) (85). By controlling the local concentrations AHL, Basu *et al* used programmed synthetic GRNs to form different pattern shapes. They employed co-cultures of engineered sender and receiver cells. The sender cells were designed to synthesize AHL under user defined gradients, while the receiver cells were engineered to operate similar to a bandpass filter. By placing sender cells in different configurations, through their fluorescent outputs the engineered cells jointly expressed different shapes such as a bullseye, ellipse, heart, and clover. Important control variables within the network topology and genetic components used are also widely studied. There are

currently two types of underlying GRNs that can sense environmental signals. One type of GRNs senses morphogen gradients and expresses weaker with increasing

distance from the region of highest concentration (85). Conversely, Payne *et al* developed a novel pattern formation schedule in *E. Coli* equipped with intracellular autoactivation and intercellular negative feedback motifs (87). By employing this mechanism, the pattern scale could self-organize into intended patterns without reliance on a morphogen gradient. The pattern could also be controlled by biological processes. Payne *et al* discussed that the metabolic burden caused by the activated synthetic circuit could actually enhance the pattern robustness (87). Liu *et al* demonstrated control of the mobility of cells and further achieved periodic striped patterns (86). They synthesized a LuxR/LuxI module to synthesize and excrete AHL when the cell density was high, which in turn further activated expression of LuxR. Additionally, the LuxR-AHL complex drove the expression of lambda repressor and further regulated *CheZ* expression, so as to reduce the mobility of the *E. Coli*.

A similar application of pattern formation is to control population density. When bacteria are used for bioproduction applications, it may be beneficial to halt cellular growth in order to force the population to focus on producing the molecule of interest (148). While pattern formation work has primarily taken place in bacteria plated in a dish, density control work seeks to understand and control the requirements for adjusting growth behavior in liquid culture. Similar topologies have been employed: cells were modified with LuxR/LuxI QS together with CcdA/CcdB toxin/antitoxin systems to study population density and individual fitness (148). QS can also be controlled by changing the specific QS regulators employed, taking into account any crosstalk which might occur

between QS components. Wu *et al* demonstrated the important role that crosstalk between signaling molecules and transcriptional regulators (signal crosstalk) or between regulators and promoters (promoter crosstalk), as well as the overall expression intensity of QS components as determined by each component's promoter, can play in a synthetic system (18). They also found that such regulation and crosstalk may induce novel host-circuit interaction in the QS system of LuxR/LuxI and LasR/LasI and can be engineered to generate varying population dynamics.

2.6 CONCLUSION

Control of GRNs in synthetic biology is a quickly expanding field covering all types of network behavior. Here we have provided examples of pioneering work on four key aspects of GRN control: its theoretical basis, stability analysis, temporal dynamics, and spatial distributions. Each aspect holds promise on its own and can be expanded into more complex, robust, and diverse applications. Beyond this, however, synthesis of these aspects of GRN control also promises powerful new tools for understanding and interacting with developmental processes, which innately possess multistable, temporal, and spatial properties.

3 MODELING GENE NETWORKS TO UNDERSTAND MULTISTABILITY

This chapter was prepared in collaboration with Xiao Wang, Ph.D. and has been submitted to Spring Publishing for inclusion as a chapter in upcoming instructional book.

3.1 INTRODUCTION

Multistability is a mathematical property describing a system capable of having two or more mutually exclusive states. These dynamical states could provide stem cells a portfolio of decision options that allow cells carrying a homogenous genetic code to express phenotypic variety (50). Through differential expression of genes in response to environmental factors, cells lock into various expression patterns, upregulating some genes while downregulating others. Network topology shapes the ways that these genes' expression influences one another, either reinforcing or inhibiting each other or themselves. These gene interactions, usually in the form of transcriptional regulation, can be modeled mathematically, and the information gained from the models can be used to more thoroughly understand the behaviors and limits of the network (41, 47). The utility of modeling for understanding complex gene networks has been demonstrated in a variety of contexts: understanding the link between external stimuli and gene expression (43), exploration of gene topologies capable of specific behavior (106, 114), and development of a functional multistable synthetic gene networks (79, 82, 149).

In this chapter, we cover the basics of constructing mathematical models using ordinary differential equations (ODEs), discuss ways to adapt them to simulate noisy biological environments, and discuss implementation for two common applications.

3.2 MATERIALS

The work we discuss is purely *in silico* and requires a personal computer, preferably one with a fast processor, and a programming environment. We prefer MATLAB because of its usability and host of built-in functions. MATLAB also can be used with the MATCONT addon, which adds additional analysis options for solving systems of ODEs. With a knowledgeable user, however, any programming environment with continuation software can be appropriate. Additionally, for some computationally intensive processes, access to a distributed computing center is necessary. This will require some knowledge of parallelization that is beyond the scope of this chapter.

3.3 METHODS

3.3.1 Applications

Before getting into details on how to build and run gene network models, we will discuss two main applications of mathematical modeling to stem cell behavior. While these are obviously not the only contexts in which modeling can be leveraged, in some ways they are fundamental for understanding stem cell behavior. The defining feature of stem cells are their ability to differentiate into various lineages which hold a diverse range of protein expression patterns, so we will focus here on two applications centered on understanding multistability.

In the first, the topology of the network is known, and it is a matter of exploring the range of parameters and conditions under which the system behaves in a desired manner. This will be referred to as parameterizing a network. A detailed example of this

type of network analysis can be found in the work of Wu, et al (82). In the case of stem cells, this means exploring the conditions under which the network exhibits multistability and/or will preferentially differentiate to one state over another. In the second application, the topology itself is unknown or not fully known, so one must interrogate a set of potential networks to discover those that could exhibit multistability. We refer to this as exploring network topologies. One can look at the work of Faucon, et al. for a good example of this type of analysis (106). There is a lot of commonality in the methods employed in both applications, so we will look at both simultaneously.

As an important note before beginning, biological networks are highly interconnected and complex systems, making them difficult to study directly due to the great number of confounding factors. Because of this, simplification methods are often used to reduce network complexity. Please see Section 3.4.1 (Reducing network complexity) for a more thorough discussion of this topic.

3.3.2 Formulating a system of ODEs

Gene network models typically begin as a system of ordinary differential equations (ODEs), each equation describing a single sub-cellular population: a specific protein, mRNA, metabolite, etc. While each individual member of these populations may exhibit unique, semi-random behavior, the population at large will behave relatively deterministically. Because of this, equations have been developed to represent the aggregate behavior of molecular populations. Two very important and inter-related

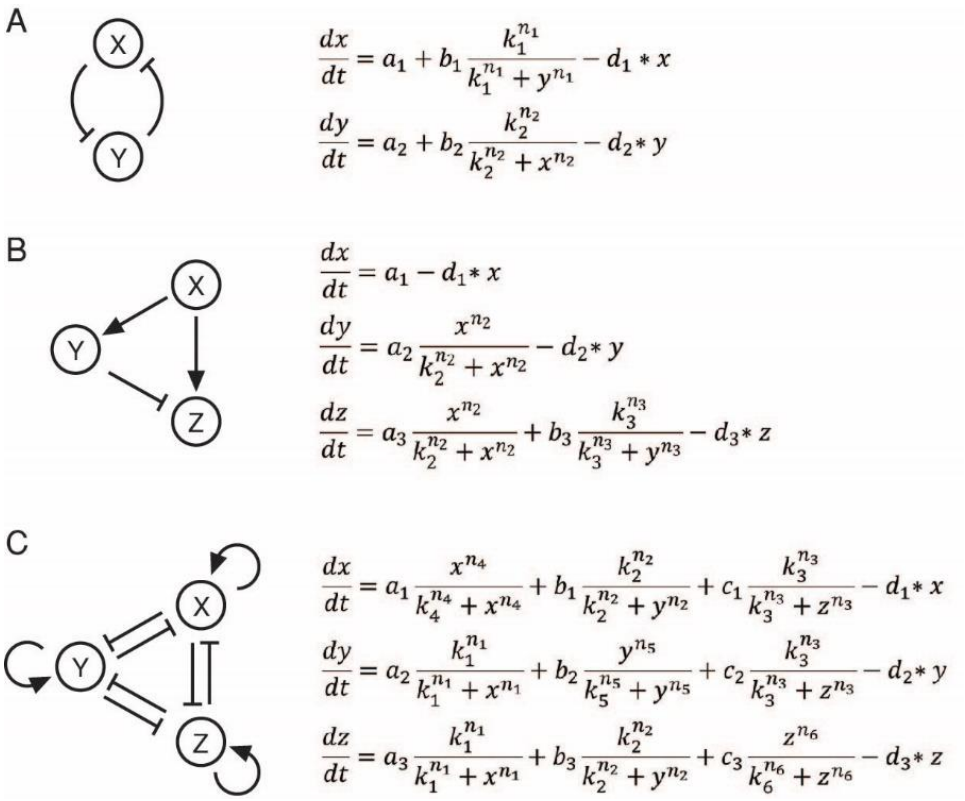

Figure 3-1 Gene network ODE formulation

Network diagrams and potential equations to describe them. Arrow links indicate activation, while barred links indicate repression. **(A)** Mutual inhibition toggle switch. **(B)** Incoherent feed-forward loop. **(C)** FCT with auto-activation and full mutual inhibition.

equations in this respect are the Michaelis-Menten equation which describes enzymatic activity (150), and the Hill equation which captures the non-linear behavior of multi-protein binding dynamics (151). We do not cover derivations of these equations here, but both are worth being familiar with, as they are the most common formulations used extensively in the biomolecular modeling literature (43, 78, 79, 82, 106, 114). With that in mind and being cognizant of the need for simplification discussed in Section 3.4.1 (Reducing network complexity), we will now explore how to formulate ODE models.

In a system of ODEs, several equations are constructed, each representing the fundamental populations that are being studied. Figure 3-1 shows several gene topologies and the related equations that describe their behavior. One may notice that

nearly all equations will have 2 parts: a production term and a degradation term. In a transcriptional regulation pathway, interacting nodes will either activate or repress the production of themselves or another population. Both interactions have basic forms, based on Hill equation kinetics:

$$\frac{dx}{dt} = a * \frac{y^n}{k^n + y^n} \qquad \text{eqn (3-1)}$$

$$\frac{dx}{dt} = a * \frac{k^n}{k^n + y^n} \qquad \text{eqn (3-2)}$$

Here, the change in our population (x) has a maximum expression rate (a) which is modified by the abundance of activators or repressors (y). In activation (eqn. 3-1), the production rate of x asymptotically approaches a as y increases. In repression (eqn. 3-2), production of x asymptotically falls from a to 0 with increasing y. A rate constant (k), sometimes referred to as the "half-max" value, serves as a sort of center point that helps determine the scale of y needed to influence x's expression. Finally, the Hill coefficient (n) determines the linearity of x's response to y, which the steady-state expression of x vs. y becoming increasingly sigmoidal as n increases. A third type of production term – constitutive production – may be included in some circumstances:

$$\frac{dx}{dt} = a \qquad \text{eqn (3-3)}$$

However, unless there is some sort of dynamic interaction with other nodes in the network, once degradation is factored in (see below), constitutively expressed components will end up expressing at a steady state. Therefore, this steady state expression can be substituted for the ODE, lowering the number of equations that must be solved and reducing the computational requirements for running the system of ODEs. Frequently, a constant production term like this will be used in conjunction with a

repression term (eqn. 3-2) to represent leaky repression. In this case, the node produces a small amount of product even when fully repressed.

Populations will also degrade with time, and there are two primary forms of that as well: constant and enzymatic. Typically, constant degradation is sufficient, but networks which produce many transcripts may see those transcripts competing for degradation machinery, in which case the enzymatic model may be more appropriate.

$$\frac{dx}{dt} = -d * x \qquad \text{eqn (3-4)}$$

$$\frac{dx}{dt} = -d * \frac{x}{x+z} \qquad \text{eqn (3-5)}$$

In the case of constant degradation, the degradation rate (*d*) is simply a fraction of the amount of *x* present in the system. In the enzymatic case, however, the total number other things (typically proteins) being degraded in the system (*z*) imposes an upper limit on the rate at which y can be degraded.

3.3.3 Alternate Equation Forms

From these basic forms, equations can be expanded as needed to integrate relevant information into the model. Leaky repressible promoters, for example, may include a low level of constitutive expression and a repression production term (82). An activator that works only in the presence of a drug may be modified by a term to indicate the fraction of functional activator at various dosages of drug (43). As a guideline, however, it is best to construct the simplest model that still captures the behavior or interactions that you care about. Simple models can always be expanded.

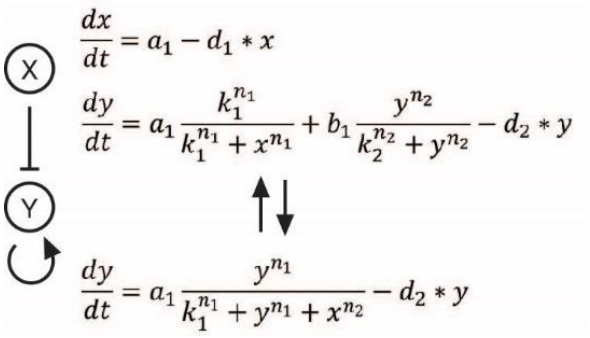

Figure 3-2 Alternate ODE construction

The ODE describing *y*'s expression be constructed in several different ways. The combined effects of repression by *x* and autoactivation could be modeled as either the sum of 2 production terms (top) or as a single term comprising both processes (bottom).

This becomes slightly less straight-forward once one has multiple regulators affecting the same gene. If a gene is both activated and repressed, it may not be obvious whether to combine the production terms additively or to make some sort of hybrid equation. Figure 3-2 shows two formulations of the same network. In practice, the differences in outputs between various formulations of the same network are often insignificant. Because modeling is an abstraction of the incredibly complex observed cellular behavior, there is no single accepted "correct" way to model a gene network. For parameterization purposes, provided that the ODEs produce the experimentally observed behavior, they are likely okay. In the case of network exploration, it is often a good idea to experiment with alternate equation formulations on a subset of your networks, to ensure that the observed behavior is not simply an artifact of the model's form.

3.3.4 ODE Formulation for Network Exploration

For parameterization of a network, once the equation form is determined, model creation is finished. However, for exploring network topologies, additional steps are required. Because of the combinatorial nature of this type of work, automation is a necessity. First, you need to generate a list of all possible networks within the bounds of whatever constraints you've imposed on the system. In the case of Faucon, et al, for example, the

constraint used was fully connected triads (FCTs) (106). To decrease downstream computation, you should identify networks which are rotations or projections of one another. These networks will behave identically, so there is no need to solve more than one of them. Once redundant networks have been eliminated, you will be left with a set of all unique networks for further analysis.

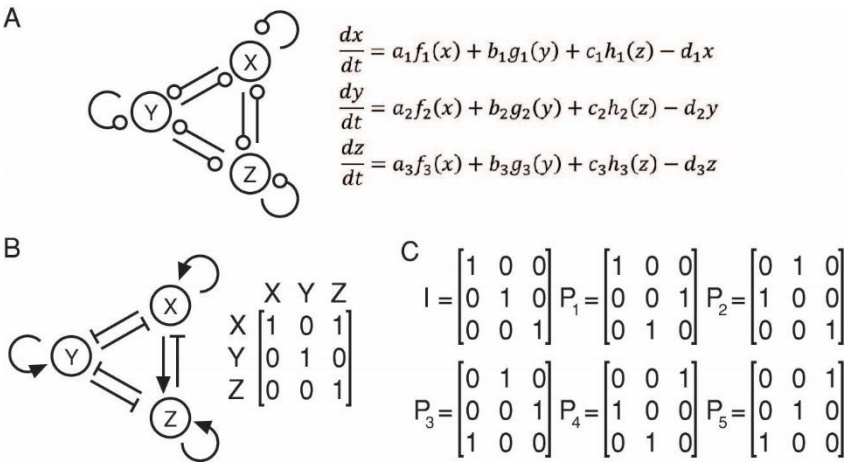

Figure 3-3 Network topology exploration

There are several stages to exploring network topologies. **(A)** A general form of the ODEs that can describe all possible networks is formulated. In the diagram, the circular ends of the regulatory links indicate that they could be either activating or repressing. **(B)** For a specific network, a connection matrix is made. A value of 0 indicates repression, while a 1 is activation. **(C)** Permutation matrixes are used to eliminate redundant networks. Here, all permutation matrices ($P_{1\text{-}5}$) for the FCT topology are shown, with the identity matrix (I) for reference.

To generate all possible networks, first formulate a generic form of your system of ODEs. For a system of fully connected triads, this might be that shown in Figure 3-3A. The parameters $a_{1\text{-}3}$, $b_{1\text{-}3}$, $c_{1\text{-}3}$, and $d_{1\text{-}3}$ will not affect the form of the equations, but the functions $f_{1\text{-}3}$, $g_{1\text{-}3}$, and $h_{1\text{-}3}$ will differ depending on whether they are activating (eqn. 3-1) or repression (eqn. 3-2). If we assume that all nodes must activate or repress (they can't not interact), the system can be represented by a 3x3 connection matrix (Figure 3-3B). Once this is realized, it is a trivial matter to automate the generation all possible matrixes. This approach can be modified easily to fit any topology with any number of interactions.

Examining the network in Figure 3-3B, one can see that if the activation from *x* to *z* were switched to be from *x* to *y,* we are left with a network which functions identically to the original, with only the names of the nodes changed. To eliminate this sort of redundancy, one can utilize linear algebra transformations to check for equivalence. In the fully connected triads example, this involves generating a set of permutation matrices that account for all possible row and column switching (Figure 3-3C). Two connection matrices (M_1 and M_2) are considered equivalent to one another if they satisfy the equation $M_1 = P_i*M_2*P_i'$ (i = 1, 2, 3, 4, 5), where i represents the permutation matrices. If matrices are found to be equivalent, one can safely be dropped to reduce downstream computation. It is worth noting explicitly that, while the algorithm for generating all possible systems of ODEs is relatively straightforward when moving into higher order and non-symmetrical networks, this is not necessarily the case for permutation matrix generation and matrix elimination. Additionally, if nodes are intended to represent specific genes, rather than generic transcriptions factors, network elimination of this type may be entirely inappropriate. In short, when working with more complex networks, thorough knowledge of linear algebra is required to determine an appropriate method for eliminating redundancy.

3.3.5 Deterministic analysis

ODEs return purely deterministic outputs; given the same initial conditions (ICs) and parameters, the output will be the same every time the equations are run. Depending on the context, a model can be used in several ways. First, it may be used to fit experimental data, providing additional information on the underlying regulatory framework. This

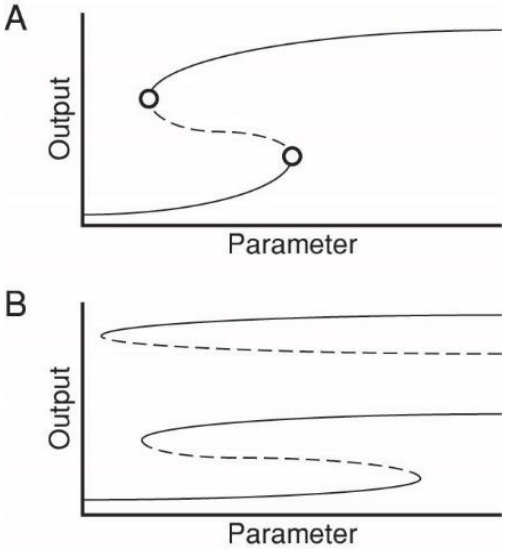

Figure 3-4 Bifurcation analysis

Bifurcation analysis can reveal regions of multistability within parameter space. **(A)** S-shaped hysteresis curve with bifurcation points indicated with circles. These are the parameter values at which the system undergoes a state change. **(B)** For more complex bifurcation analyses, a wide range of ICs should be chosen to avoid missing relevant information. Here, the topmost SSS could be missed entirely if a very large IC is not tested.

may be performed separately from or in conjunction with bifurcation analysis, which reveals multistable behaviors inherent in stem-like cell types. This technique identifies ranges of parameters under which the system exhibits hysteresis: it can hold more than one steady state expression level. The specific state the system holds is determined by the ICs provided.

3.3.6 Bifurcation analysis

At its most basic, bifurcation analysis traces steady states in parameter space. The system of ODEs is run to steady state expression, then one or more parameters are altered slightly, and the new steady state behavior is compared to the previous levels. This is done iteratively, allowing the steady states to be determined for a range of parameter values. In addition to identifying stable steady states, bifurcation analysis uncovers unstable states (which divide the space between stable attractors), a host of unique bifurcation points (specific values at which behavior changes), and more exotic behaviors (like oscillations). To perform bifurcation analysis, a continuation toolbox, like MATCONT is required.

An important factor in running a bifurcation analysis is to set a wide enough range on the parameter under observation. Figure 3-4A shows the hysteresis curve of the simple mutual inhibition switch from Figure 3-1A, made using the information from

bifurcation analysis. This "S" shape is common, indicative of a region of multistability. Within this parameter range, the system can settle into either a high- or low-expressing state, depending on the ICs. In setting the range of parameters to search for this multistable behavior, it is necessary to set it wide enough to at least include both bifurcation points. Because the location of these points is not usually known prior to analysis, setting the limits beyond what you might consider a physiologically relevant range is suggested. This may include the use of negative parameter values which, while not even physiologically attainable, may be necessary for uncovering bifurcation points and thereby tracing a complete hysteresis curve.

In addition to parameter range selection, in more complex systems it is good practice to perform bifurcation analyses from a variety of initial conditions. Because this technique essentially traces a single steady state as it changes with a given parameter, it cannot reveal additional steady states if they are discontinuous with the state being traced. As an example of this, see the hypothetical hysteresis plot in Figure 3-4B. For relatively simple networks, this may be done empirically; however, as network complexity grows, more computationally rigorous methods may be required. These are covered in Section 3.3.8 (Network topology exploration).

3.3.7 Parameter fitting

If experimental data on a specific network is available, it can be used to determine which parameter values best fit the data. For this to be effective, a range of data points, either distributed in time (a time course) or across parameter space (a dose response curve) is needed. Using a nonlinear curve fitting algorithm, such as MATLAB's *lsqcurvefit*, the

model is run with initially random parameter set and compared to the experimental data, generating a measurement of error from the observed values. The parameters are then iteratively altered, and the model rerun, improving the fit and thereby reducing error with each iteration, until it arrives at a local minimum of error. This algorithm is very powerful, but there are several ways its implementation can break down, which we will discuss.

One of the first decisions that must be made is which parameters to analyze. Running an algorithm like this on a large system of ODEs or one with many parameters can be computationally expensive and time consuming. It is therefore often beneficial to fix any parameters whose values can be determined with reasonable certainty. Protein degradation times, for example, are often closely linked to cell division and can, in many contexts, be approximated using the cellular doubling time. One must simply solve for the exponential decay constant (λ) in the exponential decay function $N(t) = N_0 e^{-\lambda t}$. Here were know the doubling time (t), and it is assumed that $N(t)$ is ½ of N_0, as the actual protein decay rate is minor in comparison to the dilution rate due to cell division. Similarly, some parameter values, such as maximum production rates, may be deducible from experimental data. In a simple repressor defined by eqns. (3-2) and (3-4), for example, if the degradation rate is deduced as discussed above, in the absence of a repressor: $a = d * x$. If data are available giving a value for x under these conditions, a can be set with some confidence. Finally, parameters may be able to be inferred from prior work.

Once any parameters which can be inferred have been fixed, then the task at hand is setting bounds on those which remain unknown. If these bounds are set too narrowly,

one risks excluding the best fit for the system. However, if the bounds are set too wide, the best fit may be missed as the algorithm converges to local minima far from the global best fit. Developing an empirical understanding of your system of ODEs is necessary for choosing proper bounds. This comes in part from understanding prior literature on similar cellular processes. Hill coefficients (n) in equations (3-1) and (3-2) are typically <10, as high degrees of nonlinearity tend to make processes unstable. Relatedly, repression or activation coefficients (k) from equations (3-1) and (3-2) are sometimes referred to as "half-max values" because they are related to the amount of activator or repressor required to transition expression to half its maximum value. A range of potential values may be deduced if additional information is known about the system. Once a good initial guess is determined, setting bounds around this guess is trivial. If the analysis is run and one or more parameter values consistently converge to the bounds that you have chosen, this is a good indicator that your bounds should be expanded.

Once all parameters have been set or bounded, the algorithm is run hundreds of times. On each run, bounded parameters' initial values are chosen randomly, and the algorithm finds a set of parameters whose error converges to a local minimum. Unlike discussed in the previous section on bifurcation analysis, ODE ICs should remain fixed across all of these runs, corresponding to whatever ICs were used to generate the underlying data against which the model is being fit. Each run will result in a set of parameters and corresponding error. These can be ranked from lowest to highest error, and a subset of those with the lowest error can be further analyzed to glean additional information on levels and interrelations between different parameters within the network.

If any confusion remains on the difference between random initial conditions and initial parameters, see Section 3.4.2 (Random initial conditions, parameters, and stochasticity).

3.3.8 Network topology exploration

If you are exploring network topologies for multistable behavior, a more computationally expensive approach is required. For this application, access to a distributed computing cluster is necessary. As discussed previously, you must first generate a minimal set of all possible networks that fit your predetermined criteria. Then, each network must be evaluated with a range of parameters and ICs to determine if it exhibits multistable behavior, and if so, how many unique states it can hold. Like in the previous section, this application explores a range of values for several parameters; however, in this case, preset values are applied in combination with a set of preset ICs across the full range of all possible networks. There is no randomness here.

This type of analysis is, by definition, not based on data; rather, it is a more theoretical exploration of topology behaviors. As such, prior discussion of setting accurate parameter bounds and selecting physiologically relevant values is less relevant here than when parameterizing a known network. Here it is more important that parameter values make sense relative to one another, with the assumption that findings would scale to an actual biological system. This altered focus typically leads researchers to select easy to work with ranges for parameters, often utilizing powers of 10, such as

(0.1, 0.3, 1, 3, 10). Due to the highly combinatorial nature of network topology exploration, only a few values can be chosen for each set of parameters, and one should attempt to fix as many parameters as possible to reduce complexity. Faucon et al, for example, reduced their fully connected triad system to 9 parameters, 3 with 3 levels and 6 with 5 levels, but this still left $3^3 * 6^5 = 421{,}875$ parameter sets to check for every IC of every possible network (106).

Similar to how parameter space was divided, initial condition space must also be divided. The range of ICs must be large enough to encompass all possible expression levels and subdivided enough to have a high likelihood of finding all stable steady states. The form of Hill equations makes finding their maximum expression easy: it is simply the value of *a* in equations (3-1) and (3-2). Maximum steady state expression is therefore the sum of all production terms divided by the degradation term. Once this is known for each variable, the range of [0 max] can be subdivided into segments covering the whole of IC space. The number of segments will vary depending on the system being analyzed but should generally be much more than the total number of theoretical roots on the right-hand side of the ODE, which is numerically equivalent to the steady states of the ODE. The goal is to have enough coverage of IC space so that it is likely that at least one IC will converge to each stable steady state (SSS). Around 10 segments per variable is a good starting value, which can be adjusted depending on the system being analyzed.

Finally, for each network, for each parameter set, for each IC, right hand side of the system of ODEs can be solved numerically. In MATLAB, there are several built-in algorithms, such as *fsolve*, for solving nonlinear equations like the ones we have constructed. After solving for all ICs within one parameter set of a given network,

unique roots can be consolidated for further analysis. Unique roots can be entered into the Jacobian matrix and the eigen values calculated. Solutions with all negative eigen values are SSS.

3.3.9 Stochastic analysis

While deterministic analysis is useful for identifying SSS and predicting average population behavior (in many circumstances), it does not recapitulate expression variance inherent in cellular systems. Some processes, like stability switching or random differentiation from a neutral state, cannot be exhibited by a simple ODE model. To explore these sorts of behavior, the model must be modified to introduce stochasticity to the genes' expression. This is frequently done in one of two ways: converting the system of ODEs to a Gillespie algorithm (152) or converting the ODEs into stochastic differential equations (SDEs) with a method like the Langevin equation (153).

In both implementations covered here, the level of noise is correlated with the size of the associated population. A population with a SSS level of 100 will exhibit much more relative variability than one with a SSS level of 10,000. For this reason, it may be necessary to rescale parameters to approximate a physiologically relevant level of gene expression variation. Many common data acquisition methods, such as flow cytometry and qPCR, give outputs which are correlated with, but not identical to, the size of intracellular protein or RNA populations. Fitting parameters to this data may give upscaled parameter values which may subsequently underestimate the role of noise on expression. Therefore, it may be helpful to modify your equations' production terms with a rescaling factor which allows you to easily up- or downscale the simulation

without changing its qualitative behavior. This can be achieved by multiplying *a* and *k* terms in equations (3-1) and (3-2) by a constant value. Adjusting the SSS value to be around 100 is a good starting point for achieving a reasonable amount of expression noise. Of course, this level can be adjusted either to explore the effects of noise or to simulate data in which the noisiness is known.

3.3.10 Gillespie algorithms

The Gillespie algorithm breaks the simulation down into discrete events in which the populations being modeled experience some change: production or degradation of individual proteins or molecules, typically. Each event is assigned a probability based on its associated rate, relative to the rates of all other events in the network. As simulation time progresses, events occur one after another with variable time between each, dependent on the current state of the system. The benefit of this approach is that it is fairly accurately simulates what is actually happening in the intracellular environment by converting all populations to discrete values. However, this algorithm tends to be relatively slow, particularly when running larger gene networks. Methods to improve simulation time have been developed (such as tau-leap methods), but these introduce approximations which may affect the accuracy of the simulation (152). Here we will cover the basics of constructing a Gillespie algorithm, but know that there are many ways that this framework can be modified for different applications.

The algorithm is a simple 4 step process: initialization, random step, update, iterate. Initialization involves defining the system and setting ICs. Then, it will take random sized steps through time, at each step selecting a single event to occur. These

random events will be used to update the state of the network. The new state will then be used to generate a new random step and further update the network state, iterating between taking random steps and updating the network state. At each step, rates of change are calculated, and two random numbers are generated to determine which event happens and at what time, making this a relatively computationally intense process.

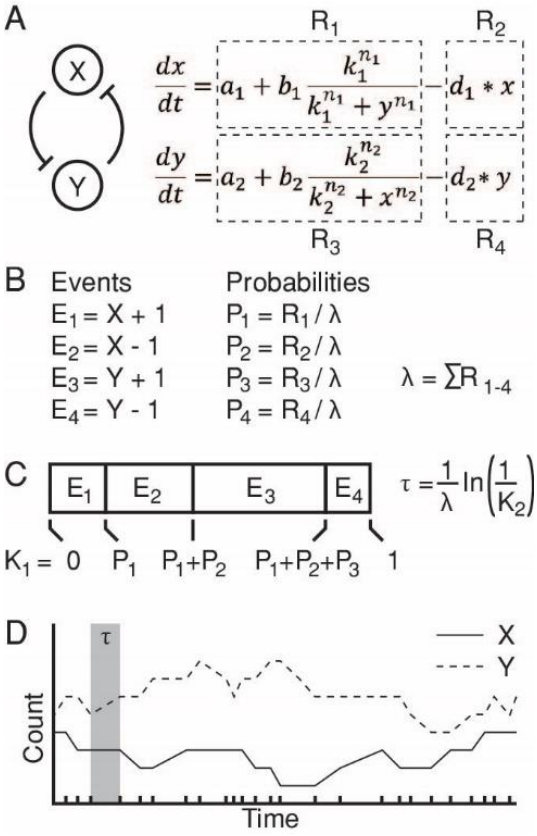

Figure 3-5 The Gillespie algorithm

Implementation of the Gillespie algorithm. **(A)** The ODE system can be deconstructed into production and degradation terms. **(B)** Discrete events (E_{1-4}), with related rates (R_{1-4}), and probabilities (P_{1-4}) are defined. At each iteration of the algorithm, rates and probabilities are recalculated from the system's current state. **(C)** Random numbers (K_1 and K_2) are used to determine which event happens and the time since the previous event (τ). **(D)** Sample output of several iterations of the algorithm. Notice that at each τ only one variable changes, and only every by ±1.

It is easy to convert a system of ODEs into a Gillespie algorithm. Each equation must be split into the component parts which cause a single change (event) to a population. In a simple transcriptional repressor network like that shown in Figure 3-5A, the unique events which can occur are: production of protein *x*, degradation of protein *x*, production of protein *y*, and degradation of protein *y*. The rates at which these events occur – and by extension the probabilities that they will happen at each time point – are obtained directly from deconstructing the ODEs into their production and degradation terms (Figure 3-5B).

Based on their rates relative to one another, each event can be assigned a probability. An event's probability of occurring is $P_i = (R_i / \lambda)$, where R_i is the rate of the individual event, and λ is the sum of all events' rates: $\lambda = \Sigma R_{1\text{-}n}$. When calculated for all events, $\Sigma P_{1\text{-}n} = 1$. At each iteration of the algorithm, a random number, K_1, chosen from the uniform distribution [0, 1] can be used to randomly select a single event to occur. The timing between events, usually referred to as τ, is governed by the exponential function: $\tau = (1 / \lambda) * \ln(1 / K_2)$, where λ is the sum of the rates of all events, as above, and K_2 is a random number from the uniform distribution [0, 1]. Figure 3-5C shows a visualization of how the random numbers, K_1 and K_2, are used to generate each time step, and Figure 3-5D shows a sample output of the mutual inhibition toggle.

Production and degradation of intracellular populations is not the only thing that can be modeled with a Gillespie algorithm. Protein multimerization, for example, could be described as a series of discrete steps from monomers to dimers to tetramers, with reversals of each of these interactions and degradation events. For more complex interactions, it should be noted that a single event might see multiple populations changing in relation to one another. For example, in a dimerization event, 2 monomers are lost and 1 dimer is gained.

3.3.11 Langevin equations

A similar outcome can be obtained by modifying the ODEs into chemical Langevin equations, which introduces some randomness with each step forward in time in the ODE solver. Relative to the Gillespie algorithm, the use of a chemical Langevin equation is much faster to run; however, the mathematical derivation is much more complex and not

as easy to grasp intuitively. To avoid abuse while implementing this method, a thorough investigation of primary literature is suggested (153). For our purposes, at each iteration of the ODE, all chemical species are updated with the formula:

$$N(t+\Delta t) = N(t) + \Delta t A\big(N(t)\big) + \alpha * S\sqrt{\Delta t(F\big(N(t)\big) + B\big(N(t)\big))}z(t) \qquad \text{eqn (6)}$$

The terms before α are standard for ODE solvers, but α and everything that follows have been added to introduce noise. Here, *N(t)* is the abundance of the population, *A* is the righthand side of the ODE, and *F* and *B* are the forward (production) and backward (degradation) reaction terms in each equation. α is a scalar term for tuning the level of noise. *z(t)* and *S* are the standard Normal variables and stochiometric matrix, respectively, of the biochemical reactions. For a fuller description of these terms, as well as software tools to implement chemical Langevin equations, see work by Adalsteinsson, McMillen, and Elston (153).

3.3.12 Stochastic analysis

With stochasticity integrated into the model, the methods for exploring the network to its fullest extent are expanded. As this is a relatively small field of research, however, there are few definitive methods for exploring stochasticity in biochemical models. Because stochastic simulations have noise, each run of the simulation will be different from every other run. For this reason, analysis of stochastic simulations can be time consuming, as the model must be run enough times to generate a characteristic picture of the effect that noise is having on expression. To account for low probability behavior, this can mean running the simulation tens, hundreds, or thousands of times.

Stochastic simulation can be used to explore the pull that different SSS exert on the system. While bifurcation analysis of an ODE model can separate state space into different attractor regions, stochastic simulations can quantify how likely those regions are to maintain stability in the face of noise. Faucon, et al, used an approach like this quantify the strength of different attractor states and measure the mean time to transition between states under variable noise conditions (106). As one example, they show that, increasing noise can cause some attractor nodes to transition from highly stable to nearly incapable of holding steady expression. Additionally, because expression noise can cause networks to transition between different SSS expression profiles, stochastic simulations can reveal preferred transition paths between SSS attractors. As computational power increases, allowing modeling of ever more complex networks, stochastic analysis of this sort can help reveal the mechanisms underlying complex differentiation pathways.

3.4 NOTES

3.4.1 Reducing network complexity

Mathematical modeling of complex cellular behavior is still a relatively new discipline, limited by both our computational power and understanding of the network topologies underlying observed cellular behavior. For this reason, we often use abstracted or synthetic networks as stand-ins for the actual networks which regulate stem cells. For example, instead of trying to model and understand the full mechanism underlying pluripotency, the network of dozens of interconnected genes is frequently condensed down to its most basic form: the Oct4, Sox2, Nanog triad (106). This sort of abstraction

can reduce the system to one that is simple enough to study and understand, while simultaneously being accurate enough to still represent the underlying regulatory complexities. Inevitably, some of the details are lost, and if one is not careful you may end up eliminating the expected behavior from your model. Nonetheless, it is common practice to reduce networks to as few moving parts as possible.

Alternatively, researchers can build gene networks from the ground up to decouple the behavior of interest from the rest of the cellular regulatory machinery. In these synthetic networks, fundamental interactions between genetic components can be more accurately observed and simulated via modeling, since they have been separated from many of the confounding factors which make natural networks so difficult to study. This is one of the fundamental differences between synthetic biology and more traditional disciplines attempting to apply mathematical methods to biology.

In addition to how the system is built and/or conceptualized, when one begins building models, it quickly becomes apparent that the model could be as complex as you could possibly make it. For a common process like transcription/translation, there are endless steps that could be added to the model: transcription factor interactions with the DNA, polymerase recognition of the promoter, assembly of the polymerase complex, transcriptional starting and pausing, incomplete transcription, termination dynamics, RNA capping and polyadenylation, nuclear export, transport through the cytoplasm to the ribosome, tRNA recruitment, and on and on. The list could be expanded nearly infinitely. Running mathematical models can be computationally expensive, and biology is endlessly complicated. If too many details are included, model run times can

skyrocket, slowing down analysis considerably. It is therefore necessary to reduce model complexity while maintaining the fundamental behavior that you are trying to simulate.

3.4.2 Random initial conditions, parameters, and stochasticity

To ensure that there is no confusion, we would like to talk briefly about the different types of randomness employed in this work. Random initial conditions refers to randomizing the starting values of the variables (each equation in the system of ODEs represents one variable). This may be used in some network exploration algorithms, but in most cases ICs are defined rationally based on experimental constraints or to cover all of potential IC space. Alternatively, parameters are all non-variable terms in the ODE. These can be randomized for curve fitting algorithms, which will then iteratively adjust the parameter values to arrive at a local best fit for a model against a set of experimental data. Finally, stochasticity refers to randomness in gene expression, introduced into the model either by converting it to a discreet Gillespie algorithm or through implementation of the Langevin equation. In both the Gillespie and Langevin cases, parameter values initial conditions remain fixed, but the changes in variables' expression as the simulation runs are subject to some randomness.

4 FLUORESCENT GUIDE RNAS FACILITATE DEVELOPMENT OF LAYERED POL II DRIVEN CRISPR CIRCUITS

This chapter was prepared in collaboration with Samira Kiani, M.D. and Xiao Wang, Ph.D. and is under consideration for publication.

4.1 INTRODUCTION

Clustered regularly-interspersed short palindromic repeats (CRISPR) technology has become a prime candidate for synthetic transcriptional regulation and creation of complex genetic networks due to its programmablity, ease of design, and modularity (56, 154, 155). Originally a bacterial immune system, engineered CRISPR is composed of two parts: the protein Cas9 and a guide RNA (gRNA). The catalytically dead Cas9s (dCas9) have been used for transcriptional repression or activation in various organisms (45, 56, 156). Because dCas9 can be directed to nearly any region of DNA by changing the sequence of the gRNA, this technology allows for rapid construction of large libraries of activators and repressors which can act orthogonally to one another. Simple design and implementation allows dCas9-based circuits to fill the need for large libraries of components for network construction.

Expression of gRNA from RNA Polymerase type II (Pol II) based promoters, enables generation of layered CRISPR-based genetic networks in which gRNAs act both as input and outputs of the circuits (74, 157). Such circuits will have tremendous value for step-wise or sequential modification of cell-fate or function in cases such as stem cell differentiation or tissue regenerative therapies. Through various RNA editing methods

which release gRNAs from primary RNA transcripts, the expression from Pol II promoters and composability problem of CRISPR-based components can be addressed (158–160). However, mRNA production rates from Pol II promoters tend to be much lower than from Pol III promoters, which could lead to overall low efficiency or even nonfunctional Pol II driven circuits in mammalian cells (74, 161). There remains a need for methods to more accurately test, quantify, and optimize these systems by directly analyzing gRNA levels and assessing their impacts on circuit functionality.

gRNA levels in CRISPR circuits have previously been measured indirectly through evaluation of circuit output or a fluorescent protein co-expressed with gRNA (74, 157). Development of RNA binding fluorescent probes (162), fluorescent protein binding RNA aptamers (163), and fluorophore-binding RNA aptamers (164, 165) have recently allowed for visualization of RNA and gRNA. It has been shown that modification of the gRNA transcript can be accomplished without destroying gRNA function, allowing insertion of fluorophore-binding RNA aptamers such as Spinach or Broccolli (164–168). Although these methods demonstrated feasibility to visualize gRNAs, their application for studying CRISPR-based synthetic gene circuits has not been explored.

Here we set out to address the need for more predictable and reliable Pol II driven CRISPR circuits by employing a gRNA modified to include the green fluorescent aptamer Broccoli (165). We demonstrate this strategy can be used to analyze and model circuit behavior. Using gRNA constructs expressed from Pol II promoters, we show that

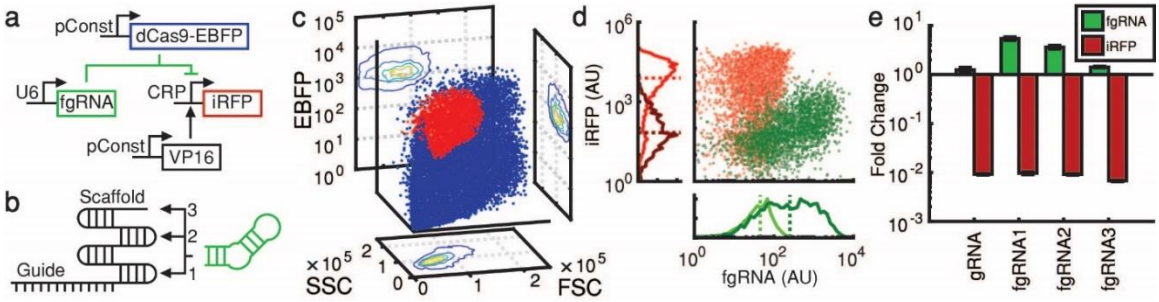

Figure 4-1 Fluorescent guide RNA construction

Fluorescent guide RNAs are visible without loss of function. **(a)** Diagram of the repressor circuit used to test Pol III driven fgRNAs. The circuit was tested with and without the inclusion of the fgRNA component to calculate fold change in fgRNA and iRFP expression. **(b)** Guide RNA diagram showing locations into which the broccoli aptamer was inserted into the gRNA scaffold. DFHBI-1T binds to the broccoli aptamer and fluoresces green when excited. **(c)** Scatter plot of flow cytometry data showing cell size (front scatter; FSC), granularity (side scatter; SSC), and transfection level (EBFP) with the population gated via GMM clustering shown in red. Contour plots of the gated population are shown on the faces. **(d)** A representative repression experiment showing expression of fgRNA and iRFP. The scatterplot shows fluorescence levels in the absence (red) or presence (green) of fgRNA. Fluorescence channels are displayed independently as histograms on the axes, in the absence (light) or presence (dark) of fgRNA. Median values are indicated with dashed lines. To calculate fold change, medians of the repressed sample were divided by the medians of the unrepressed sample. **(e)** Fold changes of three fgRNA variants compared to a non-fluorescent gRNA control. Each bar is the mean of 4 flow cytometry replicates' medians ± SD.

the visualization and analysis of circuit components can be leveraged to improve the function of a layered CRISPR-based circuit composed of only Pol II driven gRNAs.

4.2 RESULTS

4.2.1 Engineering a Fluorescent Guide RNA

To facilitate construction of diverse CRISPR based circuits, we first developed a fluorescing gRNA construct (fgRNA) without interfering with its downstream functionality. Towards this goal, various constructs were placed in a repressor circuit (Figure 4-1A), in which the infrared fluorescent protein (iRFP) gene was repressed by gRNA complexed with dCas9 fused to blue fluorescence protein (dCas9-EBFP). Previous studies suggest three candidate locations in a gRNA structure for additional sequences which do not disrupt guide function: the tetraloop, the second loop, and the 3' end (163)

(Figure 4-1B). The broccoli sequence (165) with a short hairpinning linker was inserted into each of these locations (see Table 4-1 for sequences).

Circuits were transfected into HEK293FT cells and assessed via flow cytometry after 72hrs. We first employed a 3 dimensional (3D) gating scheme, utilizing a gaussian mixed model (GMM; see Online Methods for full description), to identify cells of interest using three channels: front scatter (FSC), side scatter (SSC), and blue fluorescence (EBFP, representing dCas9-EBFP). This allowed us to accurately separate cells from debris and choose a subset of cells which were both well-transfected (high blue) and of moderate size (Figure 4-1C). Selecting cells of moderate size allowed us to reduce population variability which might influence the analysis, as cell size is highly correlated with protein production (169). We then calculated median green and infrared fluorescent intensity in each circuit to analyze the fold change of both fgRNA and iRFP, a metric which we employ throughout this research (Figure 4-1D-E).

Flow cytometry-based analysis reveals that repression of iRFP is strong across all three fgRNAs, with no significant difference between their effectiveness (Figure 4-1E). Broccoli fluorescence is strongest in fgRNA1 (tetraloop broccoli), with decreasing brightness in fgRNAs 2 (second loop broccoli) and 3 (3' tail broccoli). Due to its superior brightness, fgRNA1 is used in all further experiments, and all fgRNAs referenced hereafter are fgRNA1.

4.2.2 Circuit Optimization Improves Network Function

To characterize impacts of component abundances on circuit performances, we then generated dose response curves for each of the components within the circuit: fgRNA,

	Name	Sequence
fgRNAs	fgRNA1	NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGCGCAGACGGTCGGGTCCAGATATTCGTATCTG TCGAGTAGAGTGTGGGCTGCGCTAGCAAGTTCAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGTCGGTGC
	fgRNA2	NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGAAATAGCAAGTTCAAATAAGGCTAGTCCGTTA TCAACTTGCGCAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTGCGCAAG TGGCACCGAGTCGGTGC
	fgRNA3	NNNNNNNNNNNNNNNNNNNNGTTTGAGAGCTAGAAATAGCAAGTTCAAATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGGCACCGAGTCGGTGCGCGCAGACGGTCGGGTCCAGATATTCGTATCTGT CGAGTAGAGTGTGGGCTGCGC
		ORANGE: Target sequence GREEN: Broccoli
Editing Motifs	RGR	GAGGTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCNNNNNNNNNNGGCCGGCATGGT CCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC
	CGC	ATGTTCACCTATCTACTACCCGTTCACTGCCGTATAGGCAGNNNNNNNNNNGTTCACTGCCGTAT AGGCAGATGCCCGGAGATTATGTAGG
	TGT	AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTC CCGGCTGGTGCANNNNNNNNNNAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCAC GGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA
		ORANGE: fgRNA sequence LIGHT BLUE: Editing sequence (separated from fgRNA after cleavage) DARK BLUE: Editing sequence (connected to fgRNA after cleavage)
Promoters	U6	AAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGAC GTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCAT ATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAA CACCG
	TRE	CGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAA CGATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTACTCCCTATCAGTGAT AGAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTATCCCTATC AGTGATAGAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGGTAGGC GTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCG
	CRPa	CAGATCTCATGTGATTACGCCAAGCTACGGGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTC CGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTTCTGTCCTCCGAGCGGA GACTCTAGAGAATTCTAGGCGTGTACGGTGGGAGGCCTATATAATACCTCATCAGGAACATGTTG GTCGTTTAGTGAACCGTCAGATCGCC
	CRPb	CATGTGATTACGCCAAGCTACGGGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTTCTGTCCTCCGAGCGGAGACTCTA GAGAATTCTAGGCGTGTACGGTGGGAGGCCTATATAATATAGAACCGATCCTCCCATTGGTCGTT TAGTGAACCGTCAGATCGCCTATAGAACCGATCCTCCCAT
		BLUE: Operator Sites (rtTA or Gal4-VP16) RED: TATA Box ORANGE: gRNA Binding Site
Terminators	SUP4	TTTTTTTGTTTTTTATGTCT
	SV40	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCT G
	mPA	CTAGTAATAAAGGATCCTTTATCTTCATTGGATCCGTGTGTTGGTTTTTTGTGTGCGGCCCGTCT AGACC

Table 4-1 Synthetic DNA components
DNA components used in fgRNA plasmid construction, organized by component type: fgRNA, editing motifs, promoters, terminators. In each section, nucleotides are color-coded to indicate important functional regions.

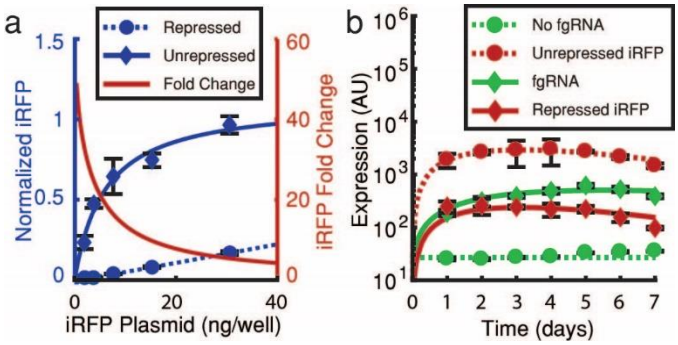

Figure 4-2 Pol III fgRNA expression

fgRNAs expressed from Pol III a promoter have a predictable response. **(a)** Normalized dose response curves of iRFP plasmid (blue), while unrepressed by fgRNA (solid line, diamonds) and while repressed (dashed line, circles). Curves were fit with an exponential function and divided to determine total fold change (red), indicating that a lower amount of plasmid leads to greater differential expression. **(b)** Flow cytometry time course (points) and model fitting (lines) of the U6-fgRNA repressor circuit. fgRNA (green) and iRFP (red) expression are tracked while fgRNA is either expressed (solid lines, diamonds) or absent (dashed lines, circles). Data in all panels are the mean of 4 flow cytometry replicates' medians ± SD.

dCas9, iRFP, the Gal4-VP16 regulator, as well as the total amount of DNA used in the transfection protocol (Figure 4-2A and Figure 4-3). As expected, both Gal4-VP16 and iRFP display decreasing repressibility, showing larger fold changes at lower concentrations. More precisely, when titrating the iRFP plasmid, fold change drops from nearly 60x to <10x as concentration increases

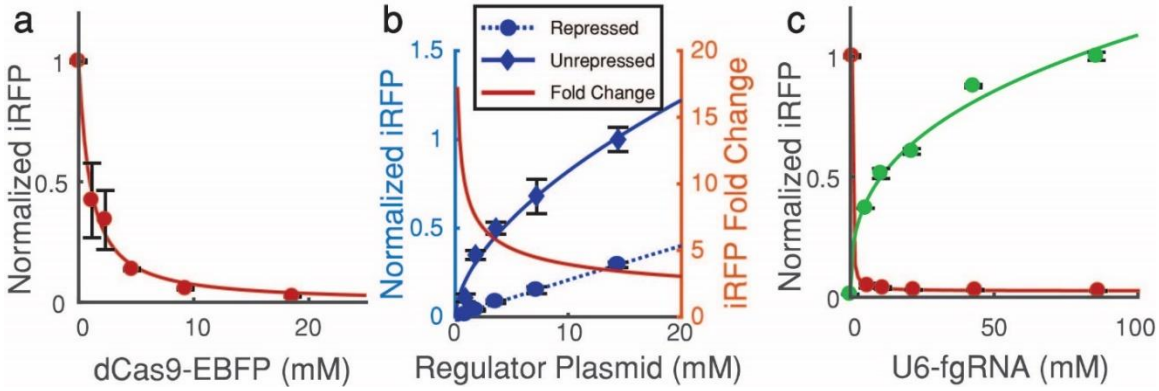

Figure 4-3 Pol III repressor optimization

Optimization of circuit components for U6-driven repression **(a)** Normalized dose response curve showing iRFP expression with increasing dCas9-BFP concentration, experimental data (circles) and exponential fit (line). **(b)** iRFP response to increasing concentrations of regulator plasmid. Experimental data of repressed (blue circles) and unrepressed (blue diamonds) were used to generate fits of repressed (blue dashed line) and unrepressed (blue solid line) iRFP expression. This was then used to calculate fold change (red line). **(c)** Dose response curve showing fgRNA (green) and iRFP (red) expression with increasing concentration of fgRNA plasmid. Exponential fits (lines) are generated from experimental data (circles). All concentrations are the amount of plasmid per 24-well plate well in a total of 600ng DNA. Data are the mean of 4 flow cytometry replicates' medians ± SD.

(Figure 4-2A). Both fgRNA and dCas9-EBFP responses saturate around 20 mM (Figure 4-3). Based on the dose response experiments, plasmid concentrations were adjusted to minimal levels (Figure 4-4, Table 4-2). This resulted in circuits with much greater iRFP dynamic range despite utilizing smaller amounts of fgRNA repressor, while also decreasing the metabolic loads on the cells.

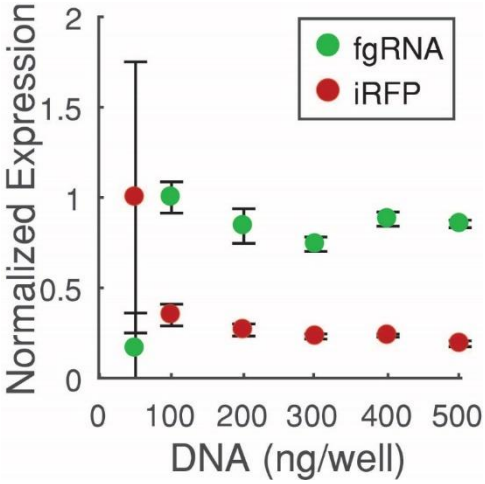

Figure 4-4 Transfection DNA optimization

Optimization of relevant transfection factors Normalized dose response curve showing fgRNA (green) and iRFP (red) expression of a minimal circuit with increasing amounts of total DNA per transfection. Data are the mean of 4 flow cytometry replicates' medians ± SD.

We then analyzed the expression of the optimized circuit over time to quantify the underlying behavior of synthetic CRISPR-based repressors in mammalian cells (Figure 4-2B). Paired with these experiments, we developed a system of ordinary differential equations (ODEs) to model the network behavior (see Online Methods for model details

	Initial Screening		Time Course		Two-Tier
	U6	TRE	U6	TRE	Cascade
dCas9-EBFP	75	75	80	80	80
Regulator	50	50	5	5	5
Csy4	~	5	~	5	5
iRFP	20	20	5	5	5
fgRNA(a)	100	100	20	80	160
fgRNA(b)	~	~	~	~	80
Empty Plasmid	to 600ng		to 300ng		to 400ng

Table 4-2 Transfection masses

Masses of plasmids used in transfections. "Initial Screening" plasmid masses were used for experiments shown in Figure 4-1 and optimization experiments in Figures 4-2A, 4-3C, 4-7, 4-8, 4-9, 4-12, and 4-13. For optimization experiments, the shown values were used except for the plasmid which was being optimized, which correspondeds to the values on the x-axis. "Time Course" plasmid masses were used for experiments shown in Figure 4-2B and Figure 4-7D. "Two-Tier Cascade" plasmid masses were used for experiments shown in Figure 4-14.

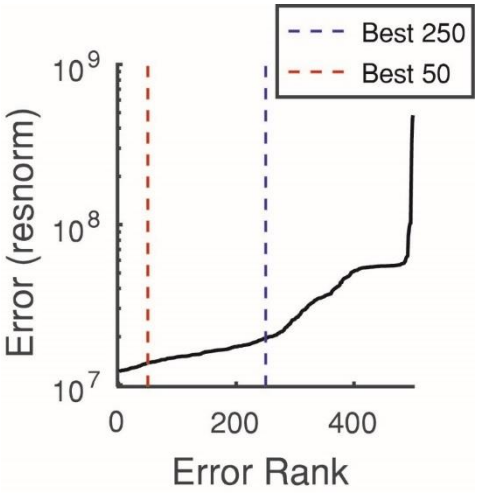

Figure 4-5 Pol III model rank fitting

Pol III time course rank fitting of model outputs generated on the Pol III repressor time course. 500 fits were ranked by their total error (residual norm) and the 50 (red) and 250 (blue) with the lowest error were selected for further analysis.

and derivation). The model consists of three ODEs representing overall plasmid levels, fgRNA, and iRFP abundances. Using these three equations, we demonstrate the relationship between fgRNA expression and iRFP regulation and scanned possible parameters governing this relationship (Figure 4-5 & 4-6). Analysis of fitted parameters indeed verified a few intuitions. For example, plasmid degradation (pDeg) and fgRNA degradation (fDeg) are inversely correlated (Figure 4-6), indicating a strict requirement for fgRNA abundance given specific dynamics. Therefore, decrease of one parameter (pDeg) needs to be compensated by the increase of the other (fDeg). Alternately, positive correlations, such as those between iRFP production (rMax) and iRFP degradation (rDeg) illustrate a need for a ratio between certain paired parameters. Moreover, the analysis reveals that fgRNA production (fMax) has a narrow distribution, suggesting it as the most critical property to tune to achieve desired dynamic behaviors. Finally, this analysis also shows that the non-linearity coefficient (b) is very close to 1. This lack of nonlinearity indicats weak cooperativity between fgRNA, Cas9, and DNA. This has ramifications for building circuits that demand nonlinearity for its function, such as noise reduction or multistability (18, 82, 149, 170).

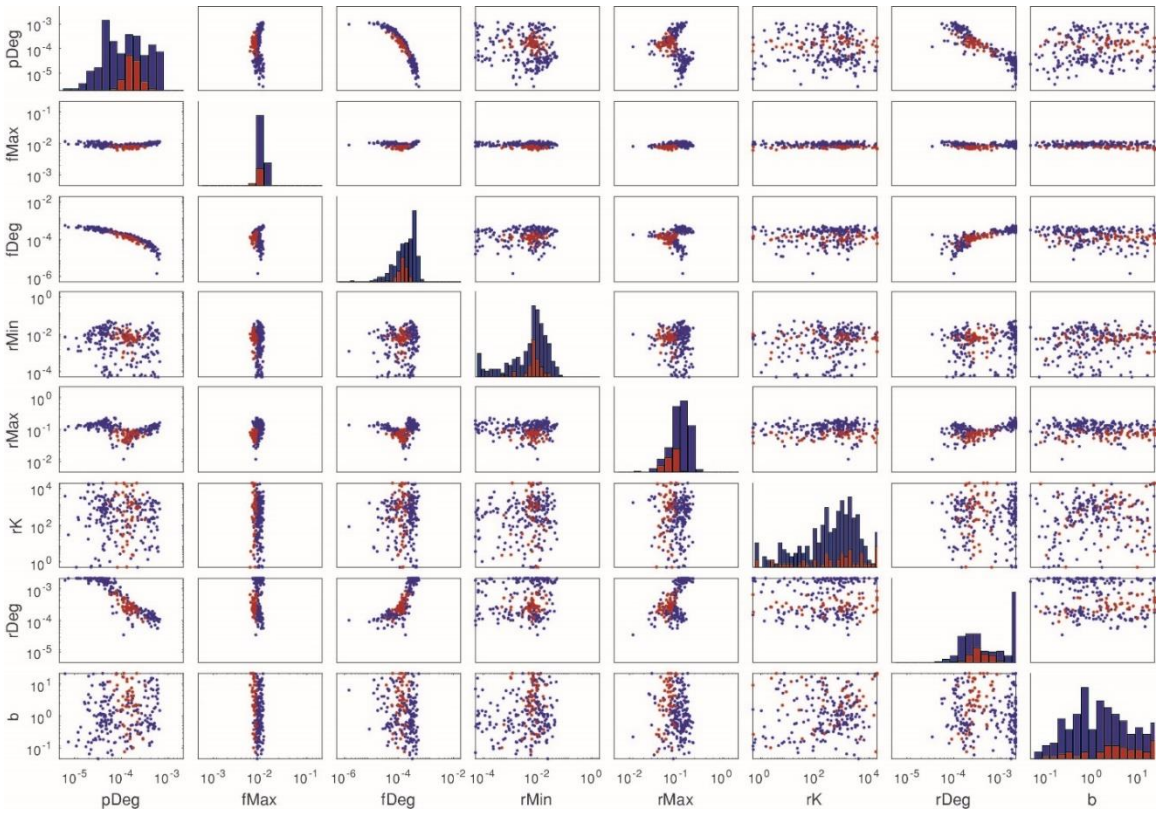

Figure 4-6 Pol III parameter fitting

Pol III time course parameter analysis of the lowest error fittings selected in Figure 4-5. On the diagonal, histogram distributions of each individual parameter are show. Scatter plots of each parameter against the others are shown in the other panels.

4.2.3 Analysis of CRISPR circuits carrying Pol II driven fgRNAs

We next utilized inferred information on engineered fgRNA dynamics to develop Pol II driven fgRNA production (Figure 4-7A). We evaluated three previously published RNA editing techniques to enable gRNA expression from Pol II promoters. The ribozyme-guide-ribozyme motif (RGR) is an fgRNA flanked by self-cleaving RNA sequences – a Hammerhead (HH) ribozyme on the 5' end and herpes delta virus (HDV) ribozyme on the 3' end – that excise the fgRNA shortly after transcription (158). The fgRNAs flanked by Csy4 editing sites (CGC) require exogenous expression of the Csy4 protein, which recognizes and cleaves a 20 nt hairpinning RNA sequencing inserted up- and downstream of the fgRNA (159). Because Csy4 cleaves on the 3' end of the hairpin, this method

leaves one of the hairpins attached to the tail of the fgRNA transcript. The fgRNA flanked by functional tRNA sequences (TGT) utilizes endogenous tRNA editing proteins RNase P and RNase Z to cut around tRNA sequences placed up- and downstream of the fgRNA, leaving a 1 nt addition to the 5' end of the fgRNA and 6 nt on the 3' end (160).

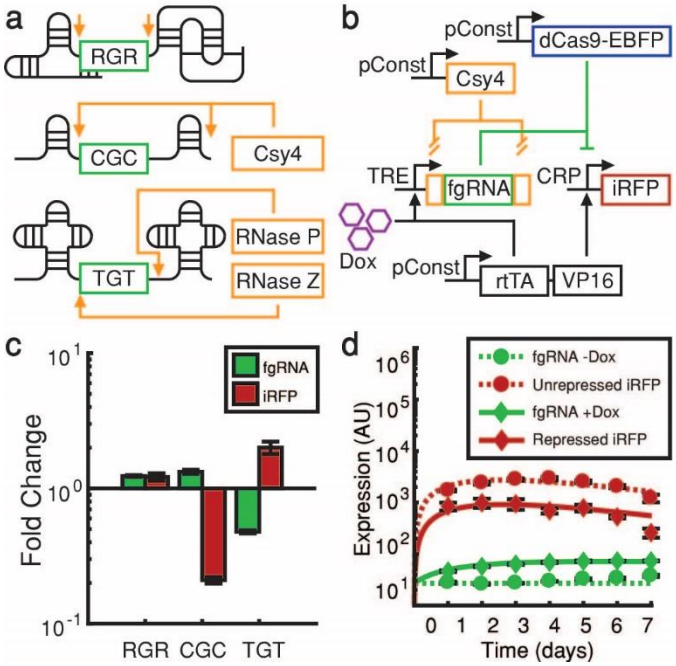

Figure 4-7 Pol II fgRNA expression

Pol II expression of fgRNAs show predictable dynamics. **(a)** Schematics for the 3 editing techniques employed in this work: self-cleaving with the hammerhead (HH) and herpes delta virus (HDV) ribozymes (RGR), targeting and cleavage of 20bp hairpins by the exogenous Csy4 protein (CGC), and excision of the guide by endogenous tRNA cleaving proteins RNase P and RNase Z (TGT). **(b)** Diagram of the doxycycline-inducible repressor circuit where the fgRNA is driven by a Pol II promoter. In this instance, Csy4 is shown editing the fgRNA transcript, though this component can be replaced with any of the editing methods shown in (a). **(c)** Fold change of the three editing methods when the fgRNA transcript is expressed from a Pol II (TRE) promoter. When expressed from a Pol II promoter, only the CGC construct shows repressive activity. **(d)** Flow cytometry time course (points) and model fitting (lines) of TRE-CGC repressor circuit. fgRNA (green) and iRFP (red) expression are tracked while fgRNA is either induced with dox (solid lines, diamonds) or without dox (dashed lines, circles). Data in all panels are the mean of 4 flow cytometry replicates' medians ± SD.

The circuits used to test these three techniques are designed to utilize tetracycline response element (TRE), a well characterized and widly used inducible Pol II promoter, for fgRNA expression (Figure 4-7B, only CGC method is shown for illustration). The TRE promoter has previously been shown to be a strong promoter for protein and gRNA expression in synthetic circuits (73, 74, 171). It requires the inclusion of a reverse tetracycline trans-activator (rtTA) protein which, when in the presence of doxycycline (Dox), activates expression of the fgRNA-containing transcript.

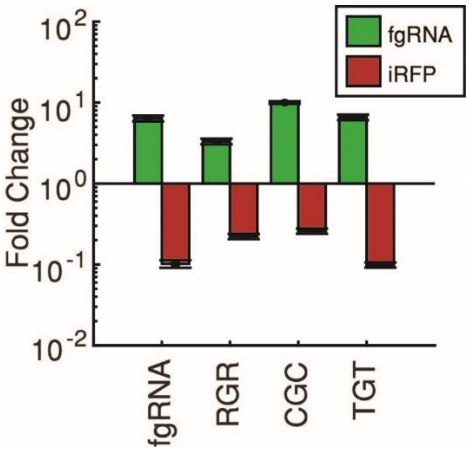

Figure 4-8 Editing method comparison

Editing sequence effect on Pol III expression When pol III transcripts are modified to include editing sequences, expression and downstream repressiveness is affected. RGR shows a decrease in both brightness and repression, while only CGC repression is affected. TGT editing does not appear to have any effect on pol III expression or gRNA repression. Data are the mean of 4 flow cytometry replicates' medians ± SD.

First, each of the three editing methods was tested using a Pol III (U6) promoter. It can be seen in Figure 4-8 that all three editing methods had no significant effects on transcriptional efficiency. Pol II (TRE) promoter driven versions were then tested to characterize their Pol II expression and inducibility. Figure 4-7C shows that Pol II expression of CGC was both visible and caused downstream repression. RGR showed little expression, while TGT lacked inducibility, expressing at a high level regardless of Dox concentration. Therefore, CGC editing was used for further parameterization experiments.

With the transition to Pol II expression, dose response curves were generated to optimize the concentrations of Dox, CGC, and Csy4 components (Figure 4-9) for maximal repression. CGC effectiveness was shown to begin saturating around 75 mM. It was also observed that Csy4 is a very efficient editor, reaching peak effectiveness at 1.5-2 mM. Dox was most effective at a relatively high concentration (2 µg/mL); however, increasing induction too far beyond this point resulted in cell sickness and network dysfunction. These experiments led us to select component concentrations yielding stronger output dynamic range (Table 4-2).

We then ran a time course to observe the direct relation between fgRNA and iRFP expression, fitting the results to our ODE model (Figure 4-7D) and quantifying the

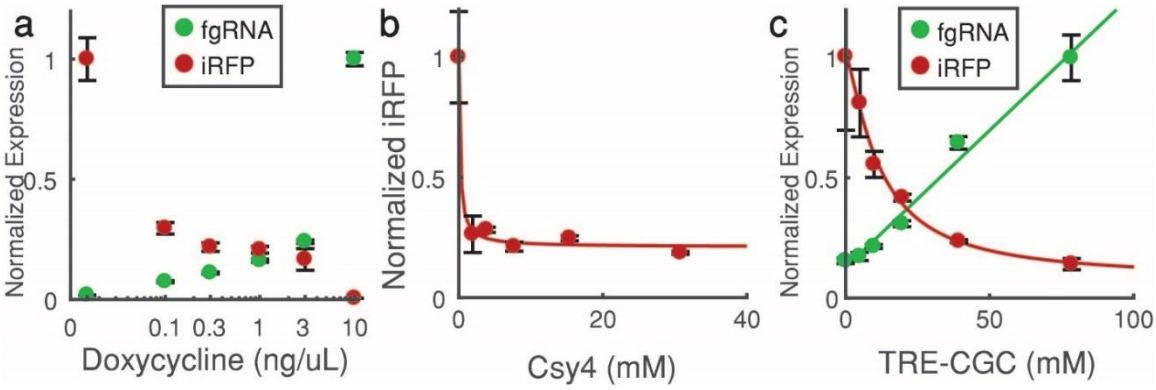

Figure 4-9 Pol II repressor optimization

Optimization of circuit components for TRE-driven repression **(a)** Inducible pol II circuit fgRNA (green) and iRFP (red) response to increasing concentrations of doxycycline. **(b)** Repressed iRFP expression (red) with increasing concentrations of Csy4 plasmid per well, experimental data (circles) and exponential fit (line). **(c)** Dose response of fgRNA (green) and iRFP expression (red) in a doxycycline-induced CGC repressor circuit, with increasing amounts of CGC plasmid. Experimental data (circles) were used to calculate fits (lines). Data are the mean of 4 flow cytometry replicates' medians ± SD.

underlying parameters determining observed behaviors (Figure 4-10 & 4-11). For this application, the ODE model was expanded to include a fourth equation representing the mRNA transcript produced by the Pol II promoter. This transcript is then edited into the functional fgRNA which binds dCas9 and inhibits iRFP production (see Online Methods for details). Fit values of parameters shared between the Pol II and Pol III circuits – pDeg, fDeg, rMin, rMax, repression coefficient (rK), rDeg, b – are similar between experiments, verifying the model's applicability to both scenarios. Examing parameters for both models allows a quantitative comparison of the promoters used, revealing that the production rate (mMax) from TRE is roughly 10-100x weaker than that of U6 (fMax). We also observed that

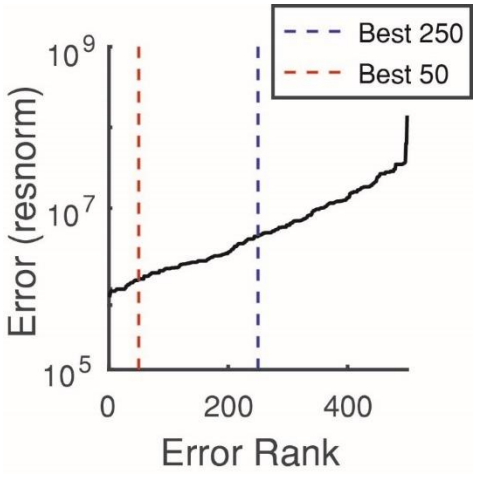

Figure 4-10 Pol II model rank fitting

Pol II time course rank fitting Rank fitting of model outputs generated on the Pol II repressor time course. 500 fits were ranked by their total error (residual norm) and the 30 (red) and 150 (blue) with the lowest error were selected for further analysis.

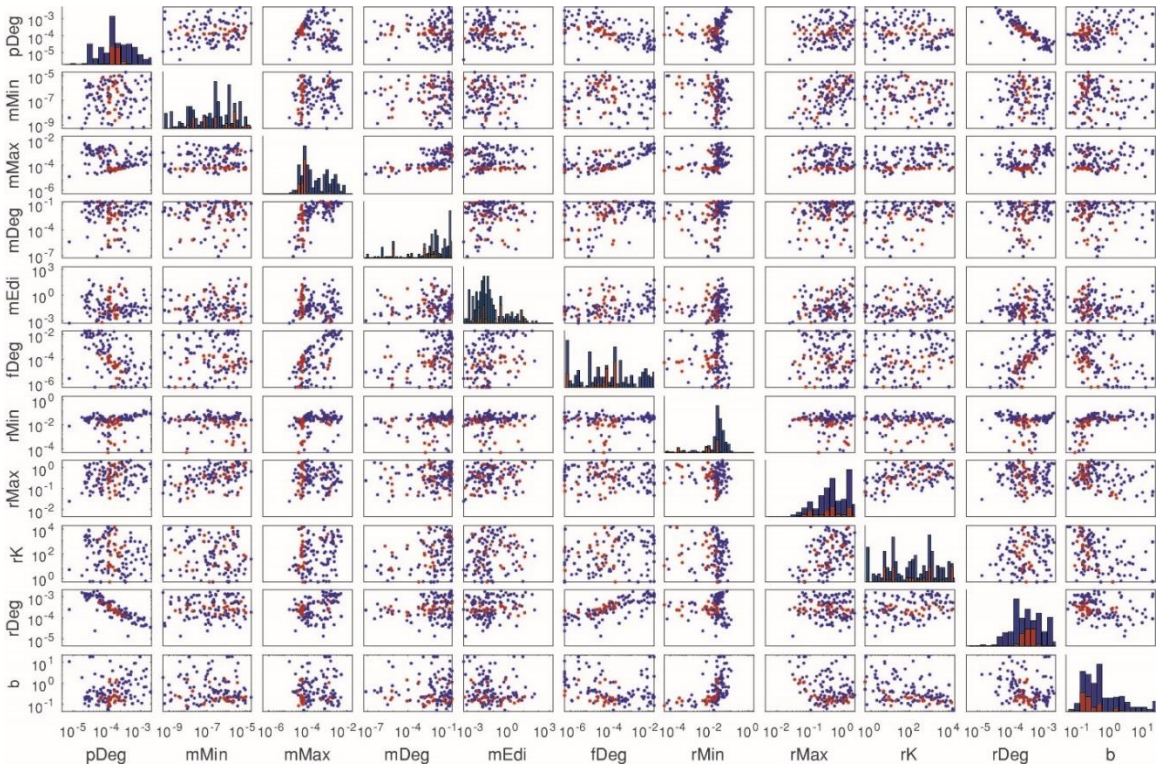

Figure 4-11 Pol II parameter analysis

Pol II time course parameter analysis of the lowest error fittings selected in Figure 4-10. On the diagonal, histogram distributions of each individual parameter are show. Scatter plots of each parameter against the others are shown in the other panels.

the degradation rate of mRNA (mDeg) was almost 100-fold higher than fgRNA degradation (fDeg). In fact, the mRNA degradation rate was similar to the rate of editing into gRNA (mEdi), indicating that mRNAs are divided relatively equally between editing and export/degradation and become stabilized once edited into gRNA, possible through complexing with dCas9. Taken together, model and guided experiments provide detailed and quantitative characteristics of Cas9 based gene expression regulation dynamics, which are otherwise hard to acquire.

4.2.4 Fluorescent Guides Drive Layered Pol II only CRISPR Circuits

Cascades are a common motif in natural regulatory systems which have been shown to act as noise filters and as memory devices in synthetic networks (44, 172, 173). The use of fgRNAs in CRISPR-based circuits allows observation of previously hidden nodes whose activity could only be inferred from network inputs and outputs. This improved resolution allows us to more accurately characterize the network's behavior and to troubleshoot more effectively. CRISPR-based layered circuits enable us to leverage the power of CRISPR and combine it with logic-based design methods for sequential gene editing or epigenetic modulation, which will aid in more sophisticated and controllable therapies. However, synthetic layered Pol II CRISPR circuits previously failed to show functionality, so we set out to devise strategies to improve them (74).

Network analysis revealed that increasing mRNA production or decreasing mRNA degradation were potential targets for improving network response. Therefore, we constructed and screened a number of fgRNA constructs modified to affect these areas (Figure 4-12 & 4-13). First, large portions of non-translated RNA are often found on the 5' end of mRNA transcripts (5' UTR), and it is believed that it plays a regulatory role. It has been shown that the length of the 5' UTR can control the expression level from Pol II and may decrease nucleosome occupancy at the +1 position (174). Therefore, we sought to unravel whether modified UTR length could influence gRNA expression from a Pol II promoter. For this we inserted an additional random 20 nt sequence within the 5' end of the gRNA transcript, immediately after the transcriptional start and compared the efficiency with the original design. Second, the mRNA Poly-A tail is strongly associated with nuclear export (175–177). Therefore, we hypothesized that

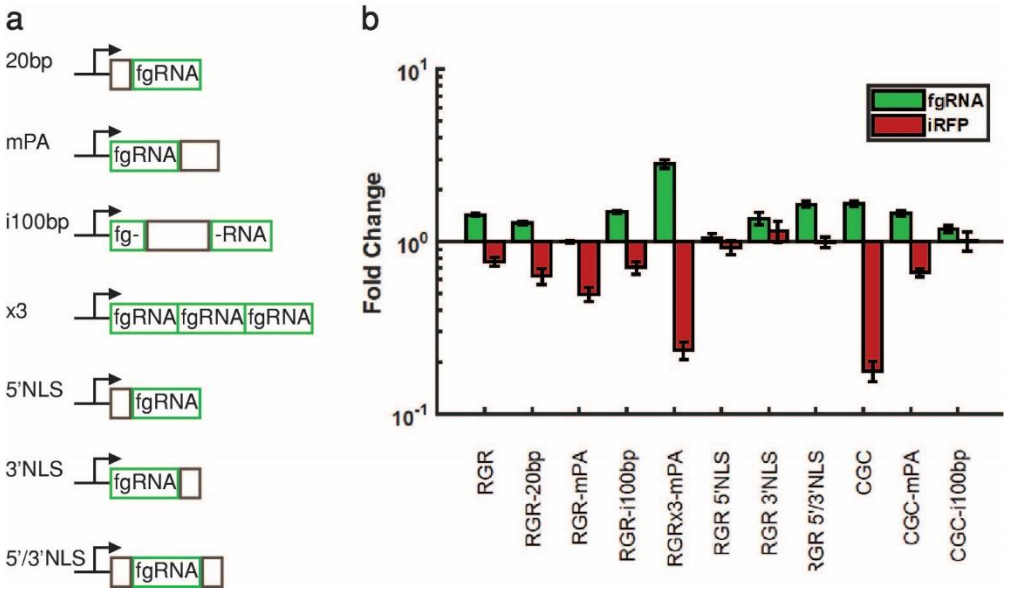

Figure 4-12 Modified fgRNA transcripts

Transcript modifications to improve fgRNA expression (a) Diagrams of various modifications employed to increase circuit functionality: 20bp spacer between transcritional start and fgRNA (20bp), mini-poly A terminator (mPA), intronic 100bp sequence in the fgRNA (i100bp), multiplexing three transcripts (3x), inclusion of an RNA nuclear localization tag on the 5', 3' or both ends (5'NLS, 3'NLS, 5'/3'NLS). (b) Fold change analysis of modified RGR or CGC constructs, showing fgRNA (green) and iRFP (red) expression. Data are the mean of 4 flow cytometry replicates' medians ± SD.

interfering with this component might lead to a larger fraction of mRNA transcripts being retained in the nucleus. A similar approach has been employed to optimize the efficiency of shRNA expression from Pol II promoters (178). So, we incorporated a truncated minimal poly-A terminator (mPA) to provide a smaller poly-A tail to the transcript and compared the efficiency with transcripts harboring the original Pol II terminator. Additionally, introns have been shown to increase mRNA accumulation when compared to similar transcripts which lack introns (179, 180). Along this line, a random, intronic 100bp sequence was added into the middle of the fgRNA sequence. Next, we made several new designs to improve localization to the nucleus. An RNA sequence shown to impart nuclear localization in long non-coding RNAs (lncRNA) was added to either the 5' or 3' end of the mRNA transcript (181). Because this sequence has been shown to

reliably localize lncRNAs to the nucleus, it was hypothesized that it could have a similar effect on the fgRNA-containing mRNAs. Finally, gRNAs were multiplexed, placing between 2 and 8 copies of the same fgRNA plus editing sequences one after another in the transcript, each separated by a short linker sequence. We hypothesized, this strategy would increase gRNA expression relative to the multiplex number per mRNA transcript.

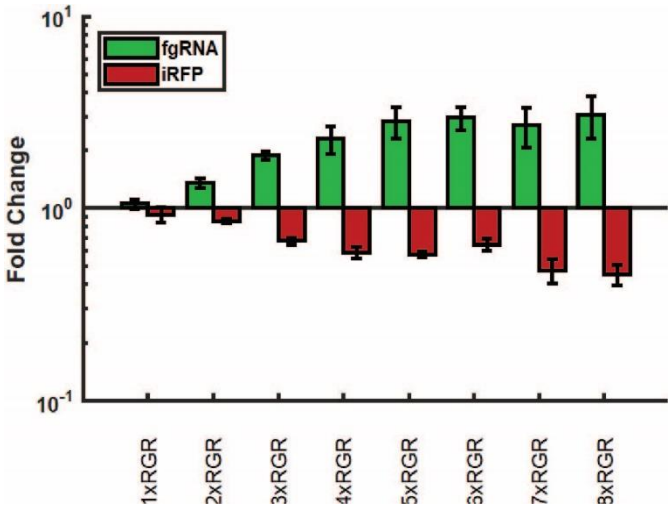

Figure 4-13 Multiplexed fgRNA transcripts

Fold change analysis of progressive multiplexing (1x through 8x) of an unmodified RGR construct, showing fgRNA (green) and iRFP (red) expression. Data are the mean of 4 flow cytometry replicates' medians ± SD.

Applied to both RGR and CGC constructs, the results of this screening are shown in Figures 4-12 and 4-13. While the CGC transcript was relatively functional to begin with, additional modifications had little effect. Conversely, initial screening of the RGR construct revealed that both expression and repression were minimal, but modification of the transcript resulted in a much more functional construct. The modifications yielding the greatest effect were addition of the mPA terminator, as well as multiplexing several copies of the fgRNA into a single transcript. Sequential addition of guides increased performance up to 4 or 5 guides, at which point continued multiplexing did not appreciably increase expression. Other modifications – changing spacing between promoter and transcript, increasing availability using intronic sequences, and inclusion of

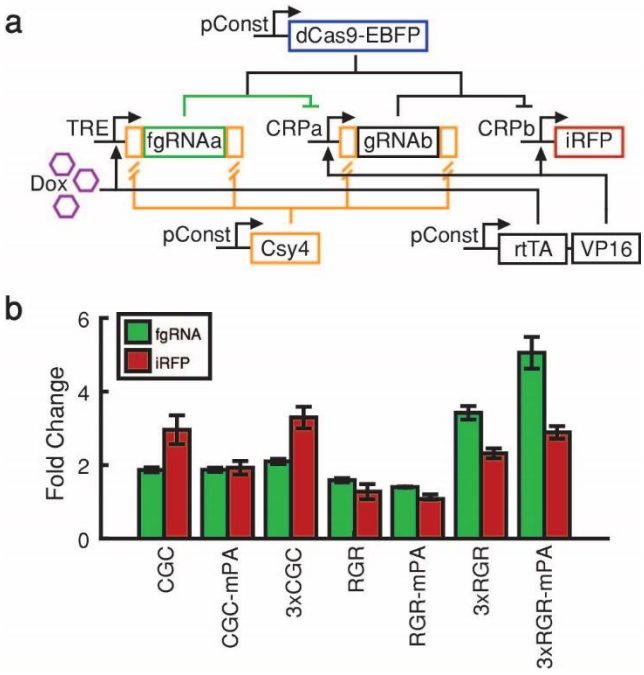

Figure 4-14 Pol II repressor cascade

Targeted fgRNA modification improves function. **(a)** Diagram of the inducible, Pol II-driven two-tier repressor cascade. Csy4 is shown editing the gRNA transcripts in both positions into functional guides. In some experiments, the inducible first node was replaced with an RGR construct. Induction of the circuit with dox should increase fgRNA and iRFP expression. **(b)** Fold changes of the two-tier cascade with various fgRNA constructs in the first position. By modifying the fgRNA transcript via multiplexing (3x) or alterations to the terminator tail (mPA), fgRNA and iRFP expression are improved using RGR constucts. Data are the mean of 4 flow cytometry replicates' medians ± SD.

lncRNA nuclear localization sequences – resulted in little improvement. As such, further efforts focused exclusively on multiplexing and terminator modification.

Next, we sought to determine whether these improved Pol II driven designs were capable of generating a functional two-layer, strictly Pol II gRNA transcriptional cascade. We transfected HEK293FT cells with the cascade circuit (Figure 4-14A), with and without Dox induction, and measured with flow cytometry 72 hours post-transfection. An unmodified, non-fluorescent CGC construct repressed iRFP expression and was itself subsequently repressed by a Dox-responsive modified fgRNA RGR or CGC. To circumvent the lack of non-linearity, twice as much fgRNA, relative to the middle-node repressor, was added to more efficiently de-repress iRFP. Combinations of 3x multiplexing and the mPA terminator were then used for the Dox-responsive input node (Figure 4-14B). These yielded a moderately functional cascade that was not achievable in previous work using a similar circuit topology (74). As with the screening, modified

CGC transcripts functioned similarly to unmodified CGCs, while modified RGRs showed marked improvement over the unmodified RGR, with iRFP fold change reaching a level similar to CGC with even greater fgRNA brightness. These results demonstrate that network improvements can be made via targeted re-engineering of circuit components directed by detailed analysis of network behavior. And, even in the absence of cooperativity, components can be adjusted to function strongly enough to exhibit cascading behavior. Furthermore, the RGR results show that such modications can be used to transform a circuit from non-functional to functional.

4.3 DISCUSSION

The ability to directly measure gRNA expression enables precise identification of single cell dynamical behaviors of CRISPR-based circuits, enabling informed optimization decisions to improve circuit functionality. We employed the fgRNA technology to interrogate the dynamics and function of otherwise hidden nodes within CRISPR-mediated synthetic gene circuits. Initial validation of fgRNA constructs shows that the placement of the fluorescent broccoli aptamer within the gRNA transcript has little effect on the function of the gRNA guided repression. The location of the insert only impacts overall fgRNA brightness. It was found that insertion of the aptamer into the gRNA tetraloop produces the highest fluorescence. We hypothesize that insertion into the second loop or tail may result in aptamer misfolding or prevent dye binding. Additionally, inclusion at the tail may be hindered by premature transcriptional termination, resulting in a functional gRNA with an incomplete aptamer. This observation may indicate design constraints for other gRNA-aptamer systems.

gRNA regulated by RNA Pol II promoters provides an attractive platform to generate libraries of composable CRISPR-based gene networks, and thereby enable scaling to more sophisticated circuits. To identify the optimum strategy for gRNA expression from Pol II promoters, we compared three different RNA editing strategies RGR, CGC, and TGT. The CGC-based strategy is shown to be more efficient than the other two in our experiments. Furthermore, while TGT editing resulted in no loss of function from a Pol III promoter, when used in a Pol II context, it exhibited a lack of inducibility. This is likely because tRNAs themselves may act as promoters (182). As tRNAs have been proposed as a means of efficiently multiplexing gRNAs in a single transcript, this is an important consideration for future studies.

As with Pol III, dynamics of the Pol II driven repressor were evaluated mathematically, yielding several intriguing findings. First, though the editing sequences may interfere somewhat with transcription, CGC editing itself is highly efficient. Second, we confirmed mathematically that Pol II expression is 10-100x lower than Pol III expression. While some of this could be the result of the flanking editing sequences, it also suggests that the Pol II promoters simply produce fewer transcripts than Pol III promoters. This is likely because mRNA transcripts can be upregulated during the translation stage in normally functioning Pol II expression systems (161). Third, we show that while parameters shared between Pol II and Pol III circuits – pDeg, fDeg, rMin, rMax, rK, b – are centered around the same values as expected, the Pol II model exhibits increased variability around this center, suggesting that Pol II driven gRNAs may be less well-regulated – in terms of production, degradation, and repression effectiveness – than their Pol III counterparts. This may be due to variability introduced by editing, as that is

the primary point of difference between the networks; however, why this would be true mechanistically is unclear. Alternatively, this may be a mathematical artefact due to the more noticeable role of stochasticity within smaller populations. Smaller changes in output, like those seen in the Pol II network, may have more combinations of parameters that still fall within a physiologically relevant range, whereas the larger changes of the Pol III network, tend to group more clearly. If increased variability were to remain after further improving gRNA expression, identifying the source of this variability and ways to control it will be an interesting and necessary route for future experimentation. Finally, our analysis and experiments demonstrate a critical property of CRISPR circuits: dCas9 regulation lacks cooperativity, resulting in a linear relationship between the amount of gRNA-complexed dCas9 and the response of the circuit. Nonlinearity is an essential component of multistable networks and is crucial for noise reduction and maintaining signal fidelity in larger networks. Engineering cooperativity is still a challenge in CRISPR circuit construction and might be accomplished through dimerization of the Cas9 protein or through inclusion of RNA aptamers which allow gRNAs to recruit additional gRNAs.

With a mathematical understanding of the dynamics underlying the fgRNA repressor network, we identified areas – gRNA production and degradation – that we could alter to improve system function to produce a functional transcriptional cascade using only Pol II driven components. We focused on improving gRNA availability by increasing gRNA production through multimerization and reducing gRNA nuclear export through terminator selection. As we demonstrate, lower Pol II production can be offset through multiplexing several identical gRNAs into the same transcript, though this

method runs into limits from transcriptional falloff and plasmid instability due to highly repetitive sequences. Additionally, alterations to the terminator impart increased fgRNA brightness and downstream repression. Through these alterations, we were able to transform the previously non-functional Pol II driven RGR-based CRISPR repressors into one capable of driving downstream derepression in a repressor cascade. Circuit component concentration optimization alone was enough to produce a functional CGC cascade. Thus, we succeeded in developing two editing methods that could produce functional Pol II gRNA transcriptional repression cascades, which was not previously achieved (74). While CGC requires coexpression of the Csy4 protein for editing, improved RGR efficiency provides an all-in-one system that works equally well and gives researchers additional flexibility to overcome experimental constraints. Many viral delivery methods, for example, impose a limit on the amount of DNA which can be packaged, so inclusion of an additional protein reduces the available space for therapeutics (183). With a process as complex as transcription, we are left with myriad angles for potential innovation in this area, such as the inclusion of enhancer sequences within the promoters, optimized RNA Pol II promoter sequences, optimized transcriptional start sites, or improved nuclear localization sequences. We used a mathematical approach to direct circuit modification, but the dissimilar response of RGR and CGC to similar modifications indicates that quantification and standardization of DNA sequence selection and assembly methods remains an area of importance for synthetic biology. There is still enough variability between research groups and experimental methods that approaches which yielded negative results in our hands cannot

be rejected outright. Universal standardization of methods for similar studies will aid in better characterization of these networks.

Constructing reliable and predictable gene networks is a nontrivial undertaking. The recent prominence of CRISPR technology promises to improve the process by offering easier generation of unique, orthogonal components and by allowing easier engineering of interactivity between network parts. With the transition to RNA-based transcriptional regulation; however, additional areas require further exploration. Here, we present a tool for visualization of gRNA dynamics within cells and demonstrate how its proper implementation can allow for improved modeling, prediction, and functionality of CRISPR-based gene circuits.

4.4 METHODS

Cell Culture and Transfection

All experiments were performed in HEK293-rtTA3 cells (cell line generation detailed by Kiani, et al (74)), a strain of HEK293FT cells with genomically integrated constitutively active rtTA activator. Cells were maintained in DMEM (Corning Life Sciences) supplemented with 10% FBS (Sigma Aldrich), 1% non-essential amino acids (NEAA; Gibco), 1% L-glutamine–streptomycin–penicillin mix (Gibco), and 1% GlutaMax (Gibco). Transfections were performed using Polyethylenimine (PEI) as a transfection reagent (Polysciences, Inc.). Cells were seeded in 24-well plates the day before so that they were at ~80% confluence at the time of transfection. Masses of various plasmids used in each set of experiments are shown in Table S2. After DNA mixes were made, the volume was brought to 25 µL by adding DMEM (no supplements) then combined with

an additional 25 µL DMEM (no supplements) with PEI equivalent to a 3:1 DNA:PEI ratio. This was vortexed twice for 1 second each and allowed to stand at room temperature for 30 minutes. While the DNA mixes sat, media was changed (DMEM with supplements above), and if necessary Dox was added to the wells as an inducer. All Dox inductions were performed at a concentration of 2 µg/mL. After 30 minutes, 50 µL of the DNA mixture was added to each well with micropipette, dipping the tip into the well's media and slowly ejecting while swirling inward, careful not to scrape the bottom of the well. Media and inducers were changed daily until analysis.

Plasmids

Plasmids were constructed using golden gate cloning methods, with pieces either copied from existing plasmids via PCR or de novo synthesis. All DNA components were purchased through Integrated DNA Technologies (IDT). The CRP-iRFP reporter plasmid was assembled using gateway cloning, combining the promoter and protein coding region in a gateway destination vector backbone. The Csy4 plasmid, PGK1p-Csy4-pA (Construct 2), was a gift from Timothy Lu (Addgene plasmid # 55196) (157). The dCas9-EBFP plasmid, pHR-SFFV-dCas9-BFP, was a gift from Stanley Qi & Jonathan Weissman (Addgene plasmid # 46910) (184).

Flow Cytometry

Prior to flow cytometry, wells were trypsinized with 100 µL 1x trypsin (Gibco) then inactivated with 200 µL Hanks Balanced Salt Solution (HBSS; Corning Life Sciences) without Calcium or Magnesium but supplemented with 2% FBS. These were transferred

to a 96-well plate and pelleted at 300g for 2 minutes at 4°C. The supernatant was aspirated, and cells were resuspended in 200 µL Phosphate Buffered Saline (PBS; Corning Life Sciences) with 4% FBS and 40 mM DFHBI-1T (Lucerna), as recommended in prior literature(165). Flow Cytometry was performed either daily (for time courses) or 72 hrs post-transfection, using a FACSCelesta flow cytometer (Becton Dickson) with HTS attachment. The cytometer was configured with Violet (405nm), Blue (488nm), and Red (640nm) lasers, used for excitation of EBFP (450/40 filter), Broccoli/DFHBI-1T (530/30 filter), and iRFP (780/60 filter), respectively. Samples were collected at 1.5 µL/s to a total of 200,000 events.

Data Analysis

Data were analyzed using MATLAB (MathWorks, Inc.). Gates were generated against a test data set using a Gaussian Mixed Model (GMM), then applied to all experimental data. The GMM used 6 clusters with 20 replicates, selecting the highest log likelihood. Channels used for gate generation were Front Scatter (FSC-A), Side Scatter (SSC-A), and EBFP (BV421-A) which was our transfection marker. Because the EBFP values were log distributed, we used a Log10 tranform of the actual values to fit the GMM. This same transformation was also performed on all experimental data before clustering, then reversed to maintain the original values. Once gated, the median green (BB515-A) and infrared (APC-Cy7-A) fluorescence of all cells with expression >0 was calculated. Fold change of these fluorescences were calculated by dividing the expression with the addition of gRNA/Dox by the expression beforehand.

Modeling

Ordinary Differential Equation (ODE) models were solved and analyzed using MATLAB run on a personal computer. We designed a system of ordinary differential equations (ODEs) to describe the expression of important components in the system. For the U6 driven fgRNA repressor, we began with equations for fgRNA (F, eqn 4-1) and iRFP (R, eqn 4-2), following standard forms for production / degradation and hill function repression.

$$\frac{dF}{dt} = pol3 * fMax - F * fDeg \quad \text{eqn (4-1)}$$

$$\frac{dR}{dt} = reg * irfp * \left(rMin + \frac{rMax}{1+(rK*cas*F)^b}\right) - R * rDeg \quad \text{eqn (4-2)}$$

Due to the transient nature of the transfection protocol used, we added another equation to describe plasmid dilution with each subsequent cell division (P, eqn 4-3), which was then integrated into the F and R equations (eqns 4-4 and 4-5). These equations were used for fitting the Pol III experimental data.

$$\frac{dP}{dt} = -pDeg * P \quad \text{eqn (4-3)}$$

$$\frac{dF}{dt} = (P * pol3) * fMax - F * fDeg \quad \text{eqn (4-4)}$$

$$\frac{dR}{dt} = (P * reg) * (P * irfp) * \left(rMin + \frac{rMax}{1+(rK*(P*cas)*F)^b}\right) - R * rDeg \quad \text{eqn (4-5)}$$

To account for both Pol II as well as Pol III RNA production, a fourth equation was added representing mRNA expression (M, eqn 4-6). This equation included an editing term, which converts some portion of M into F, requiring modification of the equation of F to account for this change (eqn 4-7). The modified equations used for fitting the Pol II data were:

$$\frac{dM}{dt} = (P * pol2) * (mMin + dox * (P * reg) * mMax) - M * mEdi - M * mDeg \quad \text{eqn (4-6)}$$

$$\frac{dF}{dt} = (P * pol3) * fMax + M * mEdi - F * fDeg \quad \text{eqn (4-7)}$$

Model fitting was performed in MATLAB software using a least-squares curve-fitting algorithm (lsqcurvefit). The fitting was first performed on the Pol III data using equations 4-3, 4-4, and 4-5 to fit the following parameters: pDeg, fMax, fDeg, rMin, rMax, rK, b, and rDeg. The algorithm was run 500 times with randomly selected initial conditions within physiologically relevant bounds (determined empirically). The fitting rank and squared error (resnorm) are shown in Figure 4-5. Because there was no clear region of noticeably superior fit, we selected the 250 best fits, as well as a smaller sub-population of the 50 best fits, for further analysis. Figure 4-6 shows these fit parameters' distributions as histograms and as scatter plots against one another.

When expanding to the 4-equation Pol II model (eqns. 4-3, 4-6, 4-7, and 4-5) the same fitting algorithm was used on the 11 relevant parameters parameters: pDeg, mMin, mMax, mDeg, mEdi, fDeg, rMin, rMax, rK, b, and rDeg. Again, fittings were ranked by resnorm (Figure 4-10) and plotted against one another (Figure 4-11).

5 INTRACELLULAR NOISE LEVEL DETERMINES RATIO CONTROL STRATEGY, SPEED, AND ACCURACY

This chapter was prepared in collaboration with Xiao-Jun Tian, Ph.D. and Xiao Wang, Ph.D. and is under consideration for publication.

5.1 INTRODUCTION

Ratio control of differentiation within isogenic populations is a ubiquitous but poorly understood phenomenon. From single celled microbes to higher organisms, many processes require mixed populations to carry out complex functions, like bacterial persistence (11, 120), bacterial competence (30, 119), nasal and ocular receptor development (185–187), differentiation of blood and vascular cells (12), immune response (188), and stem cell maintenance and differentiation (189). Several general explanations for how this phenotypic diversity arises have been proposed, such as stochastic fluctuations within gene regulatory networks (48, 190), asymmetrical sequestering of regulatory proteins during cell division (191), and differential response to spatial gradients of extracellular soluble factors (192). While each of these methods could theoretically generate a mixture of differentiated cells within a population, research has frequently focused on the types of cells yielded, rather than the quantitative control of their ratios. These processes are often tightly controlled in terms of ratio accuracy and attainment speed to avoid overspecialization or to ensure normal development (193, 194). Developing a unified understanding of the mechanisms and relevant factors to achieve and maintain precise ratio control will have widespread benefits in areas such as

countering bacterial immunity (27), treating diseases in which ratio control is disrupted, as is the case in some platelet disorders (195) or mastocytosis (196), or in developing improved protocols for stem cell differentiation (197, 198).

As with many biological systems, the highly interconnected nature of the underlying genetic regulatory circuitry makes it difficult to study phenotypic ratio control without encountering myriad confounding variables. Synthetic biology, by cutting through to the basics of transcriptional regulation in isolated and orthogonal circuits, offers an attractive route for exploring mechanisms underlying ratio tuning. Studying fundamental genetic motifs in isolation has yielded a greater understanding of key cellular behaviors, such as multistability, oscillation, and adaptation (77, 79, 82, 135). Multistable networks, specifically, are highly relevant for cellular differentiation processes and have wide applicability in a diverse range of contexts, from developmental biology (199), to targeted therapeutics (200), and cell-environment interactions (201). Synthetic toggle switches have been implemented repeatedly in multiple organisms, demonstrating the feasibility of studying differentiation with minimal, synthetic circuits (79, 82, 202, 203). Further exploration of bistable circuits can reveal how sub-population ratios can be controlled and manipulated, the limits of various control schema, and best practices for controlling state transitions on the single-cell and population levels. For example, recent work has studied how circuit component selection effects the network's hysteretic region (82) and how inclusion of additional circuit components can be used to adjust population ratios (90).

Here, we present methods for reliably tuning multimodal populations' ratios without the need for additional network components. First, using *E. coli*, we show that

positioning of a population with noisy intracellular expression dynamics within the bistable region can be enough to achieve fractional differentiation. Then, using the less noisy expression dynamics of *S. cerevisiae*, we demonstrate that temporary deviations from the bistable region can direct robust ratio differentiation in a low-noise system. By precisely modulating the stimulus strength and duration for which the network is moved towards the other state, we achieve reliable ratio tuning. From these findings, we develop a mathematical framework through which we can fully understand the roles of stimulus dosage, stimulus duration, and noise in driving fractional state switching of cellular populations. Gene expression noise acts as a global regulator of ratio tuning speed and accuracy, with noise level positively correlated with speed but negatively correlated with accuracy. Low noise systems become candidates for pulsed induction ratio tuning, in which an inverse correlation between stimulus dose and duration is observed, translating into a tradeoff between the speed to attaining a chosen ratio and ratio accuracy.

5.2 RESULTS

5.2.1 Noise-induced ratio control in bacteria

To investigate mechanisms of ratio control in bacteria, we first use the well-established bistable toggle switch circuit (204). The topology is that of mutual inhibition (Figure 5-

1a) in which TetR and LacI repress one another while LacI is co-expressed with green fluorescent protein (GFP). TetR activity can be modulated with the addition of the small molecule anhydrotetracycline (ATc), which inhibits TetR activity and hence alleviates its inhibition on the expression of LacI. When grown in the absence of induction, the system favors the TetR dominant, low GFP state. A hysteresis curve was generated as a function of ATc concentration (Figure 5-1b) to probe the cells state distribution in and out of its bistable region. Fitted with a deterministic model, it is indicated that the left bifurcation point is less than 0 ng/mL ATc while the right bifurcation point to be around 2 ng/mL ATc. Within the bistable region

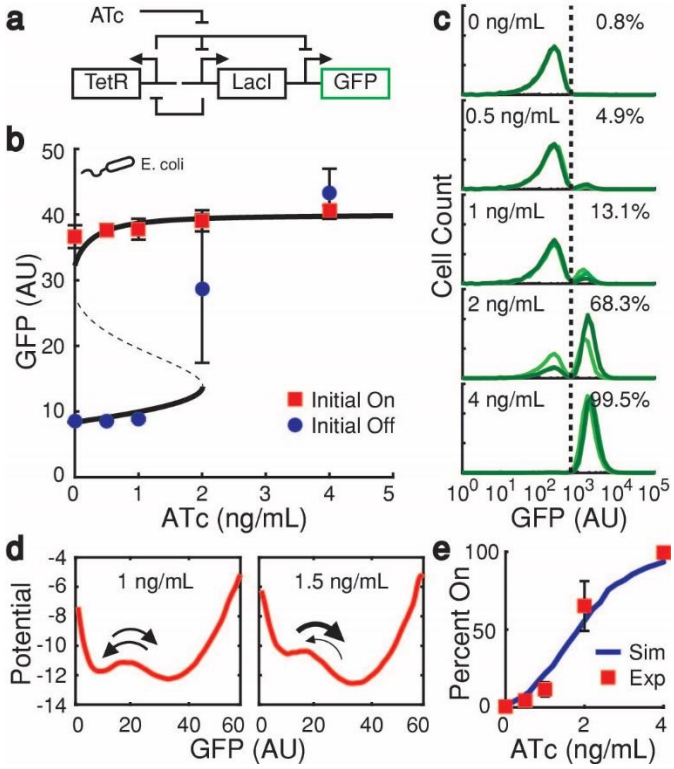

Figure 5-1 Noise induced ratio control

Gene expression noise in *E. coli* induces fractional differentiation within the bistable region. (a) Diagram of the mutual inhibition toggle switch. (b) In E. coli, this toggle switch exhibits hysteresis from 0 to roughly 2 ng/mL ATc. Within this range, cells can stably hold either the low GFP or high GFP states. The plot shows mean of 3 replicates' medians ± SD, overlaid with model-predicted hysteresis curve. (c) Flow cytometry histograms of initially GFP-off populations show that as ATc dose increase, gene expression noise causes an increasing fraction of cells to spontaneously turn on. Dashed lines indicate an empirical threshold between OFF and ON cells. Three replicates are shown on each plot with corresponding dosage and the total percentage of ON cells noted. Data of 10000 cells are collected for each experiment. (d) Energy potentials computed for different ATc dosages within the hysteretic region show the relative stability of the two steady states. At 1 ng/mL cells transition at similar rates between wells (arrows), but as ATc dosage increases, the potential well for the low state becomes shallower, allowing cells to more easily transition to the high GFP state (represented by thicker arrow head). Potential wells are generated from a stochastic simulation fit to the hysteresis curve data. (e) Stochastic model predicted ON percentage (blue line) fit experimental results (red square) accurately. Experimental ON percentage is computed as the mean ON percentage of three replicates.

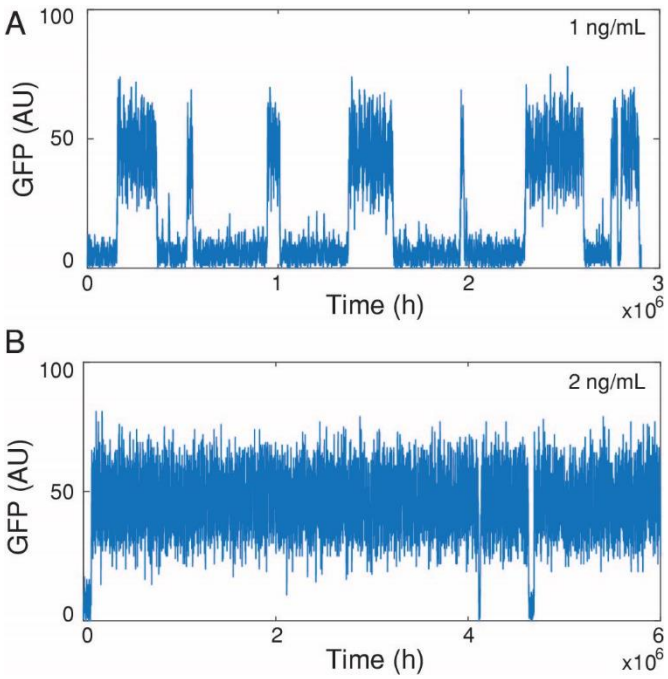

Figure 5-2 Simulation of noisy kinetics

Simulations with different levels of ATc show random state switching, with the time spent in each state indicating the relative depths of each state's energy well. (A) At 1 ng/mL ATc, cells transition relatively freely between the GFP-on and GFP-off states. (B) At 2 ng/mL, cells preferentially transition to the GFP-on state; however, gene expression noise is large enough to occasionally cause them to transition back to the GFP-off state.

bimodal behavior was observed, with a portion of cells in the high GFP state while some in low GFP state (Figure 5-1c). This bimodality is believed to be caused by noise driven spontaneous state transitions (60, 119, 205).

To quantitatively understand the relationships between noise and resulting bimodality of phenotypes, we developed a stochastic model to explore the energy landscape underlying this bistability. It is shown that as ATc concentration increases the depth of the left (low GFP) potential well decreases (Figure 5-1d; Figure 5-2). This essentially lowers the barrier of state transition from low GFP to high GFP, predicting higher ATc induction would result in a larger percentage of cells in high GFP state. Experiments indeed confirmed model predictions (Figure 5-1c, Figure 5-3a). This fraction changed little between measurements at 5 and 8 hours, suggesting it is stable. Furthermore, it is shown that our model can predict such ratio control with high quantitative accuracy (Figure 5-1e, Figure 5-3b-c), showing that high-GFP cell ratios increase monotonically but nonlinearly as ATc induction increase. Such a gradual and

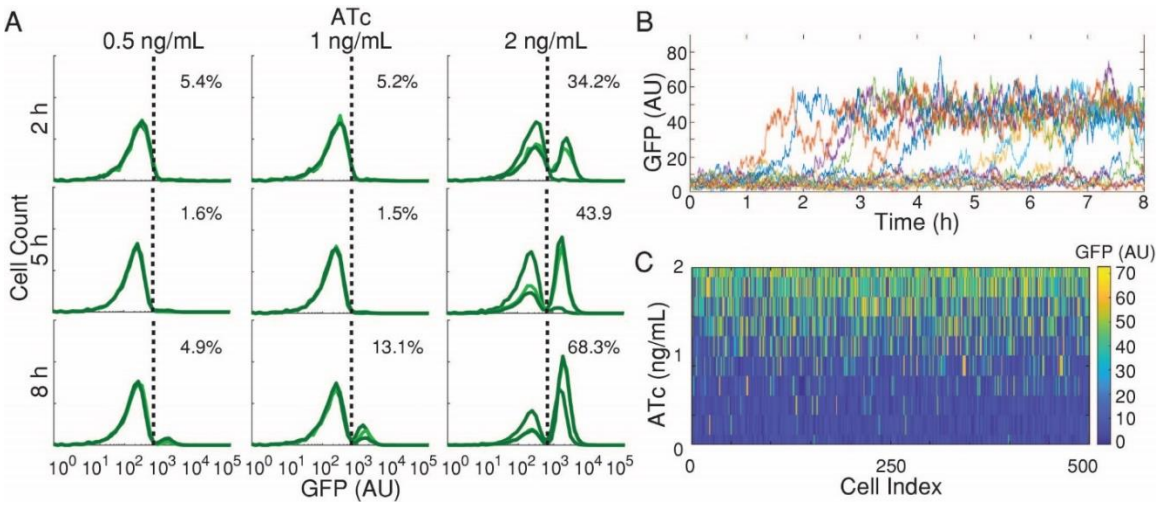

Figure 5-3 Noise within the hysteretic region

Noise induces bimodality within the hysteretic region. (a) Histograms of ATc doses within the hysteretic region show that as time progresses, cells spontaneously switch states due to intracellular noise. (b) Model results showing typical simulation results at 1.5 ng/mL ATc. By the end of the 8 hr simulation, roughly half of the simulated cells have transitioned to the GFP-on state. (c) Heat map of simulation results for 500 cells at range of ATc concentrations from 0 to 2 ng/mL, showing GFP expression after 8 hours of simulation. At higher ATc concentrations, cells show greater average GFP expression.

steady increase of ON cell percentage as a function of induction doses could enable a precise ratio control for the whole bacteria population.

This finding of tunable ratio control by adjusting induction strength within the bistable region suggests one possible ratio control strategy for systems that require precise fractional, but uncoordinated, control of population differentiation. Bacteria, for example, often keep a subset of the colony in a dormant persister state to ensure survival in the case of unexpected environmental shifts or antibiotics (11, 120). Stochastic switching provides a simple mechanism for entering and exiting this state. Similar behavior is seen in the bacterial motility and adhesion decisions (206, 207), or in switching to a mutable, competent state (193). In these cases, the overall ratio remains relatively fixed; though, individual cell's states are not. However, this strategy demands relatively high intracellular noise, which is common in plasmid-based bacterial systems

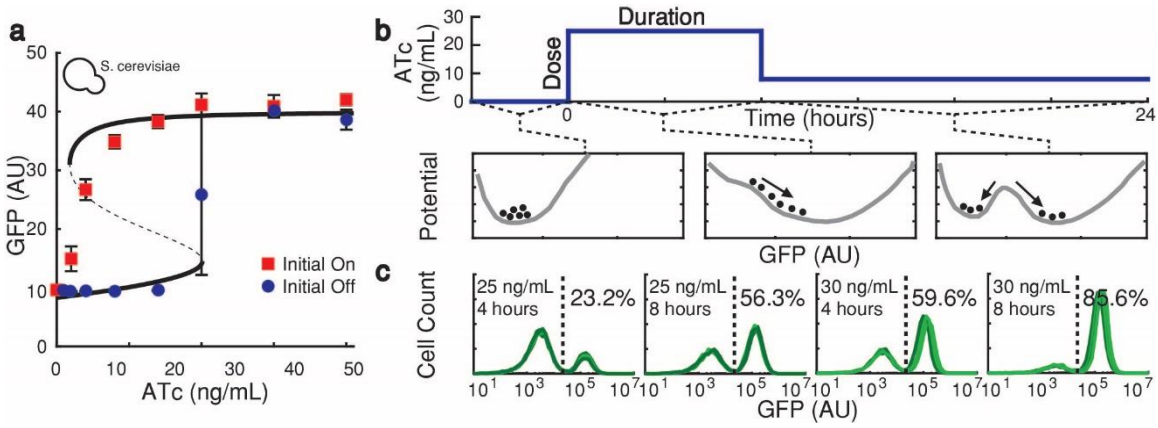

Figure 5-4 Pulsed induction ratio control

In *S. cerevisiae*, gene expression noise is lower, requiring temporary deviation from the bistable region to induce multimodality. (a) The hysteresis curve of the yeast toggle switch shows strong stability. The plot shows mean of 3 flow cytometry replicates' medians ± SD, overlaid with model-predicted hysteresis curve. (b) Yeast were exposed to pulses of ATc with varying dose and duration. Before the pulse (Phase I), all cells resided in the low-GFP state without any induction, corresponding to a single-welled potential landscape. The induction pulse (Phase 2) changes the underlying landscape to a single well in the high-GFP state, and cells begin to transition from low to high GFP state. Before all cells transition, induction is reduced to 8 ng/mL (Phase 3), at which there are two deep potential wells, low GFP and high GFP. Partially transitioned cells either transition fully or return to their initial low-GFP state. (c) Four flow cytometry experiments with four replicates each after pulse induction are shown with the total percentage of cells to the right of the dashed line. Inducer pulses for controlled durations allow for a wide range of ratios, with multiple paths to the same endpoint. Pulses of either 25 ng/mL or 30 ng/mL ATc for 4 or 8 hours achieve a range of final ratios, with the 25x8 and 30x4 pulses producing nearly identical outcomes.

(208), so that individual cells could spontaneously transition between states within reasonable amount of time.

5.2.2 Pulsed induction ratio control in yeast

Spontaneous and random back and forth switching between states is certainly not suitable for processes requiring irreversible cell fate determination, such as development and cell differentiation. In these contexts, intracellular noise would need to be low enough to avoid stochastic state switching. To identify possible ratio control strategies in such low-noise environment, we transitioned experimentally to a less-noisy system: the chromosomally integrated mutual inhibition toggle in *S. cerevisiae* (82). This circuit has

the same topology as that shown in Figure 5-1a, exhibits hysteretic behavior (Figure 5-4a), and favors the TetR-dominant, low-GFP state under no induction. However, differences in promoters, copy number, and transcription-translation processes between *E. coli* and yeast serve to reduce intracellular noise and shift the bistable region up to roughly 3-14 ng/mL ATc. Unlike in *E. coli*, the bulk of the bistable range was impervious to the effects of intracellular noise, resulting in a single peak homogeneous expression profile even when the system is operating within the bistable region (82). Instead, to induce bimodality, we hypothesized that internal variability could be utilized by temporarily forcing the cells outside of the bistable region favoring another state. Then the population would begin to transition to the other state. However, natural stochasticity would cause some cells to transition faster than others. When the population was returned to the bistable region prior to full-population transition, some fraction of the cells would finish their transition while the rest would return to their original state. Figure 5-4b schematically illustrates this process and how the various stimulus levels adjust the underlying potential landscape to induce bimodality. This diagram also shows the two variables which determine population response: dose, which measures the magnitude of the induction pulse, and duration, the length of time for which the dose is applied. Using this method, the population achieved a specific phenotypic ratio and individual cells only transition state once, which is distinct from *E. coli.*

Systematic temporal induction experiments were then designed and carried out to test our hypothesis. Using doses of 20, 25, 30, 35, and 40 ng/mL ATc, with pulse durations between 2 and 24 hours, we comprehensively explored the range of ratios yielded by various dose/duration pairs (Figure 5-5). As seen in Figure 5-4c, there exists

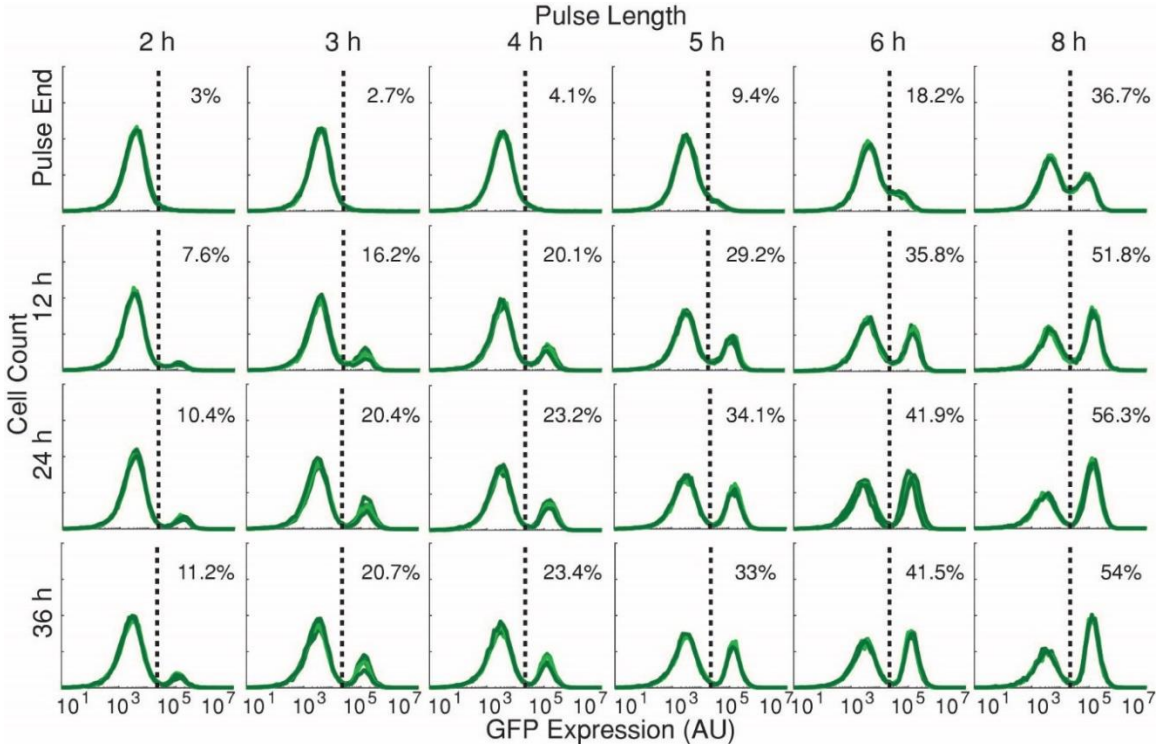

Figure 5-5 Low-noise pulsed-induction experiment

A sample experiment in yeast, showing 25 ng/mL pulses of ATc for between 2 and 8 hours. At the 12-hour mark, the population is still settling into a stable on/off ratio. By 24 hours, population ratio is fixed, as evidenced by the minimal differences between 24 and 36 hours. Plots show overlays of 5 replicates, with the total percentage from all replicates to the right of the dotted line indicated.

an inverse relationship between dose and duration, with increases in either variable causing a larger fraction of the population to transition to the high GFP state. Therefore, similar fractional responses can be obtained through multiple induction routes. For example, as demonstrated in the middle two panels of Figure 5-4c, 8 hours of 25 ng/ml induction and 4 hours of 30 ng/ml both produced about 59% of ON cells. Therefore, this pulsed induction method is experimentally verified to be able to produce tightly controlled yeast population ratios.

To further understand the tradeoff between dose and duration, we developed a system of stochastic differential equations (SDEs) to help interpret experimental results. The model fit of the experimental data (Figure 5-6a) provides insight into the range of

doses and durations appropriate for producing specific ratios. For example, with a 20 ng/mL ATc induction, it took between 18 and 24 hours to have greater than 70% of the cells to transition to the high GFP state, whereas a 40 ng/mL induction produced the same transition with a 4-hour pulse. Generally, we observe that larger doses produce fast switching dynamics, and smaller doses required progressively longer durations to produce similar switch percentages. As evidenced by the very long durations for the 20 ng/mL dose, induction pulses near the bifurcation point could require durations of a day or more to cause a majority of cells to transition to high GFP. Using experimentally validated model and parameters, more simulation data were analyzed to determine the robustness of the system to temporal perturbation of the pulse length. For this, we looked at the difference in time required to cause 30% and 70% of the population to transition, termed tuning range (TR= $T_{70} - T_{30}$). Figure 5-6b shows that while lower

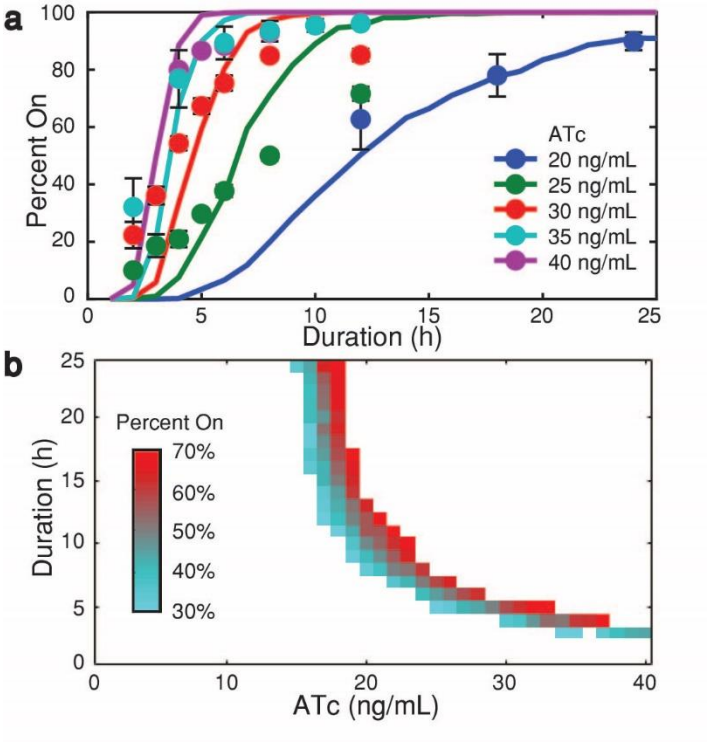

Figure 5-6 Simulations predict ratio control

Mathematical modeling reveals the relationship between dose and duration needed for precise ratio control. (a) Experimental data (circles) were used to fit the stochastic model (lines) for multiple dose-duration pairs, showing strong agreement between the model and experimental results. Experimental data are the mean of 4 or 5 replicates' medians ± SD. (b) Further simulations show induction duration needed for a specific ratio increases exponentially as dose decreases until near the bifurcation point at roughly 14 ng/mL. Additionally, the tuning range between when a small portion of the population has switched states (teal) to when a large fraction has transitioned (red) is large at low doses but shrinks with increasing doses.

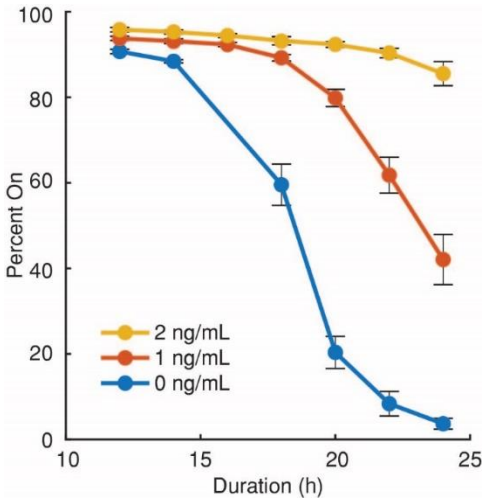

Figure 5-7 On-Off transition dynamics

Yeast exhibit similar state-transition dynamics with low ATc dosages causing on-to-off switching. Durations are longer, as at the initial stages of transitions, LacI must be outcompeted solely by TetR leakage, whereas in the off-to-on case TetR is effectively inactivated by ATc allowing immediate full-strength production of LacI.

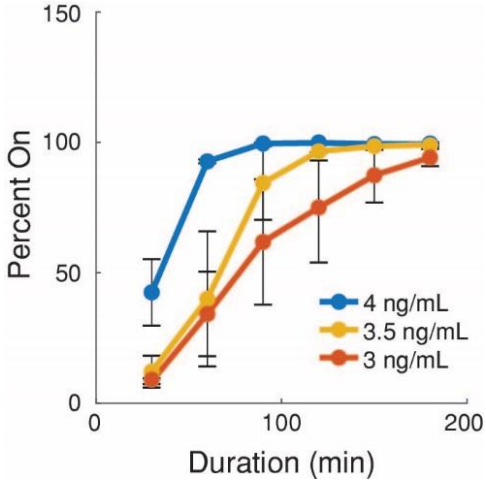

Figure 5-8 Pulsed induction in a noisy system

E.coli experience similar off-to-on transition dynamics to yeast. The ATc range for achieving tunable ratios is lower due to the location of the bistable region, and intracellular noise makes ratio attainment noisier overall.

doses require longer durations to achieve a desired ratio (higher vertical positions of colored bars towards the left), these doses also have a bigger TR (longer vertical span of colored bars), suggesting ratio tuning in this region is more robust against temporal variation of induction pulses. For example, a 25 ng/mL dose generates 30% high GFP with a 4-hour pulse and 70% with a 7-hour pulse, resulting in a TR of 3 hours. For 20 ng/ml induction, T_{30} and T_{70} both increase to 8 and 13 hours, respectively, resulting in an overall larger TR of 5 hours, indicating that there is more room for pulse-length error if a specific ratio within this range is desired. Conversely, larger doses, due to their fast switching dynamics, leave little room for error if a specific ratio is desired, with T_{30} and T_{70} being nearly the same. Results for on-off transitions show a similar relationship between induction dose and required duration (Figure 5-7), and *E. coli* exhibit similar switching behavior as well (Figure 5-8).

5.2.3 The effect of gene expression noise on ratio tuning approaches

To develop a complete understanding of the relationship between noise and ratio control strategies across the noise spectrum, we employed an *in silico* approach that allows us to adjust the noisiness of gene expression while holding other system parameters constant. The model has been shown to be able to recapitulate experimental results accurately under various conditions and therefore serves as an appropriate tool to conduct thorough *in silico* explorations so that we can meaningfully compare ratio control outcomes between high and low noise scenarios.

To investigate the effect of noise on system responsiveness to induction, simulations like those shown in Figure 5-1 were carried out in low and high noise settings. Figure 5-9a shows that increasing noise reduces the time required for a population to transition to a new steady state ratio. The orange line indicates the steady state percentage of the population which will transition to the high-GFP state at a given ATc concentration. All distributions between 0 and 100% are represented with the circuit's bistable region, with the ATc concentrations resulting in 30% and 70% on cells (grey region) defining a region of broad tunability. The steady state is strictly 0 or 100% to the left or right of the bistable region, respectively. Because the time required to reach a steady state approaches infinity in the absence of noise, we measured the time needed for the population to reach half of the steady state ($T_{halfmax}$). In the low noise setting (light blue line), $T_{halfmax}$ is very long within the bistable region, with times of 20 hours or more for steady state ratios below 80%. This is consistent with our previous studies in yeast (82). In the high noise setting (dark blue line), $T_{halfmax}$ is universally reduced, with the largest value being 15 hours when transitioning to 50% high-GFP. Beyond the

bistable region, noise plays a less prominent role, with the high and low noise conditions resulting in very similar transition times. These results indicate that constant induction within the bistable region is only a suitable ratio control strategy for systems with high enough noise to induce frequent and spontaneous state transitions.

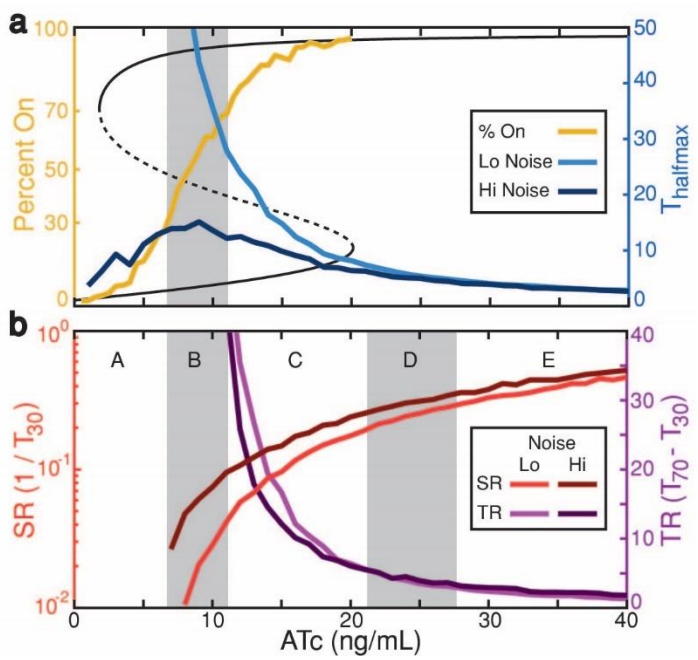

Figure 5-9 Transition speed and accuracy trade-off

Gene expression noise determines appropriate method for achieving predictable ratio control. (a) The orange line indicates the steady state population level average expression of the system, with the gray region bounded by constant induction dosage needed for 30% and 70% high GFP. Blue lines indicate Thalfmax for low noise (light blue) and high noise (dark blue) environments. High gene expression noise reduces the time required for the population to transition from off to the steady state determined by inducer concentration. The hysteresis curve (black) is included for visual reference. (b) Network transition speed (SR, red, left y-axis) and robustness (TR, purple, right y-axis) of ratio control for a range of induction dosages. Relative to the low noise setting (light lines), high intracellular noise (dark lines) increases SR, leading to faster ratio control. TR is slightly reduced with increased noise, though this effect is less prominent. Both effects are larger at lower ATc concentrations. These curves divide the induction space into 5 regions. Regions A, C, and E are unsuitable for generating controlled ratios because cells do not transition (A), transition too slowly (C), or lack robust ratio control (E). Precise ratios can be attained in region B if noise is high enough to increase SR to an acceptable level. Region D is suitable for ratio control if both SR and TR are large enough, determined empirically. Here, region D meets the criteria SR > 10-0.9 and TR > 3 in the low noise setting.

Because of the long transition times in the low noise setting, the ratio control strategy for these cells becomes transient induction outside the bistable region. To compare the impacts of noise from multiple perspectives, we developed criteria by which to measure system responsiveness. Stimulus responsivity ($SR = 1/T_{30}$) is a measure of the speed of transitions in response to a stimulus, higher SR value means faster ratio control.

Along with TR, both metrics can be calculated for pulsed inductions as used in Figure 5-4.

Both TR and SR were computed for simulations of pulsed inductions for various inducer concentrations in both low and high noise conditions. As can be seen in Figure 5-9b, while the system's responsiveness (SR, red lines) increases asymptotically as the dose increases, tolerance to error (TR, purple lines) decreases. For a low dose of 15 ng/mL, SR increases from 10^{-1} at low noise (light lines) to $10^{-0.9}$ with high noise (dark lines), whereas TR decreases from 35 to 28 hours. At 20 ng/mL, SR increases from $10^{-0.95}$ to $10^{-0.85}$ while TR sees little change with increasing noise. This is consistent with the findings from Figure 5-9a, which showed that differences in response time due to noise were most apparent within the bistable region. Generally, noise causes a noticeable increase in SR at doses below 35 ng/mL, but differences in TR are only notable below 20 ng/mL. The reason for this is shown in Figure 5-10, which indicates that increasing noise shifts the dose/duration curve left, decreasing SR, but only mildly increases the curve's slope, which determines TR.

These data divide the system's induction range into 5 broad categories, labeled A-E. Regions A, C, and E are not ideal for either method of ratio control, but regions B and D (shaded gray) may be appropriate under certain conditions. Region A is unsuitable for ratio control because there is neither enough noise nor induction strength to cause a substantial fraction of the population to turn on. The boundaries of region B are fixed mathematically by the system's parameters and indicate the range of constant induction which may result in a broad range of precisely tuned ratios. However, as discussed, large enough expression noise is required to attain these ratios within a reasonably short time

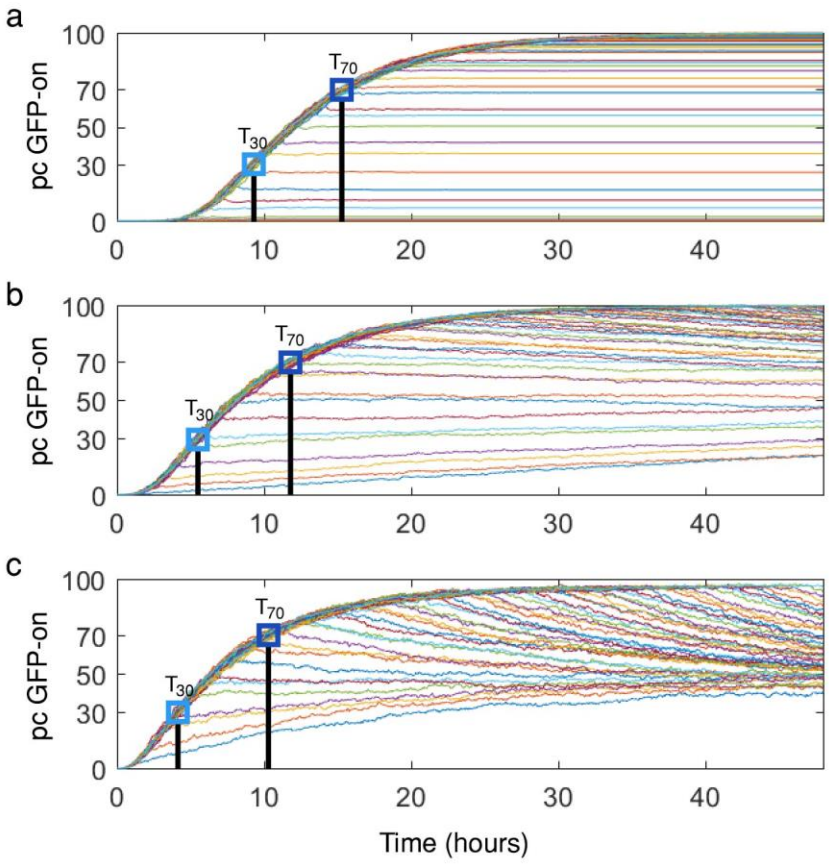

Figure 5-10 Pulsed induction simulations

Noise effects on pulse-induced ratio tuning, for 20 ng/mL ATc pulses with pulses ending every hour from 1 to 48 hours. T30 and T70 are the times at which the dose results in 30% or 70% of the population transitioning to the high GFP state, respectively. (a) In a low-noise system ($\Omega = 10$), cells maintain the fate their chosen fate after the induction pulse has finished. (b) Under mid-noise conditions ($\Omega = 2$), cells begin transitioning to the steady state of the post-pulse conditions, 8 ng/mL ATc. (c) In a high noise system ($\Omega = 1$), post-pulse transitioning is very apparent.

frame. Regions C-E cover induction levels only appropriate for pulsed induction, because constant induction at these levels will eventually cause most cells to turn on. The borders separating these three regions are determined by experimental constraints on SR and TR: the region's left edge set by SR, and its right edge set by TR. In region C, the system is highly tunable but responds so slowly as to not be viable. In region E, the response time is fast, but tuning accuracy is lost. Region D finds a compromise between

SR and TR constraints and is therefore appropriate for pulsed induction ratio control. Here, the constraints of SR > $10^{-0.9}$ and TR > 3 hours provide a good inducer range that balances speed and accuracy for the low noise condition.

Finally, region B also plays an important role in pulsed induction contexts. As seen in the mid- and high-noise simulations in Figure 5-10, once the pulse is finished, noise will cause the system to continue transitioning to the ratio defined by the final induction level. While the low-noise condition essentially "locks" in the chosen cell fate, higher noise levels increase the speed at which the population converges to the constant induction steady state. While many natural systems have built in mechanisms to "lock" in cell fate (209), in the absence of such mechanisms the final induction level chosen after a pulse is important for fulfilling that role.

5.3 DISCUSSION

In this work, we have used the mutual inhibition toggle switch – a synthetic version of a common genetic memory motif – to explore the temporal aspects of differentiation. Extracellular factors can drive a fraction of a population to switch phenotype and these factors interplay with intracellular noise to regulate population response. In the case of *E. coli*, with noisy gene expression dynamics, applying constant induction within the bistable region was enough to elicit state-switching. With the less noisy kinetics observed in *S. cerevisiae*, on the other hand, cells retained their steady state behavior unless temporarily forced out of the bistable region with a pulse of induction. Through both methods, it was demonstrated that control of population ratios could be achieved with a high degree of precision.

In the case of transiently introduced pulses of extracellular factors, we show that there exists an inverse relationship between the strength of the stimulus and duration for which it is administered. A target ratio can be achieved reliably either with a strong pulse for a short duration or vice versa. Under this framework, the only limitation for achieving precise ratio control is the temporal resolution imposed by the physical constraints in removing cells from the forcing stimulus. We also observed a tradeoff between system response speed and tuning robustness, quantified in the concepts of SR and TR. Lower pulse doses tended to be more robust to temporal variation in pulse length but required long durations to reach a desired ratio, while the opposite was true for high doses. Furthermore, gene expression noise acts as a global regulator of state switching. Increasing noise reduces transition times for both constant and pulsed inductions, increasing speed but resulting in a small reduction in tuning accuracy. Through this mechanism, noise levels determine whether constant induction will cause transitions within a reasonable time and determine the final dose at which pulsed induction should be set. Intracellular noise, therefore, is integral for the choice of ratio control strategy as well as for the chosen strategy's implementation.

In addition to the specific network studied, we also introduce a framework by which to analyze complex cellular behaviors involving a temporal component. Timing of cellular processes is becoming an increasingly important area of study (210, 211), and analogies can be drawn to applied fields of study. Though significantly more complex, the methods for deriving specific phenotypes from pluripotent progenitors are similar to the method we employ here (197, 198). Pluripotent cells are grown in a cocktail of growth factors for specific periods of time, sometimes sequentially, to force

differentiation down a desired path. While this work has been biologically and empirically driven in the past, we suggest that a mathematical approach may yield further insight into directed differentiation methods.

Additionally, while we demonstrate that noise plays an important role in shaping network behavior, practical ways to adjust noise in biological system is poorly understood. Noise can be tuned with ease in computational models (106, 212), but modification of noise in a real biological system is less common (213–215). Much useful information about how genetic networks respond to naturally stochastic expression could be gleaned by exploring all facets of this topic: the role of intrinsic vs. extrinsic noise, gene copy number vs. transcriptional or translational stochasticity, as well as the overall noise level from these various sources. Along with these dynamical concerns, network motifs – such as auto-activation (149) – may be used to shape the system response by altering noise profiles.

Through this work we have shown the importance in considering the temporal evolution and expression noise of a system when analyzing its differentiation dynamics. We developed and leveraged a mathematical understanding of the bistable toggle switch to achieve robust control of fractional differentiation ratios. Further work along these lines could have wide-ranging applications in countering bacterial persistence, developmental or stem cell biology, therapeutics, and provide guidance for *de novo* gene network synthesis.

5.4 MATERIALS AND METHODS

Plasmids and Cell Strains

E. coli experiments were performed with K12 MG1655 (American Type Culture Collection, ATCC, #700926) modified with (ΔLacI ΔAraC) deletions. The toggle switch plasmid, pKDL, was provided generously as a gift from James Collins (204). All yeast experiments were performed in YPH500 cells (Stratagene). The genomically integrated toggle system was developed previously by our group in collaboration with James Collins (82).

Flow Cytometry and Data Analysis

All cell measurements were taken with a Becton Dickinson (BD) Accuri C6 flow cytometer. Front scatter (FSC-A) and side scatter (SSC-A) were used to gate cellular populations. Only a vary course gating was used which removed debris smaller than the cell size but maintained the full range of population size variation. Samples were run on high flow rate to 10,000 captured events. The FL-1 channel (488 nm excitation; 530±15 nm filter) was used to measure GFP fluorescence. Data were analyzed using MATLAB (Mathworks, Inc.) run on a personal computer.

E. coli Experiments

E. coli were maintained in Luria Broth (LB) media with kanamycin (Sigma Aldrich). During experiments, cultures were diluted and inducers were refreshed hourly to avoid overgrowth. For hysteresis experiments, initially on and initially off cells were inoculated into media containing varying ATc concentrations, and GFP expression was

measured at 2, 5, and 8 hours. For pulsed induction experiments, initially off cells were given a dose of ATc, and a portion of the culture was diluted down to 0.125 ng/mL ATc (within the noise-resistant portion of the bistable region) every 30 minutes for 3 hours, with final measurements at 6 hours.

The pKDL toggle switch was transformed into the E. coli using a transformation kit (Zymo Research) and selected for by plating on Luria Broth (LB) agar (Sigma Aldrich) plates with added kanamycin (Sigma Aldrich). Cells were picked the day prior to performing experiments and cultured in 5mL LB medium (Sigma Aldrich) with kanamycin (Sigma Aldrich). The following day, cell density was measured with flow cytometry and diluted to 50 cells/μL in fresh LB media with kanamycin. The cells were rediluted every hour to avoid overgrowth and monitor healthy growth. For hysteresis experiments, after 2 hours, cells were rediluted into medium with anhydrotetracycline (ATc; Sigma Aldrich) forcing them to the initial off (ATc = 0; the cells favored the off state after overnight growth) or initial on (ATc = 20 ng/mL) states. These were maintained with hourly dilutions for 3 hours, then rediluted in medium containing variable ATc levels (0, 0.5, 1, 2, 4, 8, 12, 20 ng/mL). When diluting from a high ATc concentration to a lower one, medium volumes of different concentrations were mixed to avoid the potential shock of centrifugation and washing. For example, to go from 20 ng/mL to 4 ng/mL ATc, 1 part of the original culture was added to 4 parts at 0 ng/mL, yielding a final concentration of 4 ng/mL. To dilute from initial on to 0 ng/mL, cells were centrifuged and washed with fresh LB before redilution. These cultures were maintained with hourly redilutions, and portions of the culture were run on flow cytometry at 2, 5, and 8 hours. For dose/duration experiments, cells were again forced to

their initial state for 3 hours, then rediluted with medium containing the desired dose of ATc. Every 30 minutes, a portion of the culture was run on flow cytometry and another portion was rediluted to within the toggle's bistable range (0.125 ng/mL) using a similar fractional volume method of increasing or decreasing the concentration. The longest pulse given was 4 hours, and all cultures were maintained with hourly dilutions. All cultures were tested via flow cytometry 4 hours after the beginning of the pulse and again 1 hour later to unsure that the population had reached steady state expression.

S. cerevisiae Experiments

Yeast were grown in 2% glucose YPD (Sigma Aldrich) ensure healthy growth, then transferred to yeast medium containing 2% Galactose (Sigma Aldrich) and 1% Raffinose (Sigma Aldrich). During experiments, cultures were diluted every 6 or 12 hours to avoid overgrowth. For hysteresis experiments, initially on and initially off cells were inoculated into media containing varying ATc concentrations, and GFP expression was measured at 24 hours. For pulsed induction experiments, initially off cells were given a dose of ATc, and a portion of the culture was diluted down to 8 ng/mL ATc (within the noise-resistant portion of the bistable region) for pulses between 2 and 24 hours, with final measurements at 36 hours.

A single copy of the toggle switch was integrated into the yeast genome as described in previous work (82). Confirmed clones were streaked onto 2% glucose YPD agar plates (Sigma Aldrich). Colonies were picked from these plates 42 hours prior to the start of the experiment and grown in 5 mL YPD medium. After 8 hours, cultures were monitored by flow cytometry and rediluted to 1500 cells/mL in fresh YPD and allowed to

grow overnight. After 12 hours, cells were measured again and rediluted to 5000 cells/mL into yeast medium with 2% galactose (Sigma Aldrich) and 1% raffinose (Sigma Aldrich) with appropriate ATc to induce the initial off (0 ng/mL ATc) or initial on (50 ng/mL) states. These were measured and rediluted again 12 hours later and allowed to grow overnight before beginning the experiment. For hysteresis experiments, initial off and initial on cells were diluted into varying concentrations of ATc (0, 0.5, 1, 2, 4, 8, 14, 20, 30, 40, 50 ng/mL). Cells were diluted to 5000 cells/mL every 12 hours and measured via flow cytometry at 24 and 48 hours. For dose/duration experiments, cultures were induced with ATc and at the end of each duration a portion of the culture was measured with flow cytometry and a portion was diluted down to an ATc concentration within the toggle's bistable range (8 ng/mL). Cells were measured and rediluted every 12 hours after the start of the initial dose, with the final measure being at least 12 hours after the end of the initial dose, to ensure that cells had reached steady state expression.

Modeling

We used the model proposed in previous work (82). The ODEs are:

$$[L]' = \tau \cdot \left\{ crl + \frac{1.0}{1.0 + \left[\left(\frac{[T]}{kt}\right) \cdot \left(1.0 + \frac{ATc}{katc} * \frac{kt}{[T]}\right)^{-m} \right]^{nt}} \cdot (cil - crl) - delta \cdot [L] \right\} \qquad \text{eqn. (5-1)}$$

$$[T]' = \tau \cdot \left\{ crt + \frac{1.0}{1.0 + \left(\frac{[L]}{kl}\right)^{nl}} \cdot (cit - crt) - delta \cdot [T] \right\} \qquad \text{eqn. (5-2)}$$

Where [L] and [T] are the concentration of LacI and TetR. LacI is coexpressed with GFP, and thus it was used interchangeably. crl and cil are the production rate of LacI when the promoter is repressed or induced respectively, while crt and cit are the

Reaction	Description	Propensity function
Φ→LacI	Production rate of LacI	$Pl \cdot \Omega$
LacI→Φ	Degradation rate of LacI	$\tau *$ delta
Φ→TetR	Production rate of TetR	$Pt \cdot \Omega$
TetR →Φ	Degradation rate of TetR	$\tau *$ delta
$Pl = \tau \cdot \left\{ crl + \frac{1.0}{1.0 + \left[\left(\frac{[T]}{kt} \right) * \left(1.0 + \frac{ATc}{katc} * \frac{kt}{[T]} \right)^{-m} \right]^{nt}} \cdot (cil - crl) \right\}$		
$Pt = \tau \cdot \left\{ crt + \frac{1.0}{1.0 + \left(\frac{[L]}{kl} \right)^{nl}} \cdot (cit - crt) \right\}$		

Table 5-1 Stochastic transition processes and the corresponding transition rates

production rate of TetR when the promoter is repressed or induced respectively. kt represents the active TetR concentration needed to make this probability 50%, and nt describes the nonlinearity of this inhibition. kl represents the LacI concentration needed to make the promoter bound by LacI 50% of the time, and nl describes the nonlinearity of this inhibition. m is the Hill coefficient of the Hill function, which is used to describe the relationship between the active ratio of repressor TetR and the ATc inducer concentration. Here τ is the timescale of the system. The detail of the model construction can be found in (82).

For the stochastic simulation, the concentration of each molecular is converted to its number, i.e., $x = [x] \cdot \Omega$, where Ω is a system size factor. Table 5-1 lists all the reactions involved. τ-leap-based stochastic Gillespie algorithm is used for the stochastic simulation. The noise level is set by Ω, which is set to 1 for E. coil system and 10 for S. cerevisiae. The bifurcation diagrams are generated with Oscill8 (http://oscill8.sourceforge.net/).

We searched the parameter space with a customized Metropolis algorithm to fit various experimental data, including the hysteresis curve (Fig. 5-1b, 5-4a), the fraction of (Fig. 5-1e, 5-6a). It is noted that the parameter set for *E. coli* and Yeast are not same since

we found different hysteresis curves in the two systems. It is an irreversible bistable switch in *E. coli* (Fig. 5-1b), while it is a reversible bistable switch in S. cerevisiae (Fig. 5-4a). The fitted parameters can be found in Table 5-2. In Fig. 5-9, we used the parameters for *S. cerevisiae* under different noise levels to study the general strategies to achieve predictable ratio control.

	E.coli	S. cerevisiae
Timescale	1100	900
crl	0.001	0.001
crt	0.001	0.001
cil	0.08	0.08
cit	0.08	0.08
delta	0.002	0.002
m	3	3
nt	1.5	1.5
nl	3.5	3.5
kt	15	15
kl	19	20
katc	3.5	25
Ω	1	10

Table 5-2 Parameters of the model

6 CONCLUSION

6.1 DISCUSSION

The field of synthetic biology is moving toward a quantitative understanding of biological systems, one in which biological networks can be engineered from the ground up to perform desired tasks. In this work, I have detailed my (and my collaborators') steps toward realizing this goal. After presenting generally on the mathematical methods which can be employed to understand complex networks, I presented two bodies of original research.

In our CRISPR work, we introduce a new tool, the fgRNA, for visualizing the dynamics of RNA components of genetic networks. While we utilized them in the context of CRISPR transcriptional regulation, similar approaches could be used to model and predict the dynamics of a range of cellular RNA components. In addition to adding a new tool to the synthetic biology toolbox, we demonstrate several important findings. First, we show the power of circuit optimization in obtaining strong on/off ratios from genetic circuits. Particularly in human cell work, it is common to see an "add more plasmid to get better results" approach to transfection optimization. While this may work for simple up- or down-regulation, our results clearly demonstrate that as the network becomes more complex, even mildly so, this mentality does not hold up. Second, we demonstrated several ways that RNA Pol II transcripts could be modified and edited to improve their functionality for CRISPR applications. Lower transcription rates and limitations imposed by editing sequences make this a non-trivial problem, but we managed to improve efficiency enough that we could see information transmission

through a two-node, fully Pol II, CRISPR repressor cascade. Third, our modeling methods, thanks to the visibility of the guide RNA, reveal underlying parameters and trends of the component parts of the networks in use. This allowed us to quantify the difference in expression between Pol II and Pol III promoters, accounting for mRNA editing efficiency, accurately and easily. It also allows us some insight into the role of mRNA export in the editing process, suggesting further routes of research to improve Pol II gRNA expression. Lastly, it also let us confirm what was suspected about dCas9 as a transcriptional regulator: it lacks nonlinearity. Despite the obvious benefits of dCas9 as a regulator chassis, confirmation of this drawback reinforces the fact that there is still much work to be done before it can be implemented in all theorized regulatory applications.

In our ratio control work, we developed experimental and mathematical methods for understanding and predicting population level differentiation events. While much work has been done, particularly in stem cell related areas, on factors which determine differentiation, there is relatively little on the process of differentiation itself or the mathematical underpinnings thereof. Here's we integrate temporal information to better understand the differentiation process and utilize that information to control the fraction of a population which undergoes a differentiation event. Doing this required use of exclusively stochastic modeling techniques, as the process we studied was uniquely reliant on gene expression noise. We show that populations can be made to fractionally differentiate with a high degree of accuracy, and we develop metrics by which one can choose between two revealed mechanisms for achieving precise ratio control. While still a first step, a similar approach may eventually be used to better understand natural

differentiation pathways, possibly for control of stem cell differentiation in laboratory settings or for therapeutic purposes.

6.2 FUTURE DIRECTIONS

There are a multitude of questions in biological engineering that remain unanswered, and myriad directions in which future research could unfold. Here, I will cover some of those which stand out to me as most pressing or interesting topics for further pursuit.

On the biological side of synthetic biology, there remains a need for genetic part diversification and standardization. Whereas mechanical engineers may have their choice from thousands of similar but unique components – screws, girders, pipes, etc. – to fit to any specific application, bioengineers are still working with a handful of reliable genes, promoters, and terminators. The lack of components makes building larger circuits unfeasible due to the inevitable crosstalk between reused parts, and limits the tuneability of circuits, as many components operate at only a single "strength." Additional research into promoter design to allow researchers to build them from the ground up, taking into account such factors as regulatory operator domains, repressed, unrepressed and activated transcriptional rates, and insulation from neighboring genes' transcription, is of increasing importance. Relatedly, a gene's transcription is heavily influenced by the epigenetic landscape surrounding it. Building a greater understanding of the many processes which determine epigenetic state will aid in ultimately engineering reliable transcriptional regulatory networks.

There are similar concerns for transcription factors, with a lack of targetability to promoters and an overall lack of designability. There are less than a dozen well

characterized repressors and about the same number of activator domains. Zinc fingers and TALENs have be developed specifically to overcome this dearth of functional proteins; however, design difficulty and protein size have been limiting factors for adopting these methods more widely (216, 217). Utilization of Cas9 or dCas9 as a transcriptional regulator overcomes the difficulty of design issues, but introduces wrinkles of its own, as I demonstrated in the work here. While Cas9 is a protein, the gRNA that directs it is not, and must therefore either being expressed from a Pol III promoter – which lacks many of the qualities we would like in a promoter – or must be editing from a Pol II promoter's transcript – which is inefficient for gRNA production. Additionally, Cas9 lacks the cooperativity observed in many natural regulatory proteins, which allows regulation to respond sigmoidally to increasing regulator concentrations and is necessary for signal propagation through larger networks. Until these issues are resolved, Cas9 will only be appropriate for smaller circuits, despite its clear benefits for designability and orthogonality.

In addition to concerns about lacking components is a wider reproducibility concern for biological circuit construction techniques. The genetic code, while only consisting of 4 bases, is incredibly complex and still poorly understood. Particularly in non-coding regions of DNA, the effect of specific base changes is unknown, and we are likely still far from being able to design specific functional structures – like promoters or insulators – from first principles. For this reason, there is still much variation in results between groups who are ostensibly studying the same phenomena. We demonstrate this in our CRISPR work while working with Csy4 editing: for unknown reasons the sequence was resistant to any of the changes that we implemented, despite the

modifications having a noticeable effect on RGR transcripts. Thankfully, plasmid repositories, such as Addgene, have made gaining access to specific genes and sequences more convenient than ever, but the need still remains for a better understanding of how DNA sequences affects expression dynamics.

Mathematically speaking, standardization in modeling methods would aid in reproducibility and applicability of the insights gained from modeling. Currently, because the forms of models vary so much between applications, transference of parameter values between applications. There has been some work with universalizing experimental results, such as applying standardization to flow cytometry results, which is an important start (218, 219). The variability in modeling methods, various ways to simplify models, and different applications of models makes their standardization a non-trivial task. If the models themselves cannot be standardized, it may be worthwhile develop a conversion method so that parameters can still be transferred from one method or formulation to another. As things stand, aside from general correlations, and qualitative behavioral information, little can be taken away from a vast number of models, and it is very difficult to build a model from a prior model's foundation. Researchers essentially start over from scratch when approaching a circuit, even if that circuit has been studied extensively before. Standardization of modeling methods would alleviate this issue and help in standardization of the biological components themselves, as all parts' behavior could be quantified via a known metric.

In many ways, all of these potential future directions for research come back to single unifying goal: transforming biology from an exploratory science to an actionable engineering discipline. Synthetic biology is often touted as engineering, but the reality is

that it is still dependent on more trial and error than is desirable for a true engineering discipline. While this is understandable for a field which has essentially come into being in the last 20 years, it is important to not delude ourselves into believing that our field is more quantitative, predictable, or predictive than it truly is. Only by engaging with its shortcomings can we make the field into the engineering disciple that we know it can be.

REFERENCES

1. Nan B, Zusman DR (2016) Novel mechanisms power bacterial gliding motility. *Mol Microbiol* 101(2):186–193.

2. Likhoshvai VA, Khlebodarova TM (2014) Mathematical modeling of bacterial cell cycle: the problem of coordinating genome replication with cell growth. *J Bioinform Comput Biol* 12(3):1450009.

3. Wang JD, Levin PA (2009) Metabolism, cell growth and the bacterial cell cycle. *Nat Rev Microbiol* 7(11):822–827.

4. Richardson AR, Somerville GA, Sonenshein AL (2015) Regulating the Intersection of Metabolism and Pathogenesis in Gram-positive Bacteria. *Microbiol Spectr* 3(3). doi:10.1128/microbiolspec.MBP-0004-2014.

5. Bäumler AJ, Sperandio V (2016) Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535(7610):85–93.

6. Langer S, Schropp D, Bengelsdorf FR, Othman M, Kazda M (2014) Dynamics of biofilm formation during anaerobic digestion of organic waste. *Anaerobe* 29:44–51.

7. Verstrepen KJ, Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. *Mol Microbiol* 60:5–15.

8. Vital-Lopez FG, Reifman J, Wallqvist A (2015) Biofilm Formation Mechanisms of Pseudomonas aeruginosa Predicted via Genome-Scale Kinetic Models of Bacterial Metabolism. *PLoS Comput Biol* 11(10):e1004452.

9. Xu Z, Islam S, Wood TK, Huang Z (2015) An Integrated Modeling and Experimental Approach to Study the Influence of Environmental Nutrients on Biofilm Formation of Pseudomonas aeruginosa. *BioMed Res Int* 2015:506782.

10. Duan Q, Zhou M, Zhu L, Zhu G (2013) Flagella and bacterial pathogenicity. *J Basic Microbiol* 53(1):1–8.

11. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial Persistence as a Phenotypic Switch. *Science* 305(5690):1622–1625.

12. Potter CMF, Lao KH, Zeng L, Xu Q (2014) Role of biomechanical forces in stem cell vascular lineage differentiation. *Arterioscler Thromb Vasc Biol* 34(10):2184–2190.

13. Carobbio S, Rosen B, Vidal-Puig A (2013) Adipogenesis: new insights into brown adipose tissue differentiation. *J Mol Endocrinol* 51(3):T75-85.

14. Johe KK, Hazel TG, Muller T, others (1996) Single factors direct the differentiation of stem cells from the fetal and adult central nervous system. *Genes Dev* 10:3129–3140.

15. Noisa P, et al. (2014) Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. *J Cell Sci* 127(Pt 9):2083–2094.

16. Leitão AB, Sucena É (2015) Drosophila sessile hemocyte clusters are true hematopoietic tissues that regulate larval blood cell differentiation. *eLife* 4. doi:10.7554/eLife.06166.

17. Stumpf PS, et al. (2017) Stem Cell Differentiation as a Non-Markov Stochastic Process. *Cell Syst* 5(3):268-282.e7.

18. Wu F, Menn DJ, Wang X (2014) Quorum-sensing crosstalk-driven synthetic circuits: from unimodality to trimodality. *Chem Biol* 21(12):1629–1638.

19. Szymańska Z, Parisot M, Lachowicz M (2014) Mathematical modeling of the intracellular protein dynamics: the importance of active transport along microtubules. *J Theor Biol* 363:118–128.

20. Mathivanan S, Ji H, Simpson RJ (2010) Exosomes: extracellular organelles important in intercellular communication. *J Proteomics* 73(10):1907–1920.

21. Al-Shboul O, Mustafa A (2015) Effect of oxidative stress on Rho kinase II and smooth muscle contraction in rat stomach. *Can J Physiol Pharmacol* 93(6):405–411.

22. Castorena CM, Arias EB, Sharma N, Bogan JS, Cartee GD (2015) Fiber type effects on contraction-stimulated glucose uptake and GLUT4 abundance in single fibers from rat skeletal muscle. *Am J Physiol Endocrinol Metab* 308(3):E223-230.

23. Shimizu K, Fujita H, Nagamori E (2013) Evaluation systems of generated forces of skeletal muscle cell-based bio-actuators. *J Biosci Bioeng* 115(2):115–121.

24. Ghazalpour A, et al. (2011) Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLOS Genet* 7(6):e1001393.

25. Bheda P, Schneider R (2014) Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol* 24(11):712–723.

26. Huang B, Jiang C, Zhang R (2014) Epigenetics: the language of the cell? *Epigenomics* 6(1):73–88.

27. Veening J-W, et al. (2008) Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci* 105(11):4393–4398.

28. Bornstein C, et al. (2014) A negative feedback loop of transcription factors specifies alternative dendritic cell chromatin States. *Mol Cell* 56(6):749–762.

29. Ferrell JE (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14:140–8.

30. Gamba P, Jonker MJ, Hamoen LW (2015) A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence. *PLOS Genet* 11(6):e1005047.

31. Marciano DC, et al. (2014) Negative feedback in genetic circuits confers evolutionary resilience and capacitance. *Cell Rep* 7(6):1789–1795.

32. Mengel B, Krishna S, Jensen MH, Trusina A (2012) Nested feedback loops in gene regulation. *Phys Stat Mech Its Appl* 391(1–2):100–106.

33. Coulombe-Huntington J, Xia Y (2012) Regulatory Network Structure as a Dominant Determinant of Transcription Factor Evolutionary Rate. *PLoS Comput Biol* 8(10):e1002734.

34. Wilczynski B, Liu Y-H, Yeo ZX, Furlong EEM (2012) Predicting Spatial and Temporal Gene Expression Using an Integrative Model of Transcription Factor Occupancy and Chromatin State. *PLoS Comput Biol* 8(12):e1002798.

35. Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 94:128701.

36. Davidson EH, et al. (2002) A Genomic Regulatory Network for Development. *Science* 295(5560):1669–1678.

37. Faith JJ, et al. (2007) Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol* 5(1):e8.

38. Behar M, Barken D, Werner SL, Hoffmann A (2013) The Dynamics of Signaling as a Pharmacological Target. *Cell* 155(2):448–461.

39. Pattanayak V, et al. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31(9):839–843.

40. Mukherji S, van Oudenaarden A (2009) Synthetic biology: understanding biological design from synthetic circuits. *Nat Rev Genet* 10:859–71.

41. Wang L-Z, Wu F, Flores K, Lai Y-C, Wang X (2016) Build to understand: synthetic approaches to biology. *Integr Biol* 8(4):394–408.

42. Kis Z, Pereira HS, Homma T, Pedrigi RM, Krams R (2015) Mammalian synthetic biology: emerging medical applications. *J R Soc Interface* 12(106). doi:10.1098/rsif.2014.1000.

43. Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* 27(5):465–471.

44. Friedland AE, et al. (2009) Synthetic gene networks that count. *Science* 324(5931):1199–1202.

45. Nielsen AA, Voigt CA (2014) Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Mol Syst Biol* 10(11):n/a-n/a.

46. Grilly C, Stricker J, Pang WL, Bennett MR, Hasty J (2007) A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Mol Syst Biol* 3. doi:10.1038/msb4100168.

47. Menn DJ, Su R-Q, Wang X (2017) Control of synthetic gene networks and its applications. *Quant Biol* 5(2):124–135.

48. Kærn M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6(6):451–464.

49. Balazsi G, van Oudenaarden A, Collins JJ (2011) Cellular decision making and biological noise: from microbes to mammals. *Cell* 144:910–25.

50. Menn David J, Wang Xiao (2014) Stochastic and Deterministic Decision in Cell Fate. *eLS*. doi:10.1002/9780470015902.a0025319.

51. Elowitz MB (2002) Stochastic Gene Expression in a Single Cell. *Science* 297(5584):1183–1186.

52. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci* 99(20):12795–12800.

53. Choi PJ, Xie XS, Shakhnovich EI (2010) Stochastic Switching in Gene Networks Can Occur by a Single-Molecule Event or Many Molecular Steps. *J Mol Biol* 396(1):230–244.

54. Josić K, López JM, Ott W, Shiau LJ, Bennett MR (2011) Stochastic Delay Accelerates Signaling in Gene Networks. *PLoS Comput Biol* 7(11):e1002264.

55. Colman-Lerner A, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.

56. Gander MW, Vrana JD, Voje WE, Carothers JM, Klavins E (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun* 8. doi:10.1038/ncomms15459.

57. Daniel R, Rubens JR, Sarpeshkar R, Lu TK (2013) Synthetic analog computation in living cells. *Nature* advance online publication. doi:10.1038/nature12148.

58. Ebert MS, Sharp PA (2012) Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* 149(3):515–524.

59. Wang J, Xu L, Wang E (2008) Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc Natl Acad Sci U S A* 105(34):12271–12276.

60. Kim K-Y, Wang J (2007) Potential energy landscape and robustness of a gene regulatory network: toggle switch. *PLoS Comput Biol* 3(3):e60.

61. Ma W, Lai L, Ouyang Q, Tang C (2006) Robustness and modular design of the Drosophila segment polarity network. *Mol Syst Biol* 2:70.

62. Antoni D, Zverlov VV, Schwarz WH (2007) Biofuels from microbes. *Appl Microbiol Biotechnol* 77(1):23–35.

63. Dellomonaco C, Fava F, Gonzalez R (2010) The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microb Cell Factories* 9:3.

64. Harrison ME, Dunlop MJ (2012) Synthetic Feedback Loop Model for Increasing Microbial Biofuel Production Using a Biosensor. *Front Microbiol* 3. doi:10.3389/fmicb.2012.00360.

65. Krom RJ, Bhargava P, Lobritz MA, Collins JJ (2015) Engineered Phagemids for Nonlytic, Targeted Antibacterial Therapies. *Nano Lett* 15(7):4808–4813.

66. Sufya N, Allison DG, Gilbert P (2003) Clonal variation in maximum specific growth rate and susceptibility towards antimicrobials. *J Appl Microbiol* 95:1261–7.

67. DeLorenzo ME, Fleming J (2008) Individual and Mixture Effects of Selected Pharmaceuticals and Personal Care Products on the Marine Phytoplankton Species Dunaliella tertiolecta. *Arch Environ Contam Toxicol* 54(2):203–210.

68. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotechnol* 21(7):796–802.

69. Ellis T, Adie T, S. Baldwin G (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol* 3(2):109–118.

70. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.

71. Densmore D, et al. (2010) Algorithms for automated DNA assembly. *Nucleic Acids Res*:gkq165.

72. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26(10):1135–1145.

73. Chavez A, et al. (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12(4):326–328.

74. Kiani S, et al. (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* 11(7):723–726.

75. Standage-Beier K, Zhang Q, Wang X (2015) Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. *ACS Synth Biol* 4(11):1217–1225.

76. Guido NJ, et al. (2006) A bottom-up approach to gene regulation. *Nature* 439(7078):856–860.

77. Stricker J, et al. (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* 456:516–9.

78. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767):335–338.

79. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* 403(6767):339–342.

80. Shimizu Y, et al. (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19(8):751–755.

81. Pardee K, et al. (2016) Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* 165(5):1255–1266.

82. Wu M, et al. (2013) Engineering of regulated stochastic cell fate determination. *Proc Natl Acad Sci* 110(26):10610–10615.

83. Xiong W, Ferrell JE (2003) A positive-feedback-based bistable “memory module” that governs a cell fate decision. *Nature* 426:460–5.

84. Potvin-Trottier L, Lord ND, Vinnicombe G, Paulsson J (2016) Synchronous long-term oscillations in a synthetic gene circuit. *Nature* 538(7626):514–517.

85. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* 101(17):6355–6360.

86. Liu C, et al. (2011) Sequential Establishment of Stripe Patterns in an Expanding Cell Population. *Science* 334(6053):238–241.

87. Payne S, et al. (2013) Temporal control of self-organized pattern formation without morphogen gradients in bacteria. *Mol Syst Biol* 9(1). doi:10.1038/msb.2013.55.

88. Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. *Nature* 491(7423):249–253.

89. Yang L, et al. (2014) Permanent genetic memory with >1-byte capacity. *Nat Methods* 11(12):1261–1266.

90. Ishimatsu K, et al. (2014) General applicability of synthetic gene-overexpression for cell-type ratio control via reprogramming. *ACS Synth Biol* 3(9):638–644.

91. Yamaguchi M, Ito A, Ono A, Kawabe Y, Kamihira M (2014) Heat-inducible gene expression system by applying alternating magnetic field to magnetic nanoparticles. *ACS Synth Biol* 3(5):273–279.

92. Hussain F, et al. (2014) Engineered temperature compensation in a synthetic genetic clock. *Proc Natl Acad Sci U S A* 111(3):972–977.

93. Levskaya A, et al. (2005) Synthetic biology: engineering Escherichia coli to see light. *Nature* 438:441–2.

94. Levskaya A, Weiner OD, Lim WA, Voigt CA (2009) Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* 461(7266):997–1001.

95. Jogler C, Schüler D (2009) Genomics, genetics, and cell biology of magnetosome formation. *Annu Rev Microbiol* 63:501–521.

96. Shin Y-J, Bleris L (2010) Linear control theory for gene network modeling. *PloS One* 5(9). doi:10.1371/journal.pone.0012785.

97. Del Vecchio D, Dy AJ, Qian Y (2016) Control theory meets synthetic biology. *J R Soc Interface* 13(120). doi:10.1098/rsif.2016.0380.

98. Kalman RE (1963) Mathematical Description of Linear Dynamical Systems. *J Soc Ind Appl Math Control* 1(2):152–192.

99. Lin C-T (1974) Structural Controllability. *IEEE Trans Autom Control* 19(3):201–208.

100. Keasling JD (2010) Manufacturing molecules through metabolic engineering. *Science* 330(6009):1355–1358.

101. Koffas M, Roberge C, Lee K, Stephanopoulos G (1999) Metabolic engineering. *Annu Rev Biomed Eng* 1:535–557.

102. Polynikis A, Hogan SJ, di Bernardo M (2009) Comparing different ODE modelling approaches for gene regulatory networks. *J Theor Biol* 261:511–530.

103. Liu Y-Y, Slotine J-J, Barabási A-L (2011) Controllability of complex networks. *Nature* 473(7346):167–173.

104. Basler G, Nikoloski Z, Larhlimi A, Barabási A-L, Liu Y-Y (2016) Control of fluxes in metabolic networks. *Genome Res* 26(7):956–968.

105. Strogatz SH (2013) Exploring Complex Networks.

106. Faucon PC, et al. (2014) Gene Networks of Fully Connected Triads with Complete Auto-Activation Enable Multistability and Stepwise Stochastic Transitions. *PLoS ONE* 9(7):e102873.

107. Kim D hyun, Grün D, van Oudenaarden A (2013) Dampening of expression oscillations by synchronous regulation of a microRNA and its target. *Nat Genet* advance online publication. doi:10.1038/ng.2763.

108. Thorsley D, Klavins E (2012) Estimation and Discrimination of Stochastic Biochemical Circuits from Time-Lapse Microscopy Data. *PLoS ONE* 7(11):e47151.

109. Gillespie D (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.

110. Balagaddé FK, et al. (2008) A synthetic Escherichia coli predator–prey ecosystem. *Mol Syst Biol* 4(1):187.

111. Song H, Payne S, Gray M, You L (2009) Spatiotemporal modulation of biodiversity in a synthetic chemical-mediated ecosystem. *Nat Chem Biol* 5(12):929–935.

112. Song H, You L (2012) Modeling spatiotemporal dynamics of bacterial populations. *Methods Mol Biol Clifton NJ* 880:243–254.

113. Milo R, et al. (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298(5594):824–827.

114. Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* 138(4):760–773.

115. Mallet DG, De Pillis LG (2006) A cellular automata model of tumor-immune system interactions. *J Theor Biol* 239(3):334–350.

116. Huang S, Ernberg I, Kauffman S (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* 20(7):869–876.

117. Wells DK, Kath WL, Motter AE (2015) Control of Stochastic and Induced Switching in Biophysical Networks. *Phys Rev X* 5. doi:10.1103/PhysRevX.5.031036.

118. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A (2004) Multistability in the lactose utilization network of Escherichia coli. *Nature* 427(6976):737–740.

119. Leisner M, Kuhr J-T, Rädler JO, Frey E, Maier B (2009) Kinetics of Genetic Switching into the State of Bacterial Competence. *Biophys J* 96(3):1178–1188.

120. Dhar N, McKinney JD (2007) Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr Opin Microbiol* 10(1):30–38.

121. Lipshtat A, Loinger A, Balaban NQ, Biham O (2006) Genetic Toggle Switch without Cooperative Binding. *Phys Rev Lett* 96(18):188101.

122. Greber D, El-Baba MD, Fussenegger M (2008) Intronically encoded siRNAs improve dynamic range of mammalian gene regulation systems and toggle switch. *Nucleic Acids Res* 36(16):e101.

123. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U A* 100:7714–9.

124. Smits WK, et al. (2005) Stripping Bacillus: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol Microbiol* 56(3):604–614.

125. Yao G, Tan C, West M, Nevins JR, You L (2011) Origin of bistability underlying mammalian cell cycle entry. *Mol Syst Biol* 7:485.

126. Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic Properties of Network Motifs Contribute to Biological Network Organization. *PLOS Biol* 3(11):e343.

127. Singh V (2014) Recent advances and opportunities in synthetic logic gates engineering in living cells. *Syst Synth Biol* 8(4):271–282.

128. Nielsen AAK, et al. (2016) Genetic circuit design automation. *Science* 352(6281):aac7341.

129. Ausländer S, Ausländer D, Müller M, Wieland M, Fussenegger M (2012) Programmable single-cell mammalian biocomputers. *Nature* 487(7405):123–127.

130. Gaber R, et al. (2014) Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat Chem Biol* 10(3):203–208.

131. Mishra D, Rivera PM, Lin A, Del Vecchio D, Weiss R (2014) A load driver device for engineering modularity in biological networks. *Nat Biotechnol* 32(12):1268–1275.

132. Del Vecchio D (2013) A control theoretic framework for modular analysis and design of biomolecular networks. *Annu Rev Control* 37(2):333–345.

133. Harbauer AB, et al. (2014) Cell cycle–dependent regulation of mitochondrial preprotein translocase. *Science* 346(6213):1109–1113.

134. Feillet C, et al. (2014) Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proc Natl Acad Sci* 111(27):9828–9833.

135. Kim J, Khetarpal I, Sen S, Murray RM (2014) Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res*:gku233.

136. Ostojic S (2014) Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat Neurosci* 17(4):594–600.

137. Comb M, Hyman SE, Goodman HM (1987) Mechanisms of trans-synaptic regulation of gene expression. *Trends Neurosci* 10(11):473–478.

138. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* 457:309–12.

139. Xiao M, Cao J (2008) Genetic oscillation deduced from Hopf bifurcation in a genetic regulatory network with delays. *Math Biosci* 215(1):55–63.

140. Zakharova A, Vadivasova T, Anishchenko V, Koseska A, Kurths J (2010) Stochastic bifurcations and coherencelike resonance in a self-sustained bistable noisy oscillator. *Phys Rev E* 81(1):011106.

141. Lewis J (2003) Autoinhibition with Transcriptional Delay: A Simple Mechanism for the Zebrafish Somitogenesis Oscillator. *Curr Biol* 13(16):1398–1408.

142. Swinburne IA, Miguez DG, Landgraf D, Silver PA (2008) Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev* 22(17):2342–2346.

143. Izhikevich EM (2000) Neural excitability, spiking and bursting. *Int J Bifurc Chaos* 10(6):1171–1266.

144. Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 176(2):269–275.

145. Oates AC, Morelli LG, Ares S (2012) Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Dev Camb Engl* 139(4):625–639.

146. You L, Cox RS, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. *Nature* 428:868–71.

147. Balagaddé FK, You L, Hansen CL, Arnold FH, Quake SR (2005) Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* 309(5731):137–140.

148. Smith R, et al. (2014) Programmed Allee effect in bacteria causes a tradeoff between population spread and survival. *Proc Natl Acad Sci U S A* 111(5):1969–1974.

149. Wu F, Su R-Q, Lai Y-C, Wang X (2017) Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. *eLife* 6:e23702.

150. Johnson KA, Goody RS (2011) The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry (Mosc)* 50(39):8264–8269.

151. Weiss JN (1997) The Hill equation revisited: uses and misuses. *FASEB J*. doi:10.1096/fasebj.11.11.9285481.

152. Li H, Cao Y, Petzold LR, Gillespie DT (2008) Algorithms and Software for Stochastic Simulation of Biochemical Reacting Systems. *Biotechnol Prog* 24(1):56–61.

153. Adalsteinsson D, McMillen D, Elston TC (2004) Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMC Bioinformatics* 5:24.

154. Ma M, Ye AY, Zheng W, Kong L (2013) A Guide RNA Sequence Design Platform for the CRISPR/Cas9 System for Model Organism Genomes. *BioMed Res Int* 2013:e270805.

155. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32(3):279–284.

156. Farzadfard F, Perli SD, Lu TK (2013) Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth Biol* 2(10):604–613.

157. Nissim L, Perli SD, Fridkin A, Perez-Pinera P, Lu TK (2014) Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol Cell* 54(4):698–710.

158. Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing: Self-processing of ribozyme-flanked RNAs into guide RNAs. *J Integr Plant Biol* 56(4):343–349.

159. Przybilski R, et al. (2011) Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. *RNA Biol* 8(3):517–528.

160. Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci* 112(11):3570–3575.

161. Brockmann R, Beyer A, Heinisch JJ, Wilhelm T (2007) Posttranscriptional Expression Regulation: What Determines Translation Rates? *PLoS Comput Biol* 3(3):e57.

162. Raj A, Bogaard P van den, Rifkin SA, Oudenaarden A van, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5(10):877–879.

163. Ma H, et al. (2016) Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* 34(5):528–530.

164. Song W, Strack RL, Svensen N, Jaffrey SR (2014) Plug-and-Play Fluorophores Extend the Spectral Properties of Spinach. *J Am Chem Soc* 136(4):1198–1201.

165. Filonov GS, Moon JD, Svensen N, Jaffrey SR (2014) Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J Am Chem Soc* 136(46):16299–16308.

166. Paige JS, Wu KY, Jaffrey SR (2011) RNA Mimics of Green Fluorescent Protein. *Science* 333(6042):642–646.

167. Zalatan JG, et al. (2015) Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* 160(1–2):339–350.

168. Shechner DM, Hacisuleyman E, Younger ST, Rinn JL (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* 12(7):664–670.

169. Lloyd DR, Holmes P, Jackson LP, Emery AN, Al-Rubeai M (2000) Relationship between cell size, cell cycle and specific recombinant protein productivity. *Cytotechnology* 34(1–2):59–70.

170. Becskei A, Serrano L (2000) Engineering Stability in Gene Networks by Autoregulation. *Nature* 405(6786):590–593.

171. Gossen M, et al. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268(5218):1766–1769.

172. Nehlin JO, Carlberg M, Ronne H (1991) Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. *EMBO J* 10(11):3373–3377.

173. Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U A* 102:3581–6.

174. Lin Z, Li W-H (2012) Evolution of 5′ Untranslated Region Length and Gene Expression Reprogramming in Yeasts. *Mol Biol Evol* 29(1):81–89.

175. Zenklusen D, Stutz F (2001) Nuclear export of mRNA. *FEBS Lett* 498(2):150–156.

176. Vinciguerra P, Stutz F (2004) mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol* 16(3):285–292.

177. Hector RE, et al. (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J* 21(7):1800–1810.

178. Giering JC, Grimm D, Storm TA, Kay MA (2008) Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic. *Mol Ther J Am Soc Gene Ther* 16(9):1630–1636.

179. Xia H, Mao Q, Paulson HL, Davidson BL (2002) siRNA-mediated gene silencing in vitro and in vivo. *Nat Biotechnol* 20(10):1006–1010.

180. Nott A, Meislin SH, Moore MJ (2003) A quantitative analysis of intron effects on mammalian gene expression. *RNA* 9(5):607–617.

181. Zhang B, et al. (2014) A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Mol Cell Biol* 34(12):2318–2329.

182. Thompson JD, et al. (1995) Improved accumulation and activity of ribozymes expressed from a tRNA-based RNA polymerase III promoter. *Nucleic Acids Res* 23(12):2259–2268.

183. Grieger JC, Samulski RJ (2005) Packaging Capacity of Adeno-Associated Virus Serotypes: Impact of Larger Genomes on Infectivity and Postentry Steps. *J Virol* 79(15):9933–9944.

184. Gilbert LA, et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154(2):442–451.

185. Bell ML, Earl JB, Britt SG (2007) Two types of Drosophila R7 photoreceptor cells are arranged randomly: a model for stochastic cell-fate determination. *J Comp Neurol* 502:75–85.

186. Mombaerts P (2004) Odorant receptor gene choice in olfactory sensory neurons: the one receptor–one neuron hypothesis revisited. *Curr Opin Neurobiol* 14(1):31–36.

187. Wernet MF, et al. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* 440(7081):174–180.

188. Jeker LT, Bluestone JA (2013) microRNA regulation of T-cell differentiation and function. *Immunol Rev* 253(1):65–81.

189. Wu J, Tzanakakis ES (2013) Deconstructing stem cell population heterogeneity: single-cell analysis and modeling approaches. *Biotechnol Adv* 31(7):1047–1062.

190. Belete MK, Balázsi G (2015) Optimality and adaptation of phenotypically switching cells in fluctuating environments. *Phys Rev E Stat Nonlin Soft Matter Phys* 92(6):062716.

191. Wu J, Tzanakakis ES (2012) Contribution of Stochastic Partitioning at Human Embryonic Stem Cell Division to NANOG Heterogeneity. *PLoS ONE* 7(11). doi:10.1371/journal.pone.0050715.

192. Suslov ON, Kukekov VG, Ignatova TN, Steindler DA (2002) Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres. *Proc Natl Acad Sci* 99(22):14506–14511.

193. Tang L, et al. (2014) Spontaneous modulation of a dynamic balance between bacterial genomic stability and mutability: roles and molecular mechanisms of the genetic switch. *Sci China Life Sci* 57(3):275–279.

194. Pujadas E, Feinberg AP (2012) Regulated Noise in the Epigenetic Landscape of Development and Disease. *Cell* 148(6):1123–1131.

195. Fialkow PJ, Faguet GB, Jacobson RJ, Vaidya K, Murphy S (1981) Evidence that essential thrombocythemia is a clonal disorder with origin in a multipotent stem cell. *Blood* 58(5):916–919.

196. Hartmann K, Bruns SB, Henz BM (2001) Mastocytosis: Review of Clinical and Experimental Aspects. *J Investig Dermatol Symp Proc* 6(2):143–147.

197. Takahashi K, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–72.

198. Takahashi K, Yamanaka S (2016) A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol* 17(3):183–193.

199. Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C (2006) Transcriptional Dynamics of the Embryonic Stem Cell Switch. *PLoS Comput Biol* 2(9). doi:10.1371/journal.pcbi.0020123.

200. Bu P, et al. (2013) A microRNA miR-34a-regulated bimodal switch targets notch in colon cancer stem cells. *Cell Stem Cell* 12(5):602–615.

201. Oyarzún DA, Chaves M (2015) Design of a bistable switch to control cellular uptake. *J R Soc Interface* 12(113). doi:10.1098/rsif.2015.0618.

202. Ausländer S, Fussenegger M (2013) From gene switches to mammalian designer cells: present and future prospects. *Trends Biotechnol* 31(3):155–168.

203. Kramer BP, Fussenegger M (2005) Hysteresis in a synthetic mammalian gene network. *Proc Natl Acad Sci U S A* 102(27):9517–9522.

204. Litcofsky KD, Afeyan RB, Krom RJ, Khalil AS, Collins JJ (2012) Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat Methods* 9(11):1077–1080.

205. Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. *Science* 317:526–9.

206. Koirala S, et al. (2014) A nutrient-tunable bistable switch controls motility in Salmonella enterica serovar Typhimurium. *mBio* 5(5):e01611-01614.

207. Dressaire C, Moreira RN, Barahona S, Alves de Matos AP, Arraiano CM (2015) BolA is a transcriptional switch that turns off motility and turns on biofilm development. *mBio* 6(1):e02352-02314.

208. Wong Ng J, Chatenay D, Robert J, Poirier MG (2010) Plasmid copy number noise in monoclonal populations of bacteria. *Phys Rev E* 81(1). doi:10.1103/PhysRevE.81.011909.

209. Fu W, et al. (2012) A multiply redundant genetic switch “locks in” the transcriptional signature of regulatory T cells. *Nat Immunol* 13(10):972–980.

210. Cheng Y-Y, Hirning AJ, Josić K, Bennett MR (2017) The Timing of Transcriptional Regulation in Synthetic Gene Circuits. *ACS Synth Biol* 6(11):1996–2002.

211. AkhavanAghdam Z, Sinha J, Tabbaa OP, Hao N (2016) Dynamic control of gene regulatory logic by seemingly redundant transcription factors. *eLife* 5. doi:10.7554/eLife.18458.

212. Zhang L, et al. (2012) Noise drives sharpening of gene expression boundaries in the zebrafish hindbrain. *Mol Syst Biol* 8(1). doi:10.1038/msb.2012.45.

213. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.

214. Blake WJ, et al. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* 24:853–65.

215. Kim KH, Choi K, Bartley B, Sauro HM (2015) Controlling E. coli Gene Expression Noise. *IEEE Trans Biomed Circuits Syst* 9(4):497–504.

216. Wright DA, et al. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* 1:1637–52.

217. Hockemeyer D, et al. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29(8):731–734.

218. Wang L, Gaigalas AK (2011) Development of Multicolor Flow Cytometry Calibration Standards: Assignment of Equivalent Reference Fluorophores (ERF) Unit. *J Res Natl Inst Stand Technol* 116(3):671–683.

219. Wang L, Gaigalas AK, Yan M (2011) Quantitative fluorescence measurements with multicolor flow cytometry. *Methods Mol Biol Clifton NJ* 699:53–65.

APPENDIX A

AUTHORIZATION TO USE COAUTHORED WORK

I, David J Menn, verify that all co-authors – Ri-Qi Su, Samira Kiani, Xiao-Jun Tian, and Xiao Wang – have granted permission to include our joint work as part of this dissertation.

BIOGRAPHICAL SKETCH

David J Menn was born in Appleton, Wisconsin on May 16, 1984. He received his elementary education at Appleton Christian School. His secondary education was spread across Appleton Christian, Oshkosh Christian, and Fox Valley Lutheran School, from which he graduated in 2002. He attended Calvin College in Grand Rapids, Michigan, from which he graduated in 2006 with a Bachelor of Arts degree, with a major in Psychology and a minor in Philosophy. He worked several years before moving across the country to Arizona to gain a post-graduate education from Arizona State University. In 2012, he began his Doctorate degree in Biomedical Engineering under the mentorship of Dr. Xiao Wang. Most importantly, during his tenure there, he met and married his wife, Hao Hu.