Elucidating the Connection Between Cell Population Heterogeneity and Genetic Regulatory Architecture in Specific Artificial Networks

by

Stephanie Portle

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Kyriacos Zygourakis, A. J. Hartsook Professor, Chair
Chemical and Biomolecular Engineering
Bioengineering

Ka-Yiu San, E. D. Butcher Professor, Director
Bioengineering
Chemical and Biomolecular Engineering

George N. Bennett, E. Dell Butcher Professor
Biochemistry and Cell Biology

HOUSTON, TEXAS

AUGUST 2008
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI

UMI Microform 3362386
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346
Elucidating the Connection Between Cell Population Heterogeneity and Genetic Regulatory Architecture in Specific Artificial Networks

by

Stephanie Portle

Understanding the expression patterns of simple, synthetic gene regulatory networks will not only shed light into the complexity of naturally occurring networks, but it will also provide a platform for expression control that can be valuable in biotechnological applications. The expression of regulatory networks is influenced by the fact that the intracellular environment varies among the cells of a population. In turn, this variability is tightly related to the architecture of such networks. The relationship between the architecture of synthetic regulatory networks and cell population heterogeneity was studied using two model regulatory networks: a gene-switching system and an oscillatory system. A green fluorescent protein (GFP) served as the reporter for both systems, which were expressed from plasmids in the Gram-negative bacterium Escherichia coli. Inducer concentrations were varied in shake flask cultures, and GFP distributions were monitored over time with flow cytometry.

In studying the effect of GFP half-life on the gene-switching network behavior, we observed how it influences the view of the network behavior: using a lower half-life GFP reduced the inducer concentration range at which we could distinguish between network states due to lower GFP expression, but its use also showed better evidence of the fast-switching transient behavior predicted by the network architecture through wider separation of states. The oscillatory network was shown to exhibit three steady states, bi-
threshold behavior, and multiplicity, contrary to behavior predicted by an existing model. We experimentally discovered four significant nonspecific interactions between promoters and repressors within the network that, through modeling, can be shown to qualitatively create the behavior experimentally observed.

Beyond the understanding of network behavior gained through the combination of average and population-level data, the distributions demonstrated a connection between the network architecture and heterogeneity. We found heterogeneity expanded at intermediate inducer levels in both networks, when the distribution was bimodal (gene-switching network) or individual cells were displaying oscillatory behavior (oscillatory network). Both the oscillatory behavior and bimodal distributions are a result of the network architectures. We had the ability to restrict heterogeneity with multiple inducers in the oscillatory network. However, there were observable limits in doing so.
ACKNOWLEDGMENTS

I would first like to thank my advisor, Dr. Ka-Yiu San, for all the support he has shown me, particularly through all of the changes and challenges that occurred over the past year. That means as much as all of the invaluable advice and direction that helped motivate my work. I learned from him the importance of a positive attitude, to rejoice in successes and not get frustrated. I owe thanks to my other advisor Dr. Nikos Mantzaris, for also providing direction, for helping me question and make sense of my research, and for helping me to develop better presentation and technical writing skills. Thanks are also due to Dr. Bennett and Dr. Zygourakis of my thesis committee, who have pretty much acted as extra advisors. I have greatly benefited from all of the conversations and questions and suggestions that they have given to me over the past five years.

I am grateful to my family (Mom, Dad, Sean, extended family) and friends for their support. Without it, I may not have made the decision to go to graduate school in the first place. Moreover, they were always interested in how I was doing personally as well as how my work was going. I cannot neglect to mention the friends I made at Rice, particularly my roommates and those from my entering class. Their warmth and humor made my time at Rice much better than it otherwise would have been. I’d also like to thank those in my research groups for their feedback, support, and camaraderie. Lastly, I’d like to thank Blair. I never expected to meet the guy I would marry at Rice, and I still can hardly believe that I did. His molecular biology know-how benefited my work, but more importantly, his faith in me made it easier for me to have faith in myself, and his humor and love (and that of his son) drew me from a work-centered life to one that is more balanced.
### TABLE OF CONTENTS

1 INTRODUCTION .................................................................................................................. 1

1.1 OVERVIEW .................................................................................................................... 1

1.2 CELL POPULATION HETEROGENEITY ......................................................................... 2

1.2.1 Early Experimental Observations and Definition ......................................................... 2

1.2.2 Sources of Heterogeneity .......................................................................................... 3

1.3 GENETIC REGULATORY NETWORKS .......................................................................... 6

1.4 CONNECTION OF HETEROGENEITY WITH GENETIC REGULATORY NETWORKS .... 10

1.5 SYNTHETIC NETWORKS .............................................................................................. 12

1.5.1 Importance of Synthetic Networks ............................................................................ 12

1.5.2 Examples of Artificial Genetic Regulatory Networks .................................................. 13

1.6 OBTAINING CELL POPULATION INFORMATION ......................................................... 15

1.6.1 Flow Cytometry ....................................................................................................... 15

1.6.2 Green Fluorescent Protein ......................................................................................... 18

1.7 BALANCED GROWTH .................................................................................................... 19

1.8 ESCHERICHIA COLI AS A MODEL ORGANISM .............................................................. 20

1.9 THESIS ORGANIZATION ............................................................................................ 21

2 MATERIALS AND METHODS ......................................................................................... 22

2.1 OVERVIEW .................................................................................................................. 22

2.2 STRAINS AND PLASMIDS ......................................................................................... 22

2.3 PROCEDURES RELATED TO GENETIC MANIPULATIONS ....................................... 23

2.3.1 Polymerase Chain Reaction (PCR) .......................................................................... 23

2.3.2 DNA Digestion ......................................................................................................... 24

2.3.3 Agarose Gel Electrophoresis .................................................................................... 24

2.3.4 Gel Extraction ......................................................................................................... 25

2.3.5 Ligation .................................................................................................................... 25

2.3.6 Preparation of Competent Cells, Transformation Procedure .................................. 25
LIST OF FIGURES

FIGURE 1.1. SIMPLIFIED MOLECULAR CARTOON OF THE lac OPERON ................................................................. 8

FIGURE 1.2. THE THREE SYSTEMS OF FLOW CYTOMETRY ......................................................................................... 17

FIGURE 3.1. CARTOON OF THE GENETIC TOGGLE NETWORK IN PTAK117 ......................................................... 32

FIGURE 3.2: OPTICAL DENSITY (OD_{600}) AS A FUNCTION OF TIME FOR [IPTG] = 30 \mu M ............................ 36

FIGURE 3.3: DEFINING THE REFERENCE STATE ...................................................................................................... 37

FIGURE 3.4: AVERAGE GREEN FLUORESCENCE AT THE REFERENCE STATE AS A FUNCTION OF INDUCER
CONCENTRATION .................................................................................................................................................. 39

FIGURE 3.5: NUMBER DENSITY FUNCTIONS AT THE REFERENCE STATE FOR DIFFERENT INDUCER
CONCENTRATIONS ............................................................................................................................................. 41

FIGURE 3.6. TIME EVOLUTION OF THE NUMBER DENSITY FUNCTION AT [IPTG] = 40 \mu M .......................... 43

FIGURE 3.7. TIME EVOLUTION OF THE NUMBER DENSITY FUNCTION AT [IPTG] = 60 \mu M .......................... 43

FIGURE 4.1. THE OSCILLATORY NETWORK. A) THE NETWORK ARCHITECTURE, B) RELEVANT PLASMID
MAPS ............................................................................................................................................................... 49

FIGURE 4.2. AVERAGE FLUORESCENCE OVER TIME AT 150 nM aTc .............................................................. 55

FIGURE 4.3. AVERAGE FLUORESCENCE (OPEN MARKERS) AND AVERAGE FORWARD SCATTERING (CLOSED
MARKERS) OVER TIME AT 150 nM aTc ........................................................................................................... 57

FIGURE 4.4. AVERAGE GREEN FLUORESCENCE AS A FUNCTION OF aTc CONCENTRATION AT REFERENCE
STATE .............................................................................................................................................................. 58

FIGURE 4.5. TYPICAL NUMBER DENSITY FUNCTIONS AT THE REFERENCE STATE FOR LOW, MIDDLE, AND HIGH
STATES, INCLUDING DATA AT VARIOUS aTc CONCENTRATIONS ..................................................................... 59

FIGURE 4.6. AVERAGE GREEN FLUORESCENCE AS A FUNCTION OF TIME AT VARIOUS aTc CONCENTRATIONS.
........................................................................................................................................................................ 61

FIGURE 4.7. THE REGULATORY REPRESSOR-PROMOTER INTERACTIONS THAT DEFINE THE STRUCTURE OF THE
ARTIFICIAL REPRESSILATOR .................................................................................................................................. 67

FIGURE 4.8. STEADY-STATE DEPENDENCE OF THE \lambda REPRESSOR EXPRESSION LEVEL ON [aTc] PREDICTED BY
THE MODIFIED ELOWITZ AND LEIBLER MODEL .............................................................................................. 68
FIGURE 4.9. STEADY-STATE DEPENDENCE OF THE λ REPRESSOR EXPRESSION LEVEL ON [ATC] PREDICTED BY
THE MODEL INCORPORATING NONSPECIFIC PROMOTER-REPRESSOR INTERACTIONS........................................68

FIGURE 4.10. COMPARATIVE NUMBER DENSITY FUNCTIONS FOR THE FOUR CONFIRMED NONSPECIFIC
INTERACTIONS..................................................................................................................................................71

FIGURE 4.11. CHANGE IN FLUORESCENCE PROPERTIES WITH IPTG..............................................................76

FIGURE 4.12. NUMBER DENSITY FUNCTIONS AT THE REFERENCE STATE FOR MIXED INDUCER EXPERIMENTS
WHEN INDUCERS ARE ADDED SIMULTANEOUSLY .......................................................................................76

FIGURE 4.13. AVERAGE GREEN FLUORESCENCE AS A FUNCTION OF TIME FOR DIFFERENT IPTG
CONCENTRATIONS WHEN INDUCERS ARE ADDED SIMULTANEOUSLY ......................................................78

FIGURE 4.14. FLUORESCENCE PROPERTIES FOR CULTURES AT 1000 nM ATC THAT START INITIALLY WITH NO
IPTG.................................................................................................................................................................78

FIGURE A.1. PLASMID MAP OF PZS1-ITLRL.....................................................................................................89

FIGURE A.2. PLASMID MAP OF PZE21-GFPAAV..............................................................................................90

FIGURE A.3. PLASMID MAP OF PSCP102..........................................................................................................91

FIGURE A.4. PLASMID MAP OF PSCP141..........................................................................................................94

FIGURE A.5. PLASMID MAP OF PSCP151..........................................................................................................95

FIGURE A.6. PLASMID MAP OF PSCP142..........................................................................................................96

FIGURE A.7. PLASMID MAP OF PSCP152..........................................................................................................97

FIGURE A.8. PLASMID MAP OF PSCP113..........................................................................................................98

FIGURE A.9. PLASMID MAP OF PSCP123............................................................................................................99

FIGURE A.10. PLASMID MAP OF PSCP114.........................................................................................................100

FIGURE A.11. PLASMID MAP OF PSCP124.........................................................................................................101

FIGURE A.12. PLASMID MAP OF PSCP146.........................................................................................................102

FIGURE A.13. PLASMID MAP OF PSCP156.........................................................................................................103

FIGURE A.14. PLASMID MAP OF PSCP110........................................................................................................104

FIGURE B.1. AVERAGE FLUORESCENCE AND AVERAGE FORWARD SCATTER VALUES OVER TIME.............107

FIGURE C.1. SCHEMATIC OF CHEMOSTAT....................................................................................................110

FIGURE C.2. PH AND DISSOLVED OXYGEN READINGS OVER TIME..............................................................112
FIGURE C.3. AVERAGE FLUORESCENCE AND LIGHT SCATTERING VALUES OVER TIME.............................113

FIGURE C.4. GREEN FLUORESCENCE DISTRIBUTIONS AT STEADY STATE FOR EACH CONDITION......................114
LIST OF TABLES

**TABLE 2.1.** STRAINS, PLASMIDS, AND IDENTIFYING INFORMATION ........................................................................... 23

**TABLE 4.1.** LIST OF PLASMIDS USED IN HYPOTHESIS EXPERIMENTS .............................................................. 70

**TABLE 4.2.** RESULTS SUMMARY OF THE INTERACTION HYPOTHESIS EXPERIMENTS ............................................ 74

**TABLE A.1.** LIST OF PLASMIDS CONTAINING GENE-SWITCHING NETWORK PLASMIDS AND RESPECTIVE HOSTS. ........................................................................................................................................... 88

**TABLE A.2.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PZSI- LTLrL ........................................................ 89

**TABLE A.3.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PZE21-GFPAAV ................................................ 90

**TABLE A.4.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP102 ............................................................ 91

**TABLE A.5.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP141 ............................................................ 94

**TABLE A.6.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP151 ............................................................ 95

**TABLE A.7.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP142 ............................................................ 96

**TABLE A.8.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP152 ............................................................ 97

**TABLE A.9.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP113 ............................................................. 98

**TABLE A.10.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP123 ............................................................ 99

**TABLE A.11.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP114 ............................................................ 100

**TABLE A.12.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP124 ............................................................ 101

**TABLE A.13.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP146 ............................................................ 102

**TABLE A.14.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP156 ............................................................ 103

**TABLE A.15.** LOCATION OF REGULATORY ELEMENTS ON PLASMID PSCP110 ............................................................ 104

**TABLE D.1.** NOMINAL MODEL PARAMETERS FOR THE MODIFIED REPRESSILATOR DYNAMICS ......................... 118
Chapter 1

Introduction

1.1 Overview

Cell populations are heterogeneous systems. There is a distribution of phenotypes within a population of isogenic cells; therefore, not all cells behave like the average cell, as is often assumed. Cells obtain their phenotypes through gene expression. Gene expression is regulated on a variety of levels, including transcriptional, translational, and metabolic [1]. In bacteria, a majority of genes are thought to be primarily regulated at the transcriptional level [2, 3] by a variety of genetic regulatory networks. Components of genetic regulatory networks can be combined to create artificial networks, which are able not only to express foreign genes in an organism, but to exert control over their expression. The connection between genetic regulatory architecture and heterogeneity of a cell population needs to be understood in order to have as much control over expression as possible on the population level, not just on the single-cell level. Using *E. coli* as a model organism, we sought to understand this connection using artificial genetic networks with simple, well-defined architectures as model networks. Fluorescent protein expression, under the control of the networks, were assessed on a population level using flow cytometry. With this knowledge, we explored the ability to control network behavior and the distribution of phenotypes in those populations.
1.2 Cell Population Heterogeneity

1.2.1 Early Experimental Observations and Definition

Heterogeneity is present in every collection of cells with the same genotype, from bacterial cultures to the human body. Cell differentiation is one example of heterogeneity, such as in a developing embryo or in bacteria such as *Myxococcus xanthus*, which form structures called fruiting bodies. Even when there is no differentiation, cells still display differences in phenotype. These differences have been well-documented in literature since the 1940's. In the late 1930’s and 1940’s, Delbrück and colleagues performed a series of experiments to determine the burst size distribution of virus-infected bacteria by measuring the amount of viruses released from lysed cells [4]. Powell observed the amount of variation in the generation times of different strains of bacteria grown in different media in the 1950’s, incorporating theory into his analysis [5]. Around the same time, Novick and Weiner found that bacterial populations that were suboptimally-induced to produce the enzyme β-galactosidase could be divided into subpopulations that either did or did not produce the enzyme [6]. This work was further developed by Maloney and Rotman in the early 1970’s, characterizing this induction phenomenon more thoroughly [7]. A similar on or off behavior was observed by Chung and Stephanopoulos in spore formation of *Bacillus subtilis* [8]. In 1976, Spudich and Koshland published work on the variation in chemotactic states of bacteria in the presence or absence of a chemical gradient [9]. These examples do not constitute a full list, but they mark some major findings with regard to heterogeneity in cell populations.

All of these examples highlight the fact that all cells do not behave identically. Heterogeneity is defined as differences in phenotype that arise between isogenic
cells. The phenotype is determined by intracellular content of proteins and metabolites. It is influenced by all things happening in and on the cell: the type and number of genes being expressed at a given point in time, along with the intracellular reactions (metabolism) and regulatory interactions taking place inside and on the surface of a cell. Instead of all cells of a population expressing the same phenotype, there is a distribution of phenotypes within a population. In studying heterogeneity, we stop looking at the average properties of cell cultures, but instead look at the individuality of the cells within the population to find the distribution of phenotypes.

1.2.2 Sources of Heterogeneity

There are several sources of heterogeneity within populations of isogenic cells. One source of heterogeneity is the environment (or microenvironment) surrounding the cells. Since cells interact with and react to their surroundings, differences in temperature, pH, or extracellular concentrations of nutrients, toxins, quorum factors (a measure of cell density), signals from neighboring cells, or other chemicals can cause differences in the relevant behaviors related to those factors. The cells will “feel” different extracellular conditions and act appropriately, according to what concentration of molecules are already in the cell. This will lead to differences in intracellular state.

Even when cells are cultured in media with (reasonable) spatial uniformity, however, heterogeneity is still observed [4-7, 9]. Therefore, heterogeneity cannot originate only from differences in the environment. Another source of heterogeneity is the fact that cells have different intracellular content. This can arise from the operation of the cell cycle. In the cell cycle, a cell grows, duplicates its DNA, and at some point it divides into two daughter cells. During division, the mother cell partitions its content
between the two daughter cells. While chromosomal partitioning is symmetric in most organisms, most of the cellular material is partitioned unequally [10-14]. This unequal partitioning of molecules from mother cell to the two daughter cells repeats itself with every generation, further expanding the range of phenotypes in a population and creating a distribution. Phenotypic differences between daughter cells caused by unequal partitioning might be amplified when the concentration of a given molecule is small (some regulatory molecules number less than 100 per cell). The mechanism of unequal partitioning was suggested by Novick and Weiner to explain how induced cells can become uninduced over time in a non-inducing medium [6]. This type of heterogeneity is termed extrinsic.

The last source of heterogeneity is stochastic fluctuations in reaction rates between cells of (theoretically) identical content. These reactions may involve regulatory molecules interacting with DNA, or they can be general metabolic reactions. Stochastic fluctuations are more prevalent when the concentration of a molecule is very low in a cell. Since the concentration of regulatory molecules is usually low, they contribute greatly to this source of heterogeneity because their interactions with DNA affect gene expression, which helps to determine the phenotype. This kind of heterogeneity is termed intrinsic.

Furthermore, due to differences in reaction rates that occur within cells, the cells become even more unlike each other in content and this in turn becomes a source of extrinsic heterogeneity. Several studies suggest that these stochastic fluctuations are more due to translation than transcription. McAdams and Arkin concluded this using stochastic models, which predicted that proteins are produced in short bursts from one
mRNA molecule at random time intervals [15]. Other modeling and experimental studies came to the same conclusions [16, 17].

Heterogeneity has been studied both theoretically (through modeling) and experimentally to explore the different sources of heterogeneity. Many of these studies were made in the context of a specific genetic regulatory network and will be discussed later in this chapter. Other studies of heterogeneity have also been made in which observations of one type of heterogeneity or another were observed and quantified. Unequal partitioning of chromosomal content has been studied using flow cytometry [13, 18]. Fluorescence microscopy can also be used to study unequal partitioning. The two intracellular sources of heterogeneity, extrinsic and intrinsic heterogeneity, were studied in vivo by Elowitz et al. [19]. In their experiments, they placed the expression of two fluorescent protein reporters under the control of identical promoter systems (controlled by an inducer), each equidistant and on opposite sides of the origin of replication on an *E. coli* chromosome. In changing the amount of transcription by manipulating the concentration of inducer, they observed trends in the role of each source of heterogeneity: at low levels of expression both sources are present, at intermediate levels there is a maximum of extrinsic heterogeneity, which dominates over intrinsic heterogeneity, and at high expression levels there is a decrease in both types of heterogeneity. This shows a sensitivity in the level of heterogeneity to the inducer concentration (transcription rate). Quantitatively, they demonstrated that extrinsic heterogeneity plays a greater role than intrinsic heterogeneity not just in the experiment described, but with other genetic networks they tested as well.
1.3 Genetic Regulatory Networks

The expression of genes is regulated by a variety of mechanisms in different organisms, since different organisms have different tools to carry out this control. This regulation can happen at various stages, before transcription of mRNA to after translation of the full protein [1]. In prokaryotes, control is believed to be mainly at the transcriptional level [2, 3], in which regulatory elements influence at what frequency the enzyme RNA polymerase can bind to DNA and create mRNA. This in turn influences the amount of protein synthesized.

These genetic regulatory networks are an essential part of life. In controlling expression pre-transcription, cells do not waste resources on making "superfluous intermediates" [1]. In bacteria, particularly, these networks allow intracellular concentrations to be controlled in the cell (presence of an excess amount of protein can inhibit its own synthesis), direct metabolism of the cell depending on its needs (such as in the ability to switch between aerobic and anaerobic pathways), and allow the cell to react to its environment (as in chemotaxis or phototaxis), among other things.

Diauxic growth is one well-studied example of how regulatory networks allow cells to efficiently react to their environment: when bacteria are fed glucose and lactose, bacteria preferentially use first the glucose and then the lactose. The lac operon, encoding proteins that are required to uptake and process lactose, can only be turned on if both glucose is absent and lactose is present. The cAMP-CAP complex promotes transcription of the lac operon while molecules encoded by the lacI gene (lac repressors) repress it. When glucose is present, cAMP cannot bind to CAP and the complex will not promote synthesis due to binding by the lac repressor; if lactose is present, after it is
cleaved it will bind to the lac repressor and prevent it from repressing transcription [1, 20].

Genetic regulatory networks have been studied for over 50 years. A milestone came at the 1961 Cold Spring Harbor conference on cellular regulatory mechanisms, where Monod and coworkers presented an operon model that described the transcriptional control mechanisms related to the lac operon and diauxic growth [21]. They identified basic regulatory elements in bacteria that function together to achieve this control. The idea that genes could selectively be turned on and off was a new concept at the time [1]. Since then, the specifics of the transcriptional regulatory interactions and the molecular species involved have been heavily studied, particularly with the lac operon and bacteriophage lambda lysogenic and lytic cycles.

Since genetic regulatory networks are the subject of the proposed work, it is appropriate to explain the basics of how they work. Regulatory networks have four types of components: promoters, operators, regulatory proteins, and effectors. The promoters and operators are specific sites in the DNA sequence, while the regulatory proteins and effectors are free molecules that interact with DNA and each other.

A promoter is a site or set of sequences in DNA where the RNA polymerase binds to start transcription. A promoter can be constitutive, which means it is not regulated and the RNA polymerase is uninhibited in its transcription (it is always “on”), or it can be inducible, which means that transcription is dependent on the concentration of the regulatory proteins and effectors. In the lac operon, illustrated in figure 1.1, there is both a constitutive and an inducible promoter. The constitutive promoter is upstream of the inducible promoter and is responsible for expression of lacI. The inducible
promoter drives the expression of the genes lacY, lacZ, and lacA.

The operator is also a site on the DNA, found just downstream of the promoter region (can be overlapping with it), where the regulatory molecule or molecules bind to affect transcription. In the lac operon, an operator site exists just downstream of the inducible promoter, but there is no operator near the constitutive promoter.

The molecules that regulate transcription by binding to the operator site are called **regulatory proteins**, and their action can be positive or negative, enhancing the RNA polymerase action or acting as a physical obstruction to it (proteins with this negative behavior are called *repressors*). The molecules recognize specific sequences in the DNA major groove with high affinity and bind to them. The interactions are a combination of hydrogen bonds, ionic bonds, and hydrophobic interactions all along the protein-DNA surface [1]. Many of the proteins form homodimers or heterodimers (or higher multimers), which can allow them to assume a particular conformation (a Y shape is very typical) and very specifically bind the DNA sequences. In the lac system, the repressor is expressed from the gene *lacI*, and it is constitutively expressed.

The **effectors** are molecules that bind to the regulatory proteins. In binding, they cause an allosteric change in conformation of the molecule or dimer at the DNA-binding site, which changes the protein’s ability to bind at the operator site. As with regulatory
proteins, the effect can be positive or negative with regard to the gene of interest. As an *inducer*, an effector can bind to a repressor and free up the operator site, increasing expression of the gene of interest. Or an effector can be a co-repressor, allowing the regulatory protein to become active and bind DNA. Allolactose or a lactose analog like isopropyl-β-D-thiogalactopyranoside (IPTG) can act as inducers of the *lacZ*, *lacY*, and *lacA* genes by binding to the *lac* repressor and liberating the *lacO* operator site.

Once expressed, the regulated gene can exert feedback on its expression. The resulting protein can inhibit or promote its further expression, which is called negative feedback or positive feedback respectively. In the case of the *lac* operon, expression of the regulated genes provides positive feedback by producing a permease that allows more lactose or IPTG into the cell, and by producing an enzyme which cleaves lactose into two molecules, one of which is allolactose, an inducer. This positive feedback behavior is also termed autocatalytic.

The interactions of all of these elements of the regulatory network, as established by their placement in the DNA sequence (genetic architecture), determine the network function. For example, if the expressed protein acted as its own repressor, it would exert negative feedback on its own synthesis and keep the concentration of that protein tightly controlled. On the other hand, if the protein under control of an operator acts as an inducer or brings more of an inducer into the cell (as is the case in the *lac* operon), it acts positively to encourage its own synthesis, and its expression will lead to a rise in its concentration in the cell.
1.4 Connection of Heterogeneity with Genetic Regulatory Networks

Cellular content influences the expression of genes, and differences in content can lead to differences in gene expression, particularly since the concentrations of regulatory proteins controlling the expression are low (the quantity of lac repressors is 10-20 per cell [22]), so heterogeneity plays an important role in regulatory network function. Cells with different amounts of regulatory proteins might express different phenotypes. In turn, the regulatory network architecture should also influence the range of heterogeneity, since not all architectures result in the same sort of control over a gene’s expression: the architecture could enhance differences in cellular content such that the cells become more different from each other, or it could restrict the differences and control the extent of heterogeneity. Moreover, specific behaviors that are expected at the single-cell level based on the architecture might not be the same as that seen at the population level due to heterogeneity in expression. Elucidating the connection between heterogeneity and genetic regulatory architecture with regard to specific networks is the main feature of this project. Understanding the connection will have great relevance in biology and in biotechnological and biomedical applications: the connection is important not just for applications that may use the specific networks being studied, but there will be implications for natural networks with similar architectural features, too.

Modeling and experimental studies have demonstrated the connection between regulatory networks and heterogeneity for some regulatory modules. Becskei and Serrano experimentally demonstrated how stability against fluctuations in gene expression is gained by using negative feedback [23]. In contrast, the lac operon exhibits autocatalytic behavior, which led to the “all or none” behavior (two populations of cells)
described by Novick and Weiner and Maloney and Rotman [6, 7]. Through mechanistic modeling, Carrier and Keasling analyzed the autocatalytic behavior further [24]. They found that the emergence of two populations at different expression levels was a result of the autocatalytic network architecture, and is characteristic for any network where the gene encoding the transporter for the inducer is under the control of the inducer. Becskei et al. also experimentally investigated a network with positive feedback and demonstrated the qualitative difference between using autocatalytic or constitutive expression: constitutive expression produced unimodal phenotypic distributions (with graded shifts in expression with increase in inducer concentration) and expression with positive feedback could result in a bimodal distribution much like in the experimental and theoretical studies with the lac operon (with a larger percentage of cells at higher expression as inducer concentration increased) [25]. They also show that inducer concentration can be used to control the proportion of cells in either low or high expression states. A very nice comparison between a stochastic model (simulating intrinsic heterogeneity) and experimental work with another network with autocatalytic behavior was made by Isaacs et al. [26]. DNA-looping was proposed as a way to decrease heterogeneity in the lac operon by reducing transcription fluctuations (using a stochastic model) [27]. A cell population balance model incorporating extrinsic heterogeneity (unequal partitioning) was used by Mantzaris in studying the lac operon; the characteristic bimodal distributions of an autocatalytic network were found [28].

Looking more closely at the source of intrinsic heterogeneity with the help of a stochastic model, McAdams and Arkin incorporated the production of proteins in short bursts at random intervals. They suggested that this causes large differences in events in
regulatory cascades between cells, which could result and can lead to partitioning of cells down different regulatory paths (cause significant heterogeneity like in the "all or none" behavior) [15]. Other studies of intrinsic heterogeneity investigated its connection with robustness of a behavior [29-33].

1.5 Synthetic Networks

1.5.1 Importance of Synthetic Networks

A multitude of genetic regulatory networks exist in nature. Thousands of DNA sequences that serve as binding sites for different regulatory proteins (or sets of them) have been identified [1]. Global regulatory networks can be complicated, made up of a series of signal transduction reactions whose intermediates affect many different operons. One way to understand the genetic regulatory mechanisms is by studying the local regulatory networks with known components as described above. Once the properties of each network type are found, more complicated architectures can be studied. Alternatively, instead of searching for an operon that displays a given behavior in order to carefully study it or utilize it for recombinant gene expression, one can create artificial genetic networks. This gives the engineer more flexibility to define and tune a given behavior.

There is much interest in using regulatory networks to control expression of recombinant proteins. One only needs to put the gene of interest under the control of a regulated promoter system. With the advent of recombinant DNA technology in the 1970's, scientists and engineers are given the opportunity to do just that and more. We are essentially able to cut, paste, and modify genes to control expression with whatever
components we like, since tools are available to sequence and synthesize nucleotides and delete, modify, and insert genes into foreign organisms.

We can pick and choose components of different natural networks to achieve better control or a desired behavior, leading to a potentially unlimited number of platforms by which to control gene expression. One promoter can be switched for another, or an operator-repressor pair can be switched out. An example of this is the commonly-used tac system. Tac combines the trp promoter and lac operator-repressor, allowing induction by IPTG instead of trans-β-indoleacrylic acid (IAA) to decrease inducer cost, and incorporating the increased strength of the trp promoter to increase production. In order to utilize these amazing tools to rigorously design strains for biotechnological and biomedical applications, we first need to quantitatively elucidate the interplay between genetic architecture and cell population heterogeneity.

1.5.2 Examples of Artificial Genetic Regulatory Networks

Taking this idea further, several genes can be controlled by the network components separately, by connecting the components in specific architectures that may involve more than one regulatable promoter system. We would like to point out two particular examples of such artificial genetic networks.

The first network is a cross-regulatory system made up of two promoter-operator-repressor (POR) sets, such that the system can switch from expression of one gene to the other. This design was originally developed mathematically and experimentally by Chen, Kallio, and Bailey [34, 35]. The repressors are mutually inhibiting, such that expression of repressor 1 down-regulates the expression of repressor 2 and in doing so up-regulates its own synthesis by preventing the repressor 2 from inhibiting repressor 1 transcription.
The cross-regulation action controlled expression so well because, during a switch, the system up-regulated the new path and down-regulated the old simultaneously. The purpose in developing this system was to tightly regulate expression of a gene before induction (minimize "leakiness"); leakage is a practical problem in fermentation when the desired product is toxic to the cell [36]) and promote high expression when the gene was induced. Expression level was determined by using a cat reporter gene, which was assessed as an average property of a sample of culture. A more recent design was developed and tested by Gardner et al. [37], with a green fluorescent protein (GFP) as the marker gene (gfpmut3). In contrast to the method of analysis in the work by Chen et al., expression of this new reporter system was assessed on the single-cell level using flow cytometry, not averaged at the population level. An extra dimension of network expression was observed in this way since not only the average expression was determined, but the distribution of fluorescence was found. They observed interesting phenomena like bimodality in green fluorescent protein expression under certain conditions. The bimodality is attributed to a bistable system created by the characteristics and connectivity of the network. Their toggle plasmid was the basis for one major section of this work, which took the expression analysis on the population level a little further.

Another recently-constructed architecture involves an additional POR set. With three PORs, a wide variety of "regulatory motifs" are possible, such as "negative and positive feedback loops, oscillators, and toggle switches", as demonstrated by Guet et al. who combined three repressors with five different PO pairs and created a library of plasmids with various behaviors [38]. Of all the possible connectivities, the architecture
we wanted to highlight has potential behavior as an oscillator. The three POR sets are connected such that each repressor provides negative feedback to another. Two steady states of expression are possible with this system, one being stable and the other unstable. The stable steady state behavior is characterized by domination of one repressor; the unsteady steady state behavior will display oscillations as the levels of repressors cyclically rise and fall as domination passes from one repressor to the next. This design was constructed, modeled, and experimentally tested by Elowitz and Liebler [39], with a GFP as a marker for one of the repressors. Oscillations were demonstrated by fluorescence microscopy on individual cells, and it was observed that synchrony in oscillation was not often maintained after cell division, which seems to indicate a strong influence of extrinsic heterogeneity. Characterization on a population level was not done due to this asynchrony in expression.

1.6 Obtaining Cell Population Information

1.6.1 Flow Cytometry

Since we are interested in the distribution of phenotypes of the cell populations, we need to use analytical tools to obtain this information. This requires determination of a large number of individual cell phenotypes in order to obtain a distribution which is representative of the population. Through analysis of population distribution data (instead of average data), it is possible to gain insight on the "mechanistic basis of cellular phenomena" [40]. Microscopy is one way to find population data. Distributions of any number of parameters (behavior, size, or other characteristics) for even a subset population, like dividing or newborn cells, can be determined. However, microscopy is
very time consuming [41]. Instead, one extremely effective and efficient way to find
distributions is by using a method called flow cytometry. Flow cytometry is a versatile
tool used in clinical, research, pharmaceutical, and industrial applications. It can be used
to determine cell viability, determine intracellular and surface concentrations of proteins
or receptors, quantify antibodies on cells, indicate the intracellular pH and calcium levels,
analyze the cell cycle, and sort chromosomes, among other things [40, 42]. In the past
few decades, this valuable technique has been used to study cell culture with
biotechnology applications in mind, often with yeast or bacterial systems. Work has been
done in studying the cell cycle [18, 43, 44], monitoring batch or chemostat processes [45-
47], and studying recombinant protein production [45, 48-50], just to highlight a few
examples.

In this high throughput method, the flow cytometer measures light scattering and
fluorescence properties of individual cells, at a rate up to several thousand cells per
second, if not more. A schematic of different systems involved in flow cytometry can be
found in figure 1.2. First, single cells are suspended in a liquid medium (see 1.2a). Cells
in the suspension fluid are taken up into the machine and, due to the difference in
pressure with the surrounding fluid, the stream is hydrodynamically focused to form a
column of cells. The stream rapidly flows past a light source of fixed wavelength, often a
laser, which intersects the cells. As shown in 1.2b, the light is scattered by the surface
and intracellular content of the cells, and, if fluorescent molecules are present, there will
be fluorescence if the laser is set at a wavelength within the excitation spectra of the
molecules. Detectors relay the light scattering and fluorescence properties to the
electronics portion of the cytometer, which processes the signals, applies predesignated
amplification to convert the light signal to a proportional electronic signal, and records the signals (see 1.2c).

With flow cytometry, we can measure several cellular properties. Forward (low angle) scattering is related to the size of the cell, side scattering (near 90°) is related to the internal complexity, or granularity, and the fluorescence level is related to the concentration of fluorescent molecules in or on the cell. Light scattering is often used to identify the cells from debris that may be detected by the cytometer. Once a threshold is set to gate out debris, the flow cytometer records information from events exceeding the threshold. Fluorescence level and concentration of the fluorescent molecule are correlated in a linear relationship: \( \log(\text{fluorescence level}) \) is proportional to \( \log(\text{fluorophore concentration}) \). Fluorescent proteins produced intracellularly will be used in this project to provide a measure of the network behavior.

When more than one fluorescence detector is being utilized on the flow cytometer (multiple fluorescent molecules are being monitored), compensation might be necessary when fluorescence signal of one color (with an emission maximum lying within the
wavelength limits of one detector) is being read in another detector due to spectral overlap. Without proper compensation, the reported signal can be false (an artifact).

1.6.2 Green Fluorescent Protein

Fluorescent proteins are used as markers in research and medicine for a variety of different applications, such as protein localization when used in fusion tags, FRET (fluorescence resonance energy transfer) studies of protein-protein interactions, FRAP (fluorescence recovery after photobleaching) as a measure of protein diffusion, as calcium, halide, or pH indicators, or as markers for gene expression [52]. They have been successfully implemented in studies over a wide range of organisms, from mammals to yeast to bacteria. Advantages to using these proteins are that they are produced intracellularly (no staining of cells is required), do not require the addition of extra substrates or cofactors to fluoresce, and are extremely stable over time [52, 53]. Green fluorescent protein (GFP) was originally isolated from the jellyfish Aequorea victoria (although it is found in other organisms as well) and reports of its properties began to come out in the 1950’s and 60’s [52, 53]. It is the GFP gene from Aequorea that was first cloned in 1992 and expressed in 1994, and is the basis for many mutants [52].

GFP is composed of 11 beta strands in a barrel formation, enclosing an alpha helix running up its axis. The chromophore is attached to the alpha helix, near the center of the barrel. Once the protein is folded in its correct conformation, an autocatalytic reaction causes cyclization of the chromophore and finally, in the presence of molecular oxygen, the chromophore is oxidized and becomes fluorescent [52]. The slowest kinetics involved with formation of the fluorescent chromophore are during the last step of chromophore oxidation, with an estimated time ($t_{1/2}$) of 76 min [54].
Wild type GFP has been modified by many groups to change specific properties of the protein, which may be desirable for a given application. There is significant variation in oxidation times of mutants from that estimated of the wild type protein, some mutations greatly reducing the time required for the protein becomes fluorescent. The proteins have been modified to change the excitation and emission properties and half-lives for specific applications. For example, some variants have been optimized for flow cytometry analysis, with excitation maxima near 488 nm [55]. A whole variety of GFP mutants have shifted emission spectra, with blue, cyan, yellow, or further red-shifted fluorescence, and this allows more flexibility in choosing which color protein to use or in using more than one fluorescent protein simultaneously. Other variants have been created with reduced half-lives which would be more ideal for *in vivo* studies of transient gene expression [56].

### 1.7 Balanced Growth

In order to make meaningful comparisons between the fluorescence distributions for the various conditions, it is important to determine when the cells are in a similar physiological condition. Although comparisons are often made between cultures at a particular point in time [37, 57], factors such as growth phase, inducer concentration, and cell concentration can affect the physiological characteristics of cultures.

The most preferable state at which to compare the cultures is that of balanced growth, a state reached during exponential phase [58], originally defined as when the average composition of the cells (or relative concentrations of all metabolites and enzymes) does not change over time [59, 60]. More stringently, balanced growth is
reached when the number density functions (a function defined as the ratio of the distribution of a phenotype over the cell density) of every characteristic describing the culture are non-changing over time (see Fredrickson et al. [61] for a thorough discussion). This theoretical state is not possible to confirm experimentally, but the idea of reaching a stable state was mimicked: our requirements for what we call the "reference state" are that the light scattering averages, green fluorescence averages, and green fluorescence distributions are relatively time-invariant (see chapters 3 and 4 for figures depicting how this is determined experimentally). This is in contrast to the assumption often made that cultures are in a reference state after they are grown in exponential growth phase for a certain number of generations with periodic dilutions into fresh media [25, 37]. At balanced growth, normalized phenotypic distributions are independent of both cell concentration and time and can be regarded as characteristic for the specific host, system, and set of conditions [58]. We view the experimentally-determined reference state in a similar way.

The state of balanced growth is different than the classical definition of steady state. Steady state is reached when the average concentrations of all components do not change over time. This is only possible in continuous culture, when a constant cell density is maintained. Balanced growth refers not only to the average values, but also to the normalized distributions, making the definition more stringent. Also, it applies generally to cells in exponential phase, whether in continuous culture or in batch culture.

1.8 *Escherichia coli* as a Model Organism

There are several advantages to working with *E. coli* as our model organism. *E.
E. coli is a gram-negative bacterium [62] and a unicellular organism, like all bacteria. It is rod-shaped, mobile, fast-growing, and has a circular chromosome of approximately 4.6 million base pairs. The entire genomes of many E. coli strains are known (this knowledge being a requirement for genetic manipulation of a chromosome or deciding on a proper host strain for a constructed plasmid), and the tools for genetic modification of bacteria are well-developed. E. coli has historically been and is presently the species of choice when studying “microbial physiology, biochemistry, and molecular biology” [62].

1.9 Thesis Organization

The thesis is divided into 5 chapters: Chapter 1 contains relevant background information and important concepts to aid in understanding the thesis work. Chapter 2 is the materials and methods section, including all relevant protocols, strains, plasmids, and chemicals. They are described in a general way. More detailed description of materials or methods used in a particular experiment are included within the chapter explaining the experiment. Chapter 3 includes the work done with the gene-switching (toggle) system, exploring inducer-controlled behavior when using in-vivo reporters of different half-lives, which lead to different views of the population behavior. Taking on a more complicated system, the oscillatory (repressilator) system, Chapter 4 describes our exploration of its range of behavior using two inducers, and resolves our results in light of mathematical modeling, which led us to revise the accepted view of the connectivity of the system and confirm our suspicions through experimental testing. Chapter 5 concludes the thesis with summary, conclusions, and future work.
Chapter 2

Materials and Methods

2.1 Overview

This section contains information about strains, plasmids, media and chemicals, and general protocols for plasmid constructions, experiments, and analysis. More specific information about procedures used for particular experiments will be given in subsequent chapters, since those protocols differ between chapters.

2.2 Strains and Plasmids

In all experiments, it is desirable that no natural regulatory network interfere with the artificial network. Ideally, none of the repressor genes and promoter/operator sites should be duplicated on the host chromosome (unless inserted as part of an integrated artificial network).

Experiments with the gene-switching network (toggle) were performed using E.coli strain JM2.300, which produces a non-functional lac repressor. It was assumed that this non-functional repressor did not interact with the inducer, IPTG, nor with other functional repressors to form a tetramer, nor with the operator sites on the plasmids. There is a lac operator site present on the host chromosome, but any titration effect due to binding with the plasmid-encoded lac molecules was assumed to be negligible since the number of operator sites on the plasmids used was ~30 per cell [63]. One experiment was performed in MC4100, an E.coli strain described in the next paragraph.

In the oscillatory network (repressilator) experiments, the strain of E.coli used
must lack the *lac*, *lambda*, and *tet* operons and the repressor proteins controlling them.

The strain chosen is MC4100. Both stains are listed with genotype in Table 2.1.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Identifying information</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM2.300</td>
<td>CGSC strain 5002</td>
<td><em>lambda</em>-, <em>lacI</em>22 <em>rpsL</em>135 (StrR), <em>thi</em>-1</td>
</tr>
<tr>
<td>MC4100</td>
<td>ATCC® Number: 35695</td>
<td><em>F</em>- <em>araD</em>139 <em>delta</em>(argF-<em>lac</em>)U169 <em>prsL</em>150 <em>relA</em>1 <em>deoC</em>1 <em>rbsR</em> fitH5301 fruA25 <em>lambda</em>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmids¹</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTAK117</td>
<td>original toggle plasmid</td>
<td>[37]</td>
</tr>
<tr>
<td>pGNN10030²</td>
<td>same as pTAK117 with lower stability GFP, gfpaav</td>
<td>[64]</td>
</tr>
<tr>
<td></td>
<td>also known as pGFPaav</td>
<td></td>
</tr>
<tr>
<td>pZSl-ITLrL</td>
<td>main repressor plasmid</td>
<td>[39]</td>
</tr>
<tr>
<td>pZE21-GFPaav</td>
<td>reporter plasmid for repressor</td>
<td>[39]</td>
</tr>
<tr>
<td>pSCP102</td>
<td>combination of main and reporter plasmids</td>
<td></td>
</tr>
<tr>
<td>pSCP141</td>
<td>used to test for <em>tet</em> repressor - <em>λ</em> PR interaction</td>
<td></td>
</tr>
<tr>
<td>pSCP151</td>
<td>same as pSCP141, without promoter for tetR</td>
<td></td>
</tr>
<tr>
<td>pSCP142</td>
<td>used to test for <em>lac</em> repressor - <em>P</em>₁tetO-1 interaction</td>
<td></td>
</tr>
<tr>
<td>pSCP152</td>
<td>same as pSCP142, without promoter for lacI</td>
<td></td>
</tr>
<tr>
<td>pSCP113</td>
<td>used to test for cl repressor - <em>P</em>₁lacO-1 interaction</td>
<td></td>
</tr>
<tr>
<td>pSCP123</td>
<td>same as pSCP113, without promoter for cl</td>
<td></td>
</tr>
<tr>
<td>pSCP114</td>
<td>used to test for cl repressor - <em>P</em>₁tetO-1 interaction</td>
<td></td>
</tr>
<tr>
<td>pSCP124</td>
<td>same as pSCP114, without promoter for cl</td>
<td></td>
</tr>
<tr>
<td>pSCP146</td>
<td>used to test for <em>lac</em> repressor - <em>λ</em> PR interaction</td>
<td></td>
</tr>
<tr>
<td>pSCP156</td>
<td>same as pSCP146, without promoter for lacI</td>
<td></td>
</tr>
</tbody>
</table>

2.3 Procedures Related to Genetic Manipulations

2.3.1 Polymerase Chain Reaction (PCR)

PCR is performed to amplify fragments (or whole plasmids) of DNA. Reaction mixes were prepared as directed by the manufacturer for each enzyme and associated kit (Platinum *Pfx* DNA Polymerase, PfuUltra High-Fidelity Polymerase, Easy-A High-

---

¹ complete description of plasmids used during this work, including plasmid maps, is found in Appendix A
² constructed by labmate T. Bryce Causey
Fidelity PCR Cloning Enzyme, Stratagene) and performed on a Robocycler Gradient 96 (Stratagene). Annealing temperatures were approximated by using the online oligo analyzer application on the Integrated DNA Technologies website (http://scitools.idtdna.com/analyzer/Applications/OligoAnalyzer/), and were usually fine-tuned by setting a range for the annealing temperature and looking for the most concentrated PCR product. To remove the enzyme, PCR products were immediately cleaned (QIAquick PCR Purification Kit, Qiagen) and resuspended in Sigma water (free of nuclease).

2.3.2 DNA Digestion

DNA (plasmids or PCR products) was digested (cut leaving blunt or sticky ends) using restriction enzymes from New England Biolabs or Fisher Scientific, which were used in appropriate buffer at specific temperatures as recommended by the manufacturer for the particular enzyme. Digestions of large amounts of DNA for gel extraction were usually incubated for two hours; digestions to check a small amount of DNA for plasmid size or orientation of a gene were incubated at least 30 minutes.

2.3.3 Agarose Gel Electrophoresis

Gels were composed of 0.8-1% agarose in TAE buffer, with 0.5% ethidium bromide. After gels were placed in a TAE buffer, samples and appropriate standards (100 bp or 1 kb DNA ladders, New England Biolabs) were mixed with gel dye and loaded into wells. A constant voltage of 100 V was applied across the gel for 40-60 minutes to separate linear DNA fragments by size. Ethidium bromide intercalates in the DNA and fluoresces under UV light, illuminating the bands.
2.3.4 Gel Extraction

After electrophoresis, if DNA fragments were needed for subsequent ligation, fragments were visualized with UV light and cut from the gel using a clean razor blade. After measuring the gel slice weight, the procedure of the QIAquick Gel Extraction Kit (Qiagen) was followed to separate and purify the DNA from the gel.

2.3.5 Ligation

Linearized DNA fragments were circularized by using a ligation procedure. Insertion and recircularization ligations were performed as recommended by New England Biolabs, using T4 DNA Ligase with appropriate buffers at 16°C for 0.5-4 hours or using the Quick Ligase Kit at room temperature for 5 min. If fragments did not already have phosphorylated ends from a restriction digest, a kinase reaction was used as recommended by manufacturer before ligation (T4 Polynucleotide Kinase, New England Biolabs). Transformation of circularized DNA was performed within 24 hours (kept on ice or at -20°C until use). TOPO cloning and transformation was performed with the pTrcHis TOPO TA Expression Kit.

2.3.6 Preparation of Competent Cells, Transformation Procedure

The cells were made chemically competent (able to take up plasmids) and transformed with plasmid as described in RbCl Transformation Procedure for Improved Efficiency [65]. Unused competent cells were stored in a -80°C freezer. If a plasmid was difficult to transform, supercompetent cells were used, following their appropriate procedure (One Shot TOP10 Chemically Competent Cells, Invitrogen).

2.3.7 Plasmid Extraction

Isolation of plasmid from cells was performed using miniprep kits according to
the appropriate procedures (QIAprep Miniprep Handbook, Qiagen; QuickClean 5M Miniprep Kit, GenScript). To concentrate the amount of DNA from one miniprep, two cultures were processed separately (preparation and clearing of lysate), but the supernatant lysate from each was poured over the same spin column to adsorb as much DNA as possible on the membrane. Sequencing of plasmids was performed by Lone Star Labs, Inc. (Houston, TX).

2.3.8 Glycerol Stocks

For long-term storage of strains, glycerol stocks were prepared. 0.5 mL of glycerol was added to 0.5 mL of an overnight culture. After mixing, cells were left at room temperature for 0.5-1 hour to acclimate to the new environment. Then they were stored at -80°C.

2.4 Experimental Procedures

2.4.1 Shake Flask

After growing up cells from a glycerol stock in 5 mL rich media (LB for gene-switching system and LSRB for oscillatory system) with the appropriate antibiotic(s) (Ap for gene-switching system and Km for oscillatory system), without inducer(s), for 12 hours, the cells were diluted to a lower OD_{600} and grown into log phase with the appropriate inducer(s) for an extended period of time to reach the reference state. This was accomplished by either growing the cells from a very low OD_{600} in one 2000 mL Erlenmeyer flask until the cultures left log phase (gene-switching system), or by continuously subculturing the cells in 250 mL Erlenmeyer flasks every 3.2 generations within an OD_{600} range of 0.001-0.1 (oscillatory system). The cells were grown
aerobically by filling 1/5\textsuperscript{th} of the flask volume with media and capping the flasks with foam to allow oxygen transfer. Flasks were placed in an orbital shaker (New Brunswick Scientific) at 32°C and 250 rpm. Samples were taken approximately once per generation.

2.4.2 Preparing Samples for Flow Cytometry

Once a sample was taken from the shake flasks, it was kept on ice, covered from light. It was subsequently washed twice in ice cold PBS, and resuspended in a minimum of 0.5 mL PBS at an OD\textsubscript{600} of approximately 0.05 (exact concentration was not important, as it only affects how low it takes for the sample to be read in the cytometer). Samples were kept on ice, covered from light, until they were read by the cytometer.

2.4.3 Propidium Iodide (PI) staining

To test whether the membrane of individual cells was compromised, the cell suspension prepared for flow cytometry in PBS was stained with PI. PI stock solution was diluted to ~10 mg/L in tubes containing the samples; after letting the samples sit for 5 minutes at room temperature, the samples were immediately tested on the flow cytometer.

2.5 Media and Chemicals

\textit{LB} (Luria-Burtani broth): 10 g/L NaCl, 10 g/L tryptone (BD Biosciences), 5 g/L yeast extract (Fisher BioTech), autoclaved

\textit{LSRB} (low salt rich broth): 4 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract, autoclaved

\textit{PBS} (phosphate-buffered saline): 58 mM Na\textsubscript{2}HPO\textsubscript{4}, 17 mM NaH\textsubscript{2}PO\textsubscript{4}, 68 mM NaCl, in Milli-Q H\textsubscript{2}O, pH adjusted between 7.3 and 7.4, filtered with a 0.22 \textmu m filter; stored at room temperature

\textit{IPTG} (isopropyl-beta-D-thiogalactopyranoside): IPTG binds to the \textit{lac} repressor and is
therefore an inducer for anything controlled by a lacO site. Prepared stock solutions at approximately 44 mM, 2 mM, and 500 μM in Milli-Q H2O, filter-sterilized with a 0.22 μm filter; stored at −20°C. (Fisher BioTech)

\( aTc \) (anhydrotetracycline): \( aTc \) binds to the tet repressor from the Tn10 operon and is therefore an inducer for anything controlled by a tetO site. Prepared stock solutions at approximately 125 μM in Milli-Q H2O, filtered with a 0.22 μM filter; stored at 4°C. (Acros Organics)

**TAE Buffer:** 40 mM Tris-acetate, 2 mM EDTA, adjusted to pH 8

**Antibiotics:** Ampicillin (Ap; Fisher BioTech) and kanamycin (Km; Sigma-Aldrich) were prepared in stock solutions at 50 g/L, streptomycin (Str; Sigma-Aldrich) at 10 g/L, all in Milli-Q H2O, and filter-sterilized; stored at −20°C

**Propidium Iodide** (PI): PI is used to identify dead or injured cells (Lewis et al., 2004; Alsharif and Godfrey, 2002), as it can traverse the membrane of cells with compromised membranes and bind to the DNA (fluorescing red). Stock solution at 1 mg/mL in H2O, from Sigma-Aldrich.

### 2.6 Procedures Used in Analysis

#### 2.6.1 Cell Density

Cell density was measured using a spectrophotometer (Bausch & Lomb, Spectronic 1001), which reports an optical density, or an OD value. At 600 nm, an OD\(_{600}\) of 0.1 corresponds to \( \sim 1.0 \times 10^8 \) cells/mL, and the relationship is assumed to be linearly proportional both at lower and higher OD\(_{600}\) values [66]. The samples were diluted with appropriate amount of media or PBS to ensure readings below 0.4, which is the upper limit of the linear range for this spectrophotometer.
2.6.2 Flow Cytometry – set up and operation

A FACScalibur flow cytometer was used (BD Biosciences), with a 15 mW, 488 nm, air-cooled argon-ion laser. All parameters were logarithmically amplified, with settings specified for each experiment (listed in subsequent chapters). A side scatter threshold was applied to gate out much of the noise (at channel 130). When PI was used, compensation of FL1 - 0.9% FL2 was applied to correct for the PI signal that was seen by the FL1 detector. During acquisition, all events were saved until the software counted 5,000-10,000 cells (as determined by a minimum FSC channel value) for each sample. BD Calibrite Beads (BD Biosciences) were used to check that there were no significant changes in the flow cytometer performance over time.

2.6.3 Analysis of Flow Cytometer Files

Gates were applied in the post-processing of the flow cytometer files according to light scattering and red fluorescence (if PI was used). The green fluorescence value was never used in a gate. Since E. coli cells are so small and close to the size of the debris being sensed by the cytometer, light scattering gates were applied to better identify the cells. After observing the change in light scattering values while the culture was in exponential phase, non-changing FSC and SSC gates were applied to exclude as much debris as possible while including all events believed to be cells. If PI staining was used, a FL3 gate (red) was also applied (excluding events above a determined FL3 value that represent dead or injured cells).

The listmode files from the flow cytometer were converted from fcs format to ascii format with MFI software [67], after confirming that FACScalibur data can be successfully converted using MFI. Data were then processed with a FORTRAN code to
apply aforementioned gates, and calculate averages and normalized distributions (excluding events with FL1 (green) values in the first or last channel, which are viewed as not quantitatively accurate [68]).
Chapter 3
Gene-switching network

3.1 Motivation

In order to understand the connection between genetic regulatory architectures and cell population heterogeneity, we must start with simple, synthetic networks. Artificial gene regulatory networks are typically simpler than natural networks and consist of elements with well-defined, pre-determined functions, which do not interfere with the rest of cellular function. Thus, they offer the great opportunity to study in isolation the complex interplay between heterogeneity and specific regulatory architectures.

The first system we studied is a gene-switching network, with two mutually-repressing promoter-repressor sets, which was designed to switch between domination of each repressor. From a biotechnological standpoint, we want to control the system’s behavior, and learn to what extent we can control the behavior. We varied the concentration of the system’s extracellular inducer in order to control the switching between states and to observe the effect on distribution characteristics.

3.2 The Toggle Network

The network, designed and constructed by Gardner and co-workers [37], is composed of two promoter-operator-repressor sets, connected in a mutually repressive way: expression of one repressor inhibits expression of the other repressor (see figure

1 part of this chapter is published in Portle et al., Journal of Biotechnology, 128(2), 362-375 (2007)
This two-gene network has two possible states. In the absence of inducers at low temperature, lac repressors dominate (GFP expression is low). To switch to domination of λ repressors, the extracellular inducer isopropyl-β-D-thiogalactopyranoside (IPTG, a lactose analog) can be added, which will bind to the lac repressors and derepress cl expression. A similar genetic switch had also been constructed earlier by Chen and co-workers [34, 35] using the same repressor proteins but different promoters (λ P_L and tac) fused with different reporter genes (vibrio hemoglobin and chloramphenicol acetyl transferase gene). Chen’s construct was designed with non-native protein production in mind. The purpose of Chen’s design was to minimize leakage by the equivalent P2

3.1). This leads to each repressor competing for dominance in the cell. In the corresponding plasmid called pTAK117 that we used in this work, the two repressors are the lac repressor and the temperature sensitive λ cl repressor from phage lambda.

The lac repressor inhibits the function of thePtrc-2 promoter controlling expression of the λ repressor, which in turn inhibits the function of the P_Ikon promoter controlling expression of the lac repressor. The stable GFP variant GFPmut3 [55] is co-expressed with the λ repressor, thus functioning as a reporter protein for λ repressor expression levels.

Figure 3.1. Cartoon of the genetic toggle network in pTAK117, modified from Judd et al. [69]: P_1 = P_Ikon, R_2 = lac repressor (lacI) (light circles), P_2 = Ptrc-2, R_1 = λ repressor (ciIs) (dark circles), reporter = GFP (gfpmut3) (squares), inducer = IPTG (pentagons). Reproduced from Portle et al. [64].
promoter before expression from it was desired, so that if a toxic protein was being expressed by P2, it would not kill the culture before the culture had an opportunity to grow to a high cell density, at which time the protein production could be switched on.

3.3 Objective

In their original work where the pTAK117 construct was presented, Gardner et al. (2000) mainly focused on illustrating the presence of the two aforementioned states of the genetic toggle network. Using an experimental characterization assay consisting of three successive dilution steps and applying changes in IPTG inducer concentration and temperature at different points in time, they also showed the stability of the induced and uninduced states in the pTAK117 system. Moreover, their flow cytometric measurements of the entire fluorescent distributions at three IPTG concentrations revealed some interesting patterns at the cell population level and clearly illustrated high extents of heterogeneous behavior, including bimodality. However, the half-life of the GFPmut3 reporter protein is known to be more than 24 hrs [56], while the half-life of the $\lambda$ repressor is of the order of 1 hr [70]. Thus, for a more complete understanding of the relationship between the specific genetic toggle architecture and cell population heterogeneity, a reporter protein with a half-life close to that of the $\lambda$ repressor must also be used.

Multiple GFP variants have been employed by several researchers for reporting promoter activity or protein expression levels. Bi et al. [71] studied the dynamics of cell death using two GFP variants with different half-lives. Their results showed that the

\footnote{see Appendix B for confirmation of upper state stability with reduced half-life GFP plasmid}
destabilized variant exhibited better dynamic characteristics. Furthermore, Sternberg et al. [72] compared the use of a stable and unstable GFP variant in studying cell growth in biofilms. They found that the stable GFP was good for estimating growth rates during exponential growth, but for less favorable or stable conditions than a chemostat, the unstable variant had distinct advantages.

Motivated by the above considerations, we first constructed a variant of plasmid pTAK117 containing a reduced half-life GFP gene (gfpaav) in place of gfpmut3. This new gfp is a variant of gfpmut3 itself, mutated by the addition of a degradation tag at the C-terminus [56]. The excitation and emission spectra of these proteins are very similar, assuring us that the only difference between these proteins is in degradation rate. The approximate half-life of gfpaav is 60 minutes. We then performed detailed characterization experiments with both the long and reduced half-life GFP reporter proteins. The distribution measurements were carried out with flow cytometry. The parameters relevant to this study are the two light scattering properties, forward and side scatter, and green and red fluorescence (FSC, SSC, FL1, and FL3, respectively). Green fluorescence represents intracellular GFP content, while red fluorescence represents propidium iodide (PI) content, a measure of viability. Our aim was to elucidate the level of control that IPTG and protein half-life can exert on cell distribution characteristics.

3.4 Materials and Methods

**Strain:** E. coli strain JM2.300 (genotype listed in Chapter 2) was used for all experiments. This strain has a mutant lac repressor that is non-functional. It was transformed with either plasmid pTAK117 or plasmid pGFPaav (see below).
Plasmids: Two plasmids were used in this work. The first one, pTAK117 with the long half-life GFP, was a gift by Professor Collins. Standard methods were used for construction of the plasmid with the reduced half-life (called pGFPaav), performed by co-worker T. Bryce Causey.\(^3\)

Shake flask experiments: Cells were grown to exponential phase in a single shake flask. First, cells were grown overnight in 5 mL of LB medium (10 g/L NaCl, 10 g/L BactoTryptone, 5 g/L yeast extract, 100 mg/L ampicillin) for 12 hours and then subcultured at a low cell density (~1-2000 cells/mL) in 400 mL of prewarmed and aerated LSRB medium with 100 mg/L ampicillin and the appropriate concentration of IPTG in the range of 10 μM-2000 μM at 32°C, shaking at 250 rpm in an orbital shaker, covered from light. The 2 L flasks were capped with foam to allow oxygen transfer. Samples were withdrawn approximately once per generation (40 min) until the cultures started to transition to stationary phase (see figure 3.2 for a representative optical density (OD\(_{600}\)) curve. Samples were not taken until the cultures reached an OD\(_{600}\) of approximately 0.0001 due to the low initial cell concentration. Cells were washed twice in ice cold PBS. The samples were kept on ice and shielded from light before being measured with the flow cytometer. Triplicate experiments were performed for each IPTG concentration.

PI Staining: As written in Chapter 2.

Flow Cytometry: The following settings were used: FSC E01, SSC 381 V, FL1 601 V, FL2 500 V, FL3 575 V. Compensation of FL1 – 0.9% FL2 was applied to remove overlap of PI emission spectra into FL1. 20,000-40,000 events were collected for each

\(^3\) The full description of the construction can be found in Portle et al., *Journal of Biotechnology*, 128(2), 362-375 (2007).
Gating Data: Gates were applied in the post-processing of the flow cytometer files according to red fluorescence (PI) and light scattering. Cells that stained strongly with PI had lower fluorescence than the rest, possibly due to GFP leakage out of the cells [73]. Since these cells did not represent the network dynamics, all cells that stained strongly with PI were excluded. Moreover, since *E. coli* cells are so small and close to the size of the debris being sensed by the cytometer, gates were applied to better identify the cells. After observing the change in light scattering values while the culture was in exponential phase (between 300 and 700 min), identical FSC and SSC gates were applied to include events above a linear value of 30 (channel value of 378). Additionally, an FL3 (red fluorescence) gate was applied to exclude events above a relative red fluorescence level of 40 (channel value of 410), excluding dead or injured cells.

Software: Flow cytometer fcs files were processed as outlined in Chapter 2.

3.5 Results and Discussion

Figures 3.3a and 3.3b show a representative example of the time evolution for the average FSC (a measure of cell size), and average green fluorescence of the entire cell
Figure 3.3: Defining the reference state. All results at [IPTG] = 100μM. Filled circles: pTAK117; open circles pGFPav. a) Average forward scatter (FSC) as a function of time. b) Average green fluorescence as a function of time. c) Number density function for pGFPav at t = 460 min (solid line), t = 500 min (dashed line) and t = 540 min (dotted line). Number density functions practically overlap for three generations, while the average FSC and green fluorescence remain almost the same (holds true regardless of GFP half-life and [IPTG]). Reproduced from Portle et al. [64].
population for both strains and for [IPTG] = 100 μM. Notice that independent of GFP half-life, the average population properties initially increase with time. They subsequently reach a plateau lasting for approximately three generations (460-540 min), while afterwards, the average size and average fluorescence start dropping with time as the cells prepare to enter into the stationary phase. Moreover, during the period of time where the average population properties reach a plateau, the fluorescence number density functions become practically indistinguishable from each other (figure 3.3c). Very similar qualitative behavior was observed for all other IPTG conditions studied.

Thus, based on these observations, green fluorescence distributions were compared at points in time where the average properties reach this characteristic plateau and the number density functions overlap. This quasi steady state of the cell population will henceforth be referred to as the “reference state”.

It was found that for low (below 30 μM) and high (above 50 μM) IPTG concentrations the reference state was reached at either 500 or 540 minutes starting from an initial cell concentration of 2000 cells/ml, well before cells exit logarithmic phase (they start changing to stationary phase in FSC and SSC as stated above just after those values plateau). However, for intermediate IPTG concentrations (30-50 μM) more time was required for the reference state to be reached. It was found that IPTG concentration does not affect the growth rate of the culture. Moreover, no differences were found in the time to reach the reference state between the GFP reporter proteins. Therefore, this qualitative difference for intermediate IPTG concentrations was attributed to the close competition between the λ and lac repressor concentrations, which in turn leads to slower evolution of the genetic toggle dynamics at intermediate [IPTG]. Thus, in order to avoid
reaching the stationary phase before the aforementioned criteria for the reference state were met, at intermediate [IPTG], lower initial cell concentrations were used (1-10 cells/mL).

After establishing a reference state where comparisons are meaningful, we focused on the experimental characterization of the state of the entire cell population. Figure 3.4 shows the average relative green fluorescence at the reference state as a function of [IPTG] for both markers. Due to the structure of the genetic toggle network and the biological function of the inducer, the average fluorescence exhibited a monotonically increasing, sigmoidal dependence on [IPTG]. A relatively sharp rise in average fluorescence was observed for both pTAK117 and pGFPaav between 20 and 60 μM, above which the average fluorescence remained at maximal levels. Since the green fluorescent protein encoded by *gfpaav* has a shorter half-life than that encoded by the *gfpmut3* gene present in pTAK117, the average fluorescence levels of the cell cultures with the former gene were lower than those carrying the latter.

One would obtain only limited understanding of network behavior if the experimental characterization were based solely on average population properties. A much more in-depth insight was obtained by studying the fluorescence distributions of the entire cell population. Figures 3.5a and 3.5b show the

![Figure 3.4: Average green fluorescence at the reference state as a function of inducer concentration. Filled circles: pTAK117. Open circles: pGFPaav. Error bars represent the standard deviation based on triplicate experiments. Some error bars are smaller than the markers. Reproduced from Portle et al. [64].](image-url)
number density functions for different inducer concentrations at the reference state for both pTAK117 (figure 3.5a) and pGFPaav (figure 3.5b). First, notice that expression levels varied by one to even three orders of magnitude for some [IPTG]. Thus, the cell populations appear very heterogeneous while the extent of heterogeneity strongly depends on [IPTG].

Moreover, for low (below 30 μM) and high (above 50 μM) [IPTG], all number density functions were unimodal. However, in the intermediate range of [IPTG], where the transition from low to high average expression level occurs (see figure 3.4), the number density functions were bimodal. There appears to be a clear threshold at the single-cell level for these intermediate IPTG concentrations. The cells below and above this threshold formed well-defined subpopulations, which led to the observed bimodality. We note that Gardner et al., (2000) also reported bimodal distributions for pTAK117 at [IPTG] = 40 μM using a different experimental protocol, while our results show bimodal number density functions for [IPTG] between 30 and 50 μM for both reporter proteins. Therefore, these results indicate that the specific genetic architecture at the single-cell level leads to this characteristic bimodality feature at the cell population level for intermediate inducer concentrations.

Despite the fact that the range of [IPTG] for which the number density function is bimodal was the same for both reporter proteins, there were significant differences between pTAK117 and pGFPaav in the value of the single-cell threshold separating the two subpopulations forming the bimodal distributions. Specifically, as estimated by eye in figures 3.5a and 3.5b, the switch between low and high expression levels occurred at relative fluorescence values around 300 and 4 for pTAK117 and pGFPaav, respectively.
Figure 3.5: Number density functions at the reference state for different inducer concentrations. a) pTAK117. b) pGFPaav as determined by flow cytometry as described in methods. Included are scatter plots of FL1 vs. FSC for each variant, showing that each subpopulation of the bimodal distributions have the same FSC (size) range, and therefore differences in fluorescence are not due to age of the cells. Reproduced from Portle et al. [64].
Notice also that for both reporter proteins, the unimodal number density functions at high IPTG concentrations practically overlap with each other. Moreover, the high fluorescence parts of the bimodal number density functions obtained at intermediate induction levels appear to be centered around the same mean fluorescence value. Thus, it appears that for [IPTG] above 30 μM and for a range of [IPTG] nearly two orders of magnitude (up to 2 mM), all cells crossing the single-cell threshold became maximally induced. On the contrary, for low [IPTG], the population shifted toward higher expression levels as [IPTG] increased, indicating that the balance between the lac and λ repressor concentrations shifted to allow slightly higher expression levels of GFP. This behavior is more visible in the case of pTAK117 (figure 3.5a) where the single-cell fluorescence threshold for induction was approximately 75 times higher than in the case of pGFPaav. Due to the low GFP level at low [IPTG] in pGFPaav, the fluorescence level for some sub-threshold cells fell below the quantitative sensitivity of the flow cytometer, and some were below the four decade range of the cytometer, making it difficult to compare number density functions at low expression levels for the two systems. However, since the half-life of the reporter expressed by pGFPaav is much closer to that of the λ repressor, the sharper separation between the two subpopulations forming the bimodal distributions in the case of pGFPaav and the significantly larger separation between the mean low and high fluorescence values of unimodal distributions is believed to be a more accurate representation of network behavior.

A more complete view of the gene switching dynamics was obtained through transient studies. Figures 3.6 and 3.7 show the time evolution of the green fluorescent number density function for both pTAK117 and pGFPaav and for [IPTG] = 40 μM and
60 μM, respectively. In both cases, the initial number density function (t = 300 min) was unimodal. In the first case (figure 3.6), for both reporter proteins, a second subpopulation grew out of the main population in the middle of the exponential growth phase. Notice also that the subpopulation consisting of cells at high induction levels became gradually more significant until the reference state was reached, where the overall number density function obtained a stable bimodal shape. As also shown in figure 3.5, the separation

**Figure 3.6.** Time evolution of the number density function at [IPTG] = 40μM for a) pTAK117 and b) pGFPaav as determined by flow cytometry as described in methods. Reproduced from Portle et al. [64].

**Figure 3.7.** Time evolution of the number density function at [IPTG] = 60μM for a) pTAK117 and b) pGFPaav as determined by flow cytometry as described in methods. Reproduced from Portle et al. [64].
between the low and high expressing peaks in the bimodal number density function was more pronounced in the case of the reduced half-life GFP both transiently and at the reference state. This qualitative route towards bimodality is representative of all intermediate inducer concentrations where bimodal number density functions were obtained at the reference state ([IPTG] = 30-50μM).

However, bimodality was observed transiently even in cases where the number density function was unimodal at the reference state (see figure 3.7b). In the case of the reduced half-life GFP, as the population passed through the single-cell switching point, a separate subpopulation at high fluorescence was formed and continued to co-exist with the low fluorescence subpopulation, thus resulting in a bimodal number density function. However, in contrast to the case of [IPTG] = 40 μM (figure 3.6b) the percentage of cells below threshold continually decreased with time until it completely vanishes. Thus, the number density function eventually became unimodal at the reference state due to the higher concentration of IPTG, which enabled even the low expressing cells to become fully induced after some period of time. Contrary to the results for pGFPaav, transient bimodality was not observed in cells containing pTAK117 at [IPTG] = 60 μM (figure 3.7a). The number density function became distorted as it passed through the single-cell switching point of 300 but it never became bimodal. This is attributed to the much higher stability of the protein expressed by gfpmut3 compared to that of the λ repressor, which does not allow enough time for visualizing the transition of cells from low to high λ repressor expression levels that actually occurs at the single-cell level.
3.6 Summary and Conclusions

The interplay between a prototype genetic architecture and the distribution of phenotypes at the cell population level was studied with the use of flow cytometry. The network under investigation was a gene-switching network, known as the genetic toggle [37], consisting of two promoter-repressor pairs. Two green fluorescent proteins of different half-lives were placed after the λ repressor as reporters in order to study the effect of network structure on cell population dynamics as well as assess the potential of using an extracellular inducer (IPTG) to control the distribution characteristics.

Both reporter proteins exhibited the same sigmoidal induction patterns for the average expression levels. However, experimental characterization with flow cytometry offered a more in-depth view of network behavior. Specifically, for very low and very high inducer concentrations, the distribution of fluorescent phenotypes was found to be unimodal at a well-defined, quasi-time-invariant reference state. On the contrary, at intermediate [IPTG], where the average fluorescence switched from low to high expression levels, the distributions become bimodal consisting of two subpopulations below and above a specific single-cell threshold. The region of inducer concentrations where bimodality was observed at the reference state was the same for both reporter proteins. Moreover, bimodality was not only observed at the reference state. Even at inducer concentrations where the distribution eventually becomes unimodal, transient studies revealed that as the cell population passed through the single-cell induction threshold, two subpopulations were formed for some period of time. Thus, the existence of a single-cell threshold and the bimodal shape of the fluorescence distribution is a robust pattern of this particular genetic network. We say this with confidence not only
from our results, but from previous modeling work done by Mantzaris, who has also studied other networks [28, 74]. Using a cell population balance model able to capture distribution dynamics, Mantzaris confirmed the ability of this network to achieve stable bimodal distributions in a narrow range of IPTG concentrations [75], qualitatively matching our experimental results.

In studying the distributions in the case of the reduced half-life GFP protein, the single-cell induction threshold of fluorescence was found to be significantly lower. Thus, the corresponding bimodal distributions were visibly more asymmetric for the system with the higher GFP degradation rate. Moreover, unlike the case of the reduced half-life GFP, it was not possible to visualize the bimodal transition of the population from low to high expression levels for high inducer concentrations using the more stable GFP. Since the half-life of the λ repressor is much closer to that of the less stable GFP, the results obtained with the gfpaav gene downstream of the λ repressor at the reference state, as well as transiently, offer a more accurate representation of the genetic toggle behavior. However, the results obtained with the long half-life GFP are also valuable in understanding how the expression levels will be distributed amongst the cells of the population when a high-stability protein is placed under the control of the genetic toggle network downstream of the λ repressor.

The wide spread of fluorescent phenotypes from one to three orders of magnitude depending on the inducer concentration and the inability of the average population properties to fully characterize network behavior indicate the importance of taking into account cell population heterogeneity when designing such a gene-switching network for biotechnological applications.
Chapter 4

Oscillatory Network

4.1 Motivation

In the gene-switching network, we found out how to control the distribution of phenotypes at the cell population level at steady state by varying the concentration of inducer at which the system is exposed. We controlled what fraction of cells were induced to the high state.

In contrast to the gene-switching network, the oscillatory network we studied, termed the repressilator by its makers [39], has the capacity to be controlled both dynamically and at steady state. The network is composed of three promoter-operator-repressor sets from different natural regulatory networks, connected to form a cyclic, cascading network of negative feedback loops. This network can express stable or unstable steady states, leading to controlled, sustained expression of one repressor or oscillations in the three repressor levels. In the case of the oscillatory system, we look not only how to control the distribution of phenotypes by varying inducer concentrations (in this case, two inducers can be used), but to control the dynamic behavior of the system as well.

Why is this system worth studying? Both behaviors of the repressilator network will have different effects on cell population heterogeneity that will be unique, and both can be industrially useful, adding to the repertoire of existing platforms for controlled

---

1 The modeling work discussed in this chapter has been performed by Dr. Sergio Iadevaia. Part of this chapter will be submitted for publication.
gene expression. In particular, if other desired genes are transcribed downstream of the repressors, the expression of those genes will mimic that of the repressor genes. The expression will depend on what type of behavior the system is displaying. One repressor can be preferentially expressed at a given inducer concentration, and this control can be changed over time if the inducer is changed. For other inducer input concentrations, oscillatory behavior is expected, which can potentially control the genes' expression in a cyclic way over time. Several possible applications are envisioned (and many more are possible). One application is drug delivery, allowing controlled synthesis of desired molecules over time. If one introduces precursor enzymes of polyhydroxyalkanoate monomer units to the network, a block co-polymer made up of two or three monomer units could be produced [76]. If metabolic enzymes are inserted in the network, general metabolic pathways could be controlled over time. This would allow bacterial resources, for example, to be channeled toward a certain goal. Due to the extra POR set in the repressilator, more behaviors are possible than in the toggle network, and there is flexibility in determining such behavior through the presence and concentration of inducers.

In addition, oscillatory networks are imbedded in our genetic makeup, and in that of many other organisms. Examples of natural, genetic-based oscillations include circadian rhythms and the cell cycle [77, 78]. Circadian genetic networks do not have the same architecture as does the repressilator (they contain both positive and negative regulatory elements [77]), but we are not trying to replicate a circadian network. Understanding the connection between a model oscillatory genetic network such as the repressilator and cell population heterogeneity will provide valuable insights into the
4.2 The Repressilator Network

Our focus in this work is on the genetic oscillator known as the repressilator, made by Elowitz and Leibler [39]. The design involves three promoter-repressor pairs, which form a cyclic, cascading network of negative feedback loops (see figure 4.1a). The three repressors are the lac repressor, the λ cl repressor from phage λ, and the tet repressor from the tetracycline-resistance transposon Tn10. The genes are arranged such that the tet repressor inhibits the λ repressor, which inhibits the lac repressor, which in

![Diagram](image)

**Figure 4.1.** The oscillatory network. a) The network architecture, consisting of three promoter-repressor pairs, forming a cyclic cascading network of negative feedback loops. Network is controlled by two inducers, aTc and IPTG. b) Relevant plasmid maps. Repressilator and Reporter are the original two-plasmid system. They were combined to form pSCP102, which is the plasmid used in this work.
turn inhibits the \textit{tet} repressor. The \text{P}_{\text{tetO-1}} promoter controls expression of the \(\lambda\) repressor, the \(\lambda\)-\text{P}_{\text{R}} promoter controls expression of the \textit{lac} repressor, and the \text{P}_{\text{LlacO-1}} promoter controls expression of the \textit{tet} repressor. All repressors contain a \textit{ssrA} tag to reduce their half-life. Figure 4.1b shows the plasmid designs for the Repressilator plasmid holding all the regulatory genes, and for the Reporter plasmid which expresses a green fluorescent protein (GFP) mutant with a half-life of approximately 60 min [56]. GFP is a marker for the \(\lambda\) repressor, since both are controlled by the same promoter. Theoretically, this system can reach a stable steady state (dominant expression of one repressor) or an unstable steady state (exhibits oscillatory behavior) [39], and Elowitz and Leibler confirmed with fluorescence microscopy that their network did oscillate in at least 40\% of the cells.

\section*{4.3 Objective}

In the work presented by Elowitz and Leibler, the authors designed a network made from three promoter-repressor pairs via modeling, analyzed it theoretically to gain insight on how to best design the components to reach an unstable steady state, built it, and confirmed its behavior experimentally. Fluorescence microscopy was used to determine the network behavior at a single-cell level. One of the chemical inducers for the system, isopropyl-\(\beta\)-D-thiogalactopyranoside (IPTG), which complexes with the \textit{lac} repressor, was used to synchronize the network oscillations at the beginning of an experiment, and was used in control experiments. Anhydrotetracycline (aTc), which complexes with \textit{tet} repressors, is the other extracellular inducer for the system, but it was not used in experiments. The concentration of inducers is a user-controlled parameter.
that could provide additional flexibility in using this network biotechnologically, and this has not been investigated.

We performed experiments to better understand the potential of this system. We investigated this network’s full range of expression by running a series of shake flask experiments with both a single inducer, aTc, and with mixed inducers, using aTc and IPTG. Flow cytometry was used to measure the distribution of GFP in the cells. This high-throughput method quantifies light scattering and fluorescence properties of individual cells. The relevant parameters to this study are the two light scattering properties, forward scatter (FSC) and side scatter (SSC), and green fluorescence (FL1). Our aim was to find the effect of aTc and IPTG both transiently and at balanced growth.

4.4 Materials and Methods

**Strain:** *E. coli* strain MC4100 (lac') was used. The cells were made chemically competent and transformed with plasmid pSCP102 as described in *RbCl Transformation Procedure for Improved Efficiency* [65]. This procedure was used for all plasmids tested, always in MC4100.

**Plasmid construction for main experiments:** A one-plasmid construct was created from the two plasmids making up the original system. The Repressilator plasmid has a copy number of 3-4, and the Reporter plasmid has a copy number of 50-70 [79]. The authors apparently wanted to keep the number of repressors tightly regulated by keeping the copy number of the regulatory genes low. However, in order for GFP expression to be visible on the microscope, GFP had to be expressed on a higher copy number plasmid.

For several reasons, we decided to combine the two-plasmid system into one
plasmid. First, it is more difficult for cells to carry two plasmids than one. It is more of a burden to the cell, and it requires two antibiotics and resistance mechanisms operating in the cell. In choosing which antibiotic marker to use, the kanamycin marker was chosen. Since the ampicillin resistance gene encodes for an enzyme that can function in the extracellular media, cultures have been known to exhibit plasmid instability when carrying this marker. The kanamycin marker expresses an intracellular protein, so it does not have this problem. Furthermore, keeping the concentration of repressors low in the cell contributes to stochastic, intrinsic heterogeneity.

It was noted by the authors that cells often lost the ability to oscillate after division. We thought that, perhaps, if the number of regulatory molecules was increased, and partitioning was not very unequal, then the cells might be more likely to continue oscillating after cell division. Spudich et al. also applied this reasoning when considering synchronization of cell growth [9]: they hypothesized that the quick asynchrony they observed was due to low concentrations of enzymes, and that significantly increasing the concentration of rate-limiting enzymes involved with the cell cycle could allow cultures to remain synchronized. There are stochastic modeling studies that also support the idea that redundancy can help establish a consensus among cells [31, 32]. For these reasons, a high-copy plasmid was used for the construct. In this way, GFP would still have high expression levels.

As a side note, in the two-plasmid system, the number of tet operator sites (binding sites for repressors) was more than 10-fold higher than the number of lac and λ operator sites due to the inequality of the two plasmids’ copy numbers. Although this ratio of operator sites was a “perturbation” on the system due to the titration effect on tet
repressors [39], the ratio must have helped, not hindered, the ability of the network to oscillate. In combining the two plasmids, the number of operator sites becomes symmetrical, so we may need to titrate the \textit{tet} repressors in order to achieve oscillation, assuming that this behavior is still possible. The one-plasmid construct is called pSCP102, and its map is found in figure 4.1b. Its construction is outlined in Appendix A.

\textbf{Plasmid constructions for testing interaction hypotheses:} A generic construction outline is placed here. The plasmids are listed in table 4.1, and more specific plasmid information, including plasmid maps, are in Appendix A. Since appropriate restriction enzyme sites were not available in most cases, PCR (using either Platinum \textit{Pfx} DNA Polymerase or PfuUltra High-Fidelity Polymerase, Stratagene) was necessary for nearly every construct. The resulting fragment(s) was either digested with appropriate restriction enzymes or underwent a kinase reaction (T4 Kinase, New England Biolabs). Either a unimolecular or bimolecular ligation followed with T4 DNA Ligase or Quick T4 DNA Ligase (New England Biolabs). Colonies were transformed either into TOP10 SuperCompetent Cells (Invitrogen) or chemically competent DH5\textgreek{a} cells before being transformed into MC4100 (RbCl procedure [65]). All promoters, genes, and terminators were confirmed by sequencing.

\textbf{Shake flask experiments:} There were four sets of experiments, two with aTc only, and two with aTc and IPTG as inducers. All of them kept the cells in exponential phase in a narrow OD\textsubscript{600} range with frequent subculturing steps. The basic procedure is common between them. First, cells were grown overnight at 32°C, shaking at 250 rpm in an orbital shaker, in 5 mL of LSRB medium (with 50 mg/L kanamycin) for 12 hours and then subcultured at an OD\textsubscript{600} of about 0.0001 (∼200,000 cells/mL) in 50 mL of
prewarmed and aerated LSRB medium without inducers in a 250 mL Erlenmeyer flask capped with foam to allow oxygen transfer. In order for the culture to reach exponential phase, it was grown at the same conditions until it reached an OD$_{600}$ of approximately 0.01. Then the culture was exposed to inducers. It was subcultured 1:9 in 45 mL fresh, warmed, aerated LB with the appropriate concentration of aTc and IPTG. Subsequent dilutions in the same ratio were made in fresh media (4 to 6 times total) every time the OD$_{600}$ reached approximately 0.01. Samples were withdrawn about three times per flask (about once a generation) until the culture’s average fluorescence (as determined by flow cytometry) did not change. Cells were immediately washed twice in cold PBS. The samples were kept on ice and shielded from light before being measured with the flow cytometer.

**aTc-only experiments:** The culture was either started in an uninduced state or an induced state. In the first case, after reaching exponential phase, it was exposed to aTc at concentrations between 1 and 1000 nM. In the second case, it was first grown at 1000 nM aTc for two subculturing steps before being subcultured into media with a lower concentration of aTc, between 1 and 850 nM. Triplicate experiments were performed.

**aTc-IPTG experiments:** In the third set of experiments, cells were grown in media containing aTc (at 500 or 1000 nM) and IPTG (0 M to 600 μM) immediately after reaching exponential phase. In the fourth set, they were grown in media containing only aTc (1000 nM) for two subculturing steps before being subcultured into media containing the same concentration of aTc with the addition of IPTG (50 to 600 μM).

**Control experiments:** Experiments were set up to confirm that the oscillation in average fluorescence that was observed was not an artifact of the successive dilutions of the
culture. Duplicate experiments were performed – once the culture was returned to exponential phase after being grown overnight, it was diluted by 10 in one flask, and by 20 in another, such that the culture in the second flask would be subsequently diluted one generation later than the first. This yielded identical evolutions of fluorescence over time, showing that the culture was not affected by the act of dilution (see figure 4.2).

**Interaction hypothesis experiments:** The same general procedure described above in “Shake flask experiments” was followed, except no inducers were used. Only 2 dilution steps into fresh media were used for these experiments, instead of 4 or 6, long enough for the cultures to reach steady averages and distributions of green fluorescence. Triplicate experiments were performed when a non-negligible interaction was detected.

**Flow cytometry:** The flow cytometer used was a FACScalibur. All parameters were logarithmically amplified, with the following settings: FSC E01, SSC 381 V, FL1 700 V, FL2 500 V, FL3 675 V. A side scatter threshold was applied to gate out much of the noise (at channel 130). During acquisition, all events were saved until the software counted 5,000-10,000 events with a minimum FSC channel value of 512 (assumed to be cells) for each sample. In the interaction hypothesis experiments, the FL1 setting was lowered to 500 V, due to the

![Figure 4.2. Average fluorescence over time at 150 nM aTc, showing that time of dilution has no major effect on evolution of fluorescence. Open markers follow the fluorescence for each culture, while the closed markers indicate the times of dilution, which is different for the circles compared to the squares and triangles.](image-url)
strength of the GFP expression.

**Gating data:** Gates were applied in the post-processing of the flow cytometer files according to light scattering. Since *E. coli* cells are so small and close to the size of the debris being sensed by the cytometer, gates were applied to better identify the cells. After observing the change in light scattering values while the culture was in exponential phase, FSC and SSC gates were applied to include events above a linear value of 80 (channel value of 487) and 40 (channel value of 410), respectively.

**Software:** Flow cytometer fcs files were processed as outlined in Chapter 2.

### 4.5 Reaching the Reference State

As in the gene-switching network experiments, we developed a method to determine when cells are in a similar physiological condition, when their averages and distributions of green fluorescence are characteristic and comparisons are meaningful (as close to the state of balanced growth as we can experimentally define). Our requirements for the "reference state" are that the light scattering averages, green fluorescence averages, and green fluorescence distributions are relatively time-invariant.

The shake flask procedure that was established for these experiments (see Materials and Methods section) facilitates reaching this state. By repeatedly subculturing the cells, thereby keeping cells in the exponential phase at similar OD<sub>600</sub> values during the entirety of the experiment, the FSC and SSC values stay constant (except near points of perturbation). Once the fluorescence properties stabilize, the reference state is determined (see figure 4.3 for visualization of this concept with data). It can be reached in a few generations or after many generations, depending on the experimental
conditions. This is in contrast to our work on the gene-switching network, in which light scattering and fluorescence parameters plateau during the same period of time [64].

Using shake flask experiments and applying the requirements to find reference states, the oscillatory system was characterized by varying the concentration and order of inducers. The system was found to have many interesting features. The aTc-only experiments revealed three steady states with hysteresis, population level average dynamics that show oscillatory behavior and other interesting features, and distinct number density functions for each state. In the mixed aTc-IPTG experiments, greater flexibility is seen in the ability to control average fluorescence and coefficient of variation, and the order of addition of IPTG determines the inducer's effect on the dynamic behavior of the culture.

4.6 Experiments with aTc reveal three steady states with hysteresis

In the single-inducer experiments, aTc was chosen instead of IPTG. Without any inducer, expression of GFP is low. In order to differentiate the GFP expression under varying conditions, higher levels of GFP needed to be reached. As the aTc concentration increases, more tet repressors become bound to aTc. This depletes the tet repressor population bound to the tet operator sites, which, according to the network architecture...
(refer to figure 4.1a), increases the frequency of GFP expression.

The aTc concentration was varied, starting from both uninduced and induced conditions, and the average fluorescence at the reference states, average fluorescence over time, and corresponding number density functions were determined.

First we look at the data at the reference states for each condition.

As expected, as [aTc] increases, the average fluorescence of the culture increases. However, the increase is not monotonic; instead, as shown in figure 4.4, the system exhibits bi-threshold behavior, and the cultures reach one of three steady states. The low state is near a relative fluorescence value of 10, the middle state near 80, and the high state near 240. No other artificial genetic network yet made has been shown to reach three distinct states.

It is clear from figure 4.4 that there is a considerable amount of overlap between the three states. Hysteresis is seen between the low and middle states at [aTc] = 100-150 nM, and is seen between the middle and high states at [aTc] = 500-850 nM. The variations in ending state were found when experiments were performed on different days or with different initial states (uninduced or induced). Starting from different initial states allowed us to view...
more complete, expanded ranges of multiplicity.

The number density functions (ndfs) are unique and consistent in shape for each state, regardless of the initial condition. Figure 4.5 shows sample ndfs for each state: 4.5a from experiments starting with uninduced cultures, 4.5b from experiments starting with induced cultures. One can see how the two sets almost mirror each other, and ndfs within each state overlap each other.

Comparing the characteristics of the number density functions for each state leads to several observations. For one, the middle state is not a combination of the low and high states. The average fluorescence values of the middle state seen in figure 4.4 are not the result of some cells in the population being at the high state and the rest being at the low state – the

![Figure 4.5. Typical number density functions at the reference state for low, middle, and high states, including data at various aTc concentrations. The initial culture was grown at a) 0 M aTc or b) at 1000 nM aTc, yet both sets are consistent with the other. The middle state ndf is broader than that of the other two states.](image)
cell population is centered in between the high and low states.

Something that sets the middle state number density function apart is that, while the low and high state distributions look symmetrical on this logarithmic scale, the middle state distribution does not. It looks flattened (it is at some points in time separated into two peaks) and has a long right tail. The distribution of green fluorescence (and therefore that of GFP expression in individual cells) is more heterogeneous than in the low and high state.

One surprising result is that none of the number density functions are bimodal. This implies that all cells in the culture choose one path or another. Whatever combination of factors exist that cause a cell to reach one state or another, it is felt in the same way by the entire population.

4.7 Transient studies reveal the middle state is oscillatory while others are stable

Although cultures at many different aTc concentrations end up in the same state, the path they take to get there varies. Figure 4.6 shows the average fluorescence over time exhibited when aTc is first introduced to the system. Figures 4.6a, 4.6b, and 4.6c include dynamics that asymptotically reach the low, middle, and high state, respectively. The middle state is oscillatory, even at the population level, while the low and high states are not. We suspect that the middle state corresponds to an unstable steady state, and the low and high states correspond to stable steady states. aTc serves the same purpose as the extra tetO sites on the Reporter plasmid in the original two-plasmid system – it titrates out free and DNA-bound tet repressors, and, in a certain range, allows the network production of GFP to oscillate.

The single-cell behavior of cells in the middle state may help explain the wider
Figure 4.6. Average green fluorescence as a function of time at various aTc concentrations. Plots are grouped by initial and final states. Cultures that start uninduced and end in the a) low state; b) middle state; c) high state. Cultures that start induced, ending in the d) low state; e) middle state; f) high state. The low and high state trajectories appear stable while the middle state trajectories appear oscillatory.
distribution of this state. When aTc is added at time = 0 min, the culture is synchronized to an extent, and therefore oscillations, though damped over time, can be observed on average. The damping effect is an inevitable result of heterogeneity. Since there is no external control to synchronize the oscillations, individual cells may become desynchronized and still oscillate, or some cells may partition the regulatory components so unequally that oscillation ceases completely. If some cells are still oscillating at later time points (but desynchronized), even after the population average fluorescence is unchanging, then this variation in single-cell fluorescence levels could account for the larger CV of the middle state. For the low and high states, since the GFP expression is steady, the distribution of fluorescence could be narrower.

Each state has average fluorescence dynamics that give insight into the relationship between aTc concentration and network behavior.

For instance, figure 4.6a shows different typical trajectories for the low state: at 1 and 10 nM, there is very little change in average fluorescence over time, but at 50 and 100 nM, there is an initial overshoot in fluorescence. At low aTc concentrations, there is basically no change from the GFP expression with no aTc. Once the aTc concentration gets high enough, it will complex with enough tet repressors that GFP expression is increased. However, the concentration is not high enough to sustain oscillations, so once the network adjusts to the lower amount of free tet repressors, fluorescence decreases.

As shown in figure 4.6b, almost all dynamics of the middle state show a damping oscillation. At the network level, there is enough aTc in the cell to allow λ repressors to initially dominate. As the λ repressor level increases, the tet repressor expression becomes less inhibited and tet repressors start to accumulate. Even though some of the
tet repressors complex with aTc, enough are present to downregulate λ repressor expression, which in turn upregulates lac repressor production. Lac repressors reduce tet repressor expression, and the cycle continues. The internal aTc concentration in this state serves as a balancing weight, making sure that the tet repressor concentration remains low enough that, in the turn of the cycle, lac repressors can deplete the tet repressor level in order to allow λ repressor expression. Even if the oscillatory behavior does not present itself initially (for [aTc] > 500 nM), it emerges by the end of the experiment. For example, as seen at 750 nM aTc, the asymptotic average fluorescence level is overshot more than it is at lower aTc concentrations, but then it slowly returns to damped oscillations after the network adjusts to the level of free tet repressors.

The high state dynamics are shown in figure 4.6c. They show a typical first order response – a swift rise to high fluorescence. This response is the same for all experiments resulting in the high state. There is a saturating amount of aTc, depleting the concentration of free tet repressors to such an extent that the λ repressors are dominating in the system, even though the expression of lac repressors is relatively uninhibited.

The dynamics of the average fluorescence of the cultures that were fully induced at 1000 nM aTc at time = 0 min are also shown in figure 4.6, parts d through f. One can see in figure 4.6f that lowering the aTc concentration as low as 750 nM aTc provided no perturbation in GFP expression, so the cultures remained at the high state for the duration of the experiment.

The middle state dynamics, shown in figure 4.6e, are similar in nature to those from cultures that were uninduced at time = 0 min. The overshoot (in this case to lower fluorescence) is still present, as is the oscillatory nature. One noticeable trend is that as
[aTc] increases, the oscillations are lower in amplitude and are damped more quickly.

The low state dynamics (figure 4.6d) are relatively slow compared to the high state dynamics for cultures starting at an uninduced state (the fall to low fluorescence is slower than the rise to high fluorescence). However, the dynamics still appear to be a first order response, with an initial lag. The dynamics at 150 nM aTc are different than the rest; they appear almost oscillatory, but in contrast to the middle state dynamics, they tend toward the low state.

The initial lag shown in the low state dynamics is also seen in the middle state. The difference in dynamics between initially uninduced (no lag) and induced cells may be explained by aTc dynamics. Starting with uninduced cells, assuming that aTc enters the cell quickly, there should be little lag in the fluorescence change, as the network will immediately be affected (since the GFP maturation is quick). In contrast, when decreasing the aTc concentration, first the cells must divide to dilute out the existing aTc-tet repressor complexes before the network will experience any noticeable change in the complex concentration.

With aTc, the user can extracellularly control this three promoter-repressor architecture to achieve three different steady states, and the middle state is characterized by damped oscillations at the population level.

4.8 Discrepancy exists between existing model and experimental data

The aTc-only reference point results include bi-threshold behavior as well as multiplicity. Also, the aTc-IPTG results indicate multiplicity depending on initial condition. Looking to the model developed by Elowitz and Leibler, and adding the aTc-
tet repressor interaction to include the network’s dependence on inducers, the experimental and simulation results do not qualitatively match. Theoretically, if the network behaves as designed, with each repressor interacting only with the promoter containing its complimentary operator site, there is a continuum of steady states and no multiplicity [80]. The results cannot be explained by that model. One could suppose that the experimental data are incorrect due to some artifact, but we believe the results are genuine, especially the features of multiplicity and bi-threshold behavior that put the results at odds with the original model. After carefully looking over the data and chronology of experiments, some factors that might have caused an artificial multiplicity result were considered then discarded, such as effects due to aTc degradation, the initial OD600, and the exact length of the growth period without inducer. So, we then set out to find why the network behaves in this manner.

One of the possible causes that differentiate the gene expression designed at the single-cell level from the gene expression that is measured at the population level is cell population heterogeneity. Heterogeneous cells exist in different physiological states and therefore exhibit cell-to-cell phenotypic variability. The variability of cells harboring the artificial repressilator, however, must be very profound to justify such differences in the network dynamics at the single-cell and population level. Thus, instead of regarding heterogeneity as the sole cause that creates multiplicity of the repressilator gene expression at the population level, it appears more reasonable to assume that the cell population exhibits the observed bi-threshold behavior because each individual cell of the population has the ability to display the exact same behavior. From this perspective, heterogeneity can be regarded as the source of variability that drives the cells towards one
of the multiple states.

The only way to see multiplicity of steady states at the single-cell level is for there to be extra interactions in the network influencing the action of the repressors, either from some external force or from within the network. For example, some molecules external to the network could influence the interactions between the repressors and operator sites. Or, the repressors themselves might act in unexpected ways, for example, by binding to promoter-operator sites without their specific operator sequences. If the interaction is strong enough, it could influence the network's behavior.

Since repressors are known to have measurable nonspecific interactions with non-operator DNA ([81-83] and others), we continued with this idea. We modeled the inhibition of each repressor on its own expression and on the expression of the repressor upstream of it. Concurrently, we experimentally tested the existence of these interactions. The "self-promoting" interactions of each repressor binding to the promoter of the upstream repressor are: tet repressor-\(\lambda\) P\(_R\), lac repressor-P\(_L\)tetO-1, and \(\lambda\) repressor-P\(_L\)lacO-1, as shown in figure 4.7 (dashed lines). The "self-inhibiting" interactions of each repressor binding to its own promoter are: tet repressor-P\(_L\)lacO-1, lac repressor-\(\lambda\) P\(_R\), and \(\lambda\) repressor-P\(_L\)tetO-1 (dotted lines).

4.9 **Modeling extra network interactions leads to qualitative match with experimental data**

In order to model the dynamics of the repressilator in the presence of nonspecific repressor-promoter interactions, Dr. Sergio Iadevaia modified the mathematical model derived by Elowitz and Leibler to explicitly account for a) the ability of each protein to
nonspecifically complex with the upstream promoter as well as its own promoter, b) the effect of the aTc and IPTG inducers, and c) the dilution effect due to cellular growth [84]. His explanation and derivation of the governing equations can be found in his thesis [85], with an abbreviated version included in appendix D.

Also included in Iadevaia’s thesis are extensive modeling studies and explanation of how the repressilator system is affected by the addition of these extra interactions. First he simulated the system without addition of the extra interactions (using a modified Elowitz and Leibler model). After linking the system behavior to aTc and IPTG, he found that the GFP expression level is a monotonically increasing function of [aTc] (see figure 4.8). The expression can change from an unstable steady state to a stable steady state. There is no bi-threshold behavior, no multiplicity, so the main features of the experimental results are not captured by this model. This is in sharp contrast to what he observed after adding the extra nonspecific interactions (see figure 4.9, with parameters in appendix D) – he can qualitatively capture the behavior of the experimental system, with three states that have multiplicity between them, bi-threshold behavior, and an oscillatory middle state. In his work, he experimented with the various possibilities of certain nonspecific interactions

Figure 4.7. The regulatory repressor-promoter interactions that define the structure of the artificial repressilator. We hypothesize the existence of six nonspecific interactions. Three are “self-promoting” interactions due to the binding of 1) the tet repressor to the λ-Pr promoter, 2) the λ repressor to the P_lac0-1 promoter, and 3) the lac repressor to the P_tet0-1 promoter (dashed lines). The other three are “self-inhibiting”, due to the binding of 1) the lac repressor to the λ-Pr promoter, 2) the tet repressor to the P_lac0-1 promoter, and 3) the λ repressor to the P_tet0-1 promoter (dotted lines). Courtesy of Dr. Sergio Iadevaia.
being present while others are absent.

This summarizes what he learned: the addition of two self-promoting interactions is required to create the low and high states (with the bi-threshold behavior) from the middle state (which is the state resulting from the known interactions); the addition of a third self-promoting interaction allows the three states to be reached for a wider range of parameter values; addition of self-inhibiting interactions limits the range of multiplicity between the states and brings the two additional states in a range of expression levels that more closely resembles experimental data.

4.10 Experiments show existence of extra network interactions

With the knowledge from simulations that adding extra interactions between a
repressor and its own promoter and the promoter of the upstream repressor to the model can qualitatively capture the network behavior, we set off to find the existence of these interactions experimentally.

A simple network was created to test each hypothesized interaction, in which a constitutively expressed repressor negatively controls GFP expression by (possibly) binding to the GFP promoter. The plasmid pSCP102 was simplified using standard molecular biology techniques into separate test plasmids for each interaction. Each plasmid consists of a repressor being expressed from the same promoter as in pSCP102 and GFP being expressed from a promoter with the hypothesized interaction. We test for a difference in GFP expression when the repressor protein is being expressed and when it is not.

In order to prevent repressor expression, a second plasmid was created for each hypothesized interaction by removing the promoter in front of the repressor (including its translation start site), which should significantly reduce its expression and relieve any GFP repression. This approach to knocking out repressor expression was chosen in order to minimize the difference in size between plasmids, since significantly lowering the plasmid size by removing the full repressor gene could lessen the burden on the cells and create an artificial increase in GFP expression. This approach was applied for each interaction to keep the method consistent, to avoid any inducer effects on cell growth, and to avoid the question of how much inducer is required to reach maximum GFP expression.

The gene order on each plasmid was kept consistent (gfpav upstream of repressor gene) to prevent any leaky expression of the repressor due to promoter activity
in the area upstream of the first promoter [86]. The plasmids retain a ColEl origin of replication and kanamycin resistance, just like pSCP102. The test plasmids are summarized in table 4.1 with the arrangement of their respective promoters and genes. If the hypothesized interaction exists, there will be an increase in GFP expression when the promoter for the repressor is removed. If there is no interaction, there will not be any difference in GFP expression level.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Regulatory elements of each hypothesis plasmid, in order</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$\lambda P_R$ gfpaav terminator P$l$acO-1 TetR terminator</td>
<td>pSCP141</td>
</tr>
<tr>
<td>II</td>
<td>$\lambda P_R$ gfpaav terminator TetR terminator</td>
<td>pSCP151</td>
</tr>
<tr>
<td>III</td>
<td>P$l$tetO-1 gfpaav terminator $\lambda P_R$ LacI terminator</td>
<td>pSCP142</td>
</tr>
<tr>
<td>IV</td>
<td>P$l$tetO-1 gfpaav terminator P$l$tetO-1 $\lambda$ cI terminator</td>
<td>pSCP143</td>
</tr>
<tr>
<td>V</td>
<td>$\lambda P_R$ gfpaav terminator P$l$acO-1 TetR terminator</td>
<td>pSCP154</td>
</tr>
<tr>
<td>VI</td>
<td>$\lambda P_R$ gfpaav terminator P$l$acO-1 TetR terminator</td>
<td>pSCP145</td>
</tr>
</tbody>
</table>

4.10.1 lac repressor interactions (II and V)

The interaction between the lac repressor and P$l$tetO-1 is tested with plasmids pSCP142 and pSCP152. The first plasmid contains the $\lambda P_R$ promoter in front of LacI, while in the second one it is deleted. A similar growth method to that used in the main repressilator experiments was used, repeatedly subculturing cells containing each plasmid in media without inducer. A difference is observed between GFP levels. With expression of LacI, the relative GFP expression level is $147.7 \pm 6.1$. When the promoter for LacI is removed (uninhibited GFP expression), the expression level rises to $264.5 \pm$
10.9. The increase is significant (79%), and can easily be seen when comparing the number density functions (figure 4.10a).

Similarly, testing the interaction between the lac repressor and $\lambda P_R$ with pSCP146 and pSCP156 shows a difference between the GFP expression levels (figure 4.10b). The expression level of GFP increases from $111.4 \pm 7.9$ to $235.3 \pm 32.3$ when the promoter in front of LacI was removed, an increase of 111%. Therefore, the lac...
repressor affects the expression of its two nonspecific promoters to an extent that we can experimentally verify.

4.10.2 λ repressor interactions (III and IV)

The interaction between the λ repressor and P_LlacO-1 is tested with pSCP113 and pSCP123, with and without the promoter for λ CI, respectively. The results, shown with comparative number density functions in figure 4.10c, demonstrate a jump in GFP expression once the promoter is removed, from an average expression level of 279.2 ± 10.3 to 874.9 ± 60.9, an increase of over 200%. Moreover, there is a distortion in the shape of the number density function: with the λ repressor being expressed, the distribution of fluorescence is much wider than it is when the repressor is not expressed. This experiment provides evidence that a relatively strong interaction exists between the λ repressor and P_LlacO-1, stronger than either of the lac repressor interactions.

pSCP114 was created to test the interaction between the λ repressor and P_LtetO-1, along with the control plasmid without the promoter in front of λ CI, called pSCP124. These plasmids are similarly tested, and again a difference in GFP expression is found – rising from 396.9 ± 3.0 with the λ CI promoter to an average without the promoter of 666.5 ± 21.6, an increase of 67%. Number density functions are shown in figure 4.10d. Clearly, an interaction between the λ repressor and P_LtetO-1 also exists, though not as strong as that with P_LlacO-1.

The existence of the λ repressor-P_LlacO-1 and λ repressor-P_LtetO-1 interactions may not be that surprising since the λ repressor binds to the wild type P_L promoter, upon which the two promoters are based. Even though the λ operator sites are removed, since
the $\lambda$ repressor binds over a large region of DNA, there must be some affinity for what remains of the $P_L$ promoter in the hybrid promoters.

4.10.3 *tet repressor interactions (I and VI).*

Plasmids pSCP141 and pSCP151 were created to test the interaction between the *tet* repressor and $\lambda P_R$. pSCP151, not producing *tet* repressors, grows at a normal rate; however the cells containing pSCP141, expressing TetR highly, experience stunted growth both overnight and after dilution into fresh media. Not knowing whether the growth problem is due to *tet* repressors or some other mutation, cultures containing pSCP141 are grown overnight with varying amounts of aTc. It is determined that higher amounts of aTc removed the inhibition of the cells’ growth to an extent, indicating that high expression of TetR was likely the culprit for pSCP141’s reduced growth. A similar problem was reported in 1984, when it was found that high expression of TetR from a plasmid in bacteria led to growth arrest (Oehmichen et al). The growth problem prevented us from testing this interaction.

Knowing that high expression of tetR inhibits growth, the *tet* repressor interaction with its own promoter, $P_LlacO-1$ (hypothesis VI), was not tested. Given that all other tested interactions showed at least a small difference between expression levels, even though *tet* repressor interactions could not be found, it would not be surprising if the interactions did exist, at least to the extent that the *lac* repressor’s interactions do.

Results for the interaction hypothesis experiments are summarized in table 4.2.

According to our modeling studies, the $\lambda$ repressor-$P_LlacO-1$ interaction helps to explain the existence of the high state and the multiplicity between middle and high states. The interaction between the $\lambda$ repressor and the $P_LtetO-1$ promoter has only a
### Table 4.2. Results summary of the interaction hypothesis experiments.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Interaction</th>
<th>Strength of Interaction</th>
<th>Increase in Average Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>TetR - $\lambda$ P_R</td>
<td>not tested, inhibited growth</td>
<td>n/a</td>
</tr>
<tr>
<td>II</td>
<td>LacI - $P_I$ tetO-1</td>
<td>weak</td>
<td>79%</td>
</tr>
<tr>
<td>III</td>
<td>$\lambda$ cI - $P_I$ lacO-1</td>
<td>relatively strong</td>
<td>213%</td>
</tr>
<tr>
<td>IV</td>
<td>$\lambda$ cI - $P_I$ tetO-1</td>
<td>weak</td>
<td>68%</td>
</tr>
<tr>
<td>V</td>
<td>LacI - $\lambda$ P_R</td>
<td>weak</td>
<td>111%</td>
</tr>
<tr>
<td>VI</td>
<td>TetR - $P_I$ lacO-1</td>
<td>not tested</td>
<td>n/a</td>
</tr>
</tbody>
</table>

marginal effect on the stationary behavior of the network and solely brings the high state expression level closer to that of the middle state, better resembling our experimental results than if only the self-promoting interaction with $P_I$ lacO-1 was found. Incorporating the lac repressor interactions into the model can add the lower steady state and multiplicity, though only for a narrow range of parameters. However, if the suspected tet repressor interactions exist, particularly that with $\lambda$ P_R, the range of parameters is significantly extended for which we can match modeling to experimental behavior. Even if the tet repressor interaction does not exist, and the region of [aTc] where the low state is found would be quite small, cell population heterogeneity could expand the region to the range seen experimentally.

#### 4.11 Using aTc and IPTG as co-inducers allows finer control over distribution characteristics

Since IPTG is also an extracellular inducer of the network, its effect on the network was tested in conjunction with aTc. IPTG was not used alone because GFP expression without any inducers is at a low level, and IPTG would decrease expression further: referring to figure 4.1, IPTG depletes the number of lac repressors bound to lacO sites, which increases the expression of tet repressors, which in turn suppresses GFP
expression. Testing IPTG with aTc allowed its effect on the network to be seen through comparison with aTc-only experiments.

Two concentrations of aTc were used: 500 nM (at middle state) and 1000 nM (high state), and [IPTG] was varied from 0 M to 600 μM. We also explored the two scenarios of exposing the system to both inducers concurrently, or introducing IPTG after cells were already exposed to aTc (1000 nM only).

First we examine the results of introducing IPTG with aTc from the beginning of the experiment. At the reference point, we see how IPTG influences the average fluorescence as well as the coefficient of variation (CV). Figure 4.11a shows a linear decrease in fluorescence as [IPTG] increases. The CV has a similar trend, shown in figure 4.11b.

At 1000 nM aTc, one can see that the CV decreases between 0 and 50 μM IPTG, even though the average does not change. IPTG first affects the shape of the distribution before it affects the average. In the opposite way, at high IPTG, even though the average decreases between 300 and 600 μM IPTG, the CV plateaus. This implies that there is a limit to how compact the distribution can become.

To visualize the changes in the distributions with the addition of IPTG, the number density functions at the reference state are shown in figure 4.12. The distributions with IPTG and 500 nM aTc are in figure 4.12a and those with IPTG and 1000 nM aTc are in figure 4.12b. In figure 4.12a, as [IPTG] increases, the fluorescence of the highest-fluorescing cells decreases, resulting in a lower average and a sharper distribution. In figure 4.12b, the whole distribution shifts to lower fluorescence and the number density functions become sharper as [IPTG] increases, but there is no change in
shape between 300 and 600 μM IPTG ndfs.

Although IPTG affects the fluorescence value and the shape of the distributions at the reference state, it was found to have no effect on the network’s dynamic behavior when introduced at the same time as aTc. The oscillatory middle state expression at 500 nM aTc in figure 4.13a and the stable high state expression at 1000 nM aTc in figure 4.13b is merely attenuated in the average fluorescence value and/or amplitude of oscillation.
Experiments performed by adding IPTG after the cultures were already exposed to aTc yielded different results. Average fluorescence versus time is shown in figure 4.14a and the corresponding number density functions at the reference state is shown in figure 4.14b. The dynamics of the culture at 50 and 100 μM IPTG were similar to that of the previous set of experiments, reaching slightly lower average values as [IPTG] increases, but qualitative behavior of steady expression is unchanged. However, the dynamics at 200 to 600 μM IPTG were slower than in the previous experiments, and they asymptotically reach different average values (displaying the same trend of reaching lower values at higher [IPTG]). The slower dynamics were probably due to two things: the fact that GFP degrades at a slower rate than it becomes fluorescent, and slower cell growth. The inducer-initiated shock to the cells was most acutely felt after adding 600 μM IPTG, at which point cell growth was delayed for more than 4 hours. This growth delay stalled the dynamics because the loss of GFP was not accelerated by division (which would drop the amount of GFP per cell roughly in half).

The reference state number density functions explain even more. The ndfs at 50 and 100 μM IPTG are not very different than those in figure 4.12b. However, the ndfs at 200, 300, and 600 μM IPTG resemble the distributions of the middle state, closer to those in figure 4.5 or 4.12a. Those distributions are not sharper, but broader than those at lower [IPTG]. The trend of the distributions sharpening as [IPTG] increases is true, but only within the 200-600 μM IPTG group. Once the [IPTG] was in this range, it constituted enough of a perturbation to the network to change its behavior from the high state to the middle state. The reason that this was not seen in the experiments introducing IPTG and aTc at the same time is because the perturbation made by IPTG was made concurrently.
with aTc, so the network could adjust to both at the same time.

So, IPTG can help to control characteristics of average fluorescence and CV when used with aTc, but the order of introduction makes a difference in whether the network state is affected. When added together with aTc, IPTG attenuates the average, CV, and oscillation amplitude. When added after aTc is introduced, a higher concentration of
IPTG can also perturb the system to a lower state

4.12 Summary and conclusions

We investigated the interaction between cell population heterogeneity and the repressilator architecture with the aid of flow cytometry to discover the distribution of phenotypes. We explored the ability of the repressilator system, a cyclic negative feedback loop of three promoter-repressor pairs, to be controlled by the extracellular inducers aTc and IPTG. The original two-plasmid system designed by Elowitz and Leibler was combined into a one-plasmid system, purposely raising the intracellular concentration of the repressors. With this new plasmid, four sets of shake flask experiments were performed: 1) starting at an uninduced state, growing cells in media with aTc only; 2) starting at an aTc induced state, growing cells in media with a lower aTc concentration; 3) starting at an uninduced state, growing cells in media containing either both aTc and IPTG; and 4) starting at an aTc induced state, adding IPTG to the media while keeping the aTc concentration constant.

In the aTc-only experiments, it was discovered that, at the reference state, the one-plasmid system exhibited bi-threshold behavior (three steady states), with significant hysteresis regions. The number density functions demonstrated that the middle state is not a bimodal combination of the low and high states, but a new state. The average fluorescence values and CV's are extremely consistent within each state. The middle state is unsteady (at the single-cell level) while the low and high states are steady. The ability of the cultures in the middle state to oscillate in GFP fluorescence may explain the larger CV value of the middle state.

The bi-threshold behavior and multiplicity of the system observed experimentally
are not features that are possible with the original model accompanying the repressilator system, in which each repressor only interacts with the promoter containing its respective operator sites (Elowitz and Leibler, 2000). In order to obtain these features with a model, we must include extra interactions, namely interactions between a repressor and the nonspecific promoters known to interact with the other two repressors. Two nonspecific interactions between a repressor and the nonspecific promoter of the upstream repressor are needed to create the high and low states, as well as the multiplicity observed. Experimentally, four significant nonspecific interactions (between the lac and λ repressors and their nonspecific promoters) are observed, which according to our model can create the high state and perhaps the low state. We suspect that significant interactions also exist between the tet repressor and its nonspecific promoters (further solidifying the low state), but could not verify them due to toxicity of high tet repressor concentration. Cell population heterogeneity could expand the low and high state regions to match the experimental results. Finding the extra interactions in this network is quite significant, for it shows that even small nonspecific interactions can have a major effect on network behavior. Moreover, it lends credibility to our results, since it helps to explain why the network did not behave as expected from its design.

The aTc-IPTG mixed inducer experiments demonstrated that when IPTG is added concurrently with aTc, up to 600 μM, it fine tunes the qualities of the GFP distribution in comparison to growing cultures in aTc only. IPTG lowers the average and CV, though there is a limit to how compact the distribution can get. However, this fine tuning does not affect the corresponding behavior that the culture has with aTc alone – in the middle state, IPTG attenuates the average and amplitude of oscillation, and in the high state,
IPTG simply lowers the average. If IPTG is added after the culture is already stable in the high state, then IPTG can perturb the system enough for the network to reach the middle state if added in a high enough concentration.

There is a large amount of flexibility in the behavior of this genetic architecture, and it can be controlled extracellularly. The fine tuning that can be achieved when IPTG and aTc are used together shows that it is possible to reach two distributions with the same average fluorescence value but different CV values, and this could have use in biotechnology. We also observed the importance of the order in which the culture is exposed to various levels of aTc or IPTG in impacting the culture's asymptotic behavior, since hysteresis was observed after starting the cultures at an aTc-induced high state. Looking at the distributions of fluorescence as well as the averages gave a more complete view of the nature of the three states that are reached with this system, giving us a better understanding of how this genetic architecture functions under different inducer conditions. Moving beyond the singular oscillatory behavior previously observed, we show that this system can behave in a much richer manner than previously discovered.
Chapter 5

Summary, Conclusions, and Future Directions

5.1 Summary and Conclusions

Through several series of shake flask experiments, two artificial genetic regulatory networks were carefully studied with respect to transient gene expression and reference state expression, as a function of extracellular inducer concentration. We found that extracellular inducers were effective at controlling the average and distribution of gene expression in a cell population, as measured by flow cytometry with a GFP marker. Studying both transient and reference state behavior over different conditions allowed us to more fully understand how each regulatory network operates in the cell population as a whole.

We observed that the degradation rate of the gene marker used can affect perception of network behavior with the gene-switching network. With insights from modeling work [75], we understand more of how the fast-switching action of the gene-switching network contributes to the bimodal distributions we observe experimentally. In our experiments, we also discovered the full range of oscillatory network behavior, which has three steady states (two stable, one unstable at the single-cell level) with characteristic GFP distributions that can be modified according to inducer concentrations. This behavior was unexpected based on modeling of the designed network connectivity. So, we hypothesized via modeling and proved experimentally that significant nonspecific interactions between repressors and promoters within the network exist and are plausibly leading to the unexpected behavior. The discovery of these extra interactions was
completely unanticipated, and it highlights the importance of understanding all interactions of the components of a system when designing an artificial gene network.

We learned that genetic architecture plays a large role in phenotypic variability within a population. We found that heterogeneity expanded in the gene-switching network when only some of the population became induced, resulting in a bimodal distribution, and when individual cells displayed oscillatory behavior in the oscillatory system. Heterogeneity contracted when we could impose more constraints on the oscillatory system by using multiple inducers. What we noticed in both systems is that heterogeneity can only be limited up to a point – there appears to be a minimum variation in phenotype that the systems cannot go beyond, regardless of increasing inducer concentration. Each set of distributions that we found for both systems were unique and externally controllable, giving hope that they may be used in some biotechnological applications one day.

All of this work highlighted how valuable flow cytometry is in assessing the expression of these networks. Without the distributions from flow cytometry, we would not have differentiated so well between conditions and therefore would not have gained the insights into network behavior that we did.

5.2 Future Directions

5.2.1 Chemostat studies

In any future use of these regulatory networks for biotechnological purposes (for instance, to create biodegradable polymers), the cells carrying the networks will be most likely grown in bioreactors. To keep cells in log phase, continuous culture would likely
be used.

With the gene-switching network, it would be beneficial to see how stable the distributions (particularly the bimodal distributions) found in shake flask experiments are when cultures are grown for extended generations. Also, based on the short chemostat study for the oscillatory system in Appendix C, settings such as growth rate would have to be controlled. To produce large amounts of a product with this system, it would be necessary to grow cultures to a high OD$_{600}$. We would have to assess what effect this change (from very low OD$_{600}$ in shake flask experiments) would have on network behavior.

The same is true of the oscillatory system. More extensive chemostat studies are needed to clarify the effect of aTc concentration (if it produces 3 states as in the shake flask experiments), dilution rate, and other operating conditions on the GFP distributions.

5.2.2 Two-color studies

Realizing that using one GFP marker can only relay part of the story when it comes to system behavior (that of $\lambda$ cI expression), work should be done to elucidate the expression levels of the other proteins to complete the picture and confirm our assumptions about their expression.

The ideal scenario to carry out these studies is to find more fluorescent protein markers that have similar degradation and monomer characteristics to our GFP but have a shifted fluorescence emission spectrum so that we can differentiate between them in the flow cytometer. Assuming a suitable protein is found, it should be inserted after the lacI or tetR sequences in the two plasmids carrying the networks. One caveat is that the addition of another gene will make the plasmids significantly larger, and may affect the
cellular growth rate and consequently the networks’ expression. The results may not be
directly comparable to those with GFP only.

Another scenario is to simply move the GFP from behind the λ cI gene to be
expressed with the lacI or tetR genes instead. In this way, if no 2nd fluorescent protein is
found, we can still discover the comparative expression of the other genes involved in the
networks.

5.2.3 Fluorescence microscopy studies

One assumption that is made of the middle state in the oscillatory network studies
is that cells are individually oscillating, though not in concert. It is important to confirm
this with fluorescence microscopy studies, in which we can track changes in fluorescence
level of individual cells on a slide over time, even through divisions. An effort will have
to be made to mimic shake flask conditions, to make the results as comparable as possible
(in part due to the autofluorescence levels that may exist with rich media and aTc, which
should be avoided).

5.2.4 Genetic manipulations

Another way to exert control over network behavior is to change the
characteristics of the individual components of the network or of the plasmid itself at the
genetic level.

An interesting study would be to perform the shake flask experiments with the
oscillatory system again, varying inducer concentration, but after having changed the
network copy number. Some brief studies done with the original repressilator system
with the regulatory elements carried on a low copy plasmid showed sharper distributions
and even some transient bimodality compared to the single-plasmid system with all genes
carried on a 50-70 plasmid. It would be interesting to change the genes from pSCP102 to a lower copy plasmid (say the 30 copy plasmid carrying the gene-switching network), and see the effects on the distributions of fluorescence.

Another manipulation we can make is to add a specific degradation tag to a protein, elongating or shortening its half-life. We can do this with both the gene-switching and oscillatory networks, to see the effect of changing one or more repressor degradation rates on the network behavior and distribution characteristics.

We can also change the promoter strength of any promoter in either network. If specific mutagenic sites to affect a given promoter’s strength are not known, preliminary studies can be done with random mutagenesis to find suitable sites. Similar to the study with changing degradation rates of repressors, we can change one or more promoter strengths to discover its effect on network behavior and distribution characteristics.

These genetic manipulations could lead to more desirable distribution characteristics for a particular biotech application.

5.2.5 Electromobility Shift Assay (EMSA)

The extra interaction experiments performed for the oscillatory network are incomplete – the tet repressor interactions are unknown because they were untested. Also, in order to model the extra interactions more quantitatively, we need to know the magnitudes of the known interactions (lac repressor-P_LlacO-1, tet repressor-P_LtetO-1, λ repressor-λ P_R) as well. This information can be determined by performing an EMSA. In this procedure, the promoter DNA sequences are mixed with repressor proteins, and run on a gel against promoter DNA without proteins (if repressors are bound, the DNA moves more slowly down the gel). The whole matrix of nine interactions, and their
relative strengths, could be discerned. With this information, more complete modeling work can be done that may conclusively confirm our hypothesis that extra interactions create the behavior observed experimentally.
Appendix A

Plasmid Constructions and Maps

Included in this appendix are outlines of important plasmid information (genes, antibiotic resistance markers, size, glycerol stock information, etc.). Plasmid maps for most oscillatory network (repressilator) plasmids are included as well.\(^1\)

**Gene-Switching Network Plasmids**

<table>
<thead>
<tr>
<th>plasmid</th>
<th>host</th>
<th>size</th>
<th>resistance(^{host/plasmid})</th>
<th>glycerol stock</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTAK117</td>
<td>JM2.300</td>
<td>6.1 kb</td>
<td>Str(^R)/Ap(^R)</td>
<td>GNB10741..47 SCP1-8-1..-6</td>
<td>[37]</td>
</tr>
<tr>
<td>pTAK117</td>
<td>MC4100</td>
<td>6.1 kb</td>
<td>Str(^R)/Ap(^R)</td>
<td>GNB10665</td>
<td>[37]</td>
</tr>
<tr>
<td>pGNB10030</td>
<td>JM2.300</td>
<td>6.1 kb</td>
<td>Str(^R)/Ap(^R)</td>
<td>GNB11037</td>
<td>[64]</td>
</tr>
<tr>
<td>pGNB10030</td>
<td>MC4100</td>
<td>6.1 kb</td>
<td>Str(^R)/Ap(^R)</td>
<td>SCP4-22-2</td>
<td>[64]</td>
</tr>
</tbody>
</table>

\(^1\) plasmid maps were constructed using the Savvy version 0.1 online program, http://www.bioinformatics.org/savvy/
pZS1-ITLrL (main plasmid for Repressilator system, gift from Elowitz [39])

Table A.2. Location of regulatory elements on plasmid pZS1-ITLrL.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLlacO-1</td>
<td>6</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>tetR-lite</td>
<td>107</td>
<td>763</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>770</td>
<td>874</td>
<td></td>
</tr>
<tr>
<td>lambda PR</td>
<td>901</td>
<td>945</td>
<td></td>
</tr>
<tr>
<td>lacI-lite</td>
<td>988</td>
<td>2112</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2136</td>
<td>2240</td>
<td></td>
</tr>
<tr>
<td>PLtetO-1</td>
<td>2263</td>
<td>2332</td>
<td></td>
</tr>
<tr>
<td>clI-lite</td>
<td>2359</td>
<td>3105</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>3136</td>
<td>3240</td>
<td></td>
</tr>
<tr>
<td>SC101</td>
<td>3273</td>
<td>5425</td>
<td>complement</td>
</tr>
<tr>
<td>ApR</td>
<td>5601</td>
<td>6461</td>
<td>complement</td>
</tr>
</tbody>
</table>

"lite" indicates reduced half-life due to addition of degradation tag [39]

Figure A.1. Plasmid map of pZS1-ITLrL.
pZ2E1-gfpaav (reporter plasmid for Repressilator system, gift from Elowitz [39])

Table A.3. Location of regulatory elements on plasmid pZ2E1-gfpaav.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLtetO-1</td>
<td>6</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>109</td>
<td>864</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>915</td>
<td>1029</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1031</td>
<td>1839</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>1976</td>
<td>2770</td>
<td>complement</td>
</tr>
</tbody>
</table>

gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.2. Plasmid map of pZ2E1-gfpaav.

Both Repressilator plasmids are found in MC4100 host in glycerol stocks GNB10852-57.
pSCP102

Table A.4. Location of regulatory elements on plasmid pSCP102.

6008 total bp
MC4100 glycerol stocks: 3-75-1, GNB11078

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLlacO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>tetR-lite</td>
<td>108</td>
<td>764</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>771</td>
<td>875</td>
<td></td>
</tr>
<tr>
<td>lambda PR</td>
<td>902</td>
<td>946</td>
<td></td>
</tr>
<tr>
<td>lacI-lite</td>
<td>989</td>
<td>2113</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2137</td>
<td>2241</td>
<td></td>
</tr>
<tr>
<td>PLtetO-1</td>
<td>2264</td>
<td>2333</td>
<td></td>
</tr>
<tr>
<td>cl-lite</td>
<td>2360</td>
<td>3106</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>3136</td>
<td>3891</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>3942</td>
<td>4056</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>4058</td>
<td>4866</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>5003</td>
<td>5797</td>
<td>complement</td>
</tr>
</tbody>
</table>

“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.3. Plasmid map of pSCP102.
Construction of pSCP102

To raise the copy number of the regulatory genes and to use a kanamycin resistance marker, the regulatory genes from the Repressilator plasmid were inserted into the Reporter plasmid. Since compatible restriction enzyme sites were not found in the locations desired, PCR amplification was used for both plasmids to add the desired restriction enzyme site at one end of the fragments to be combined. Standard methods were used for construction of the plasmid and polymerase chain reaction (Miller, 1992; Sambrook and Russell, 2001). The regulatory genes were amplified by PCR from the Repressilator plasmid, from P_lacO-1 to cI, ending before the terminator. The following primers (obtained from Integrated DNA Technologies, INC.) were used: N-terminus, 5'-CAC CTC GAG AAT TGT GAG CGG -3' (all bases match template, XhoI is in italics), C-terminus, 5'-GCA CGT ACG CTC TAG ATC AAG CTG CTA AAG CG -3' (bases in bold match template, added a BsiWl site (in italics) upstream of Ncol). In the Reporter plasmid, the entire plasmid was replicated between the beginning of the gfpaav gene and after KmR, excluding the P_tetO-1 sequence, since gfpaav would be placed directly after the cl gene. The following primers were used: N-terminus, 5'-GCA CGT ACG AGG AGA AAG GTA CCG C -3' (bases in bold match template, added a BsiWI site (in italics) before the gfpaav, keeping the gfpaav RBS sequence), C-terminus, 5'-GGA CTC GAG GTG AAG ACG -3' (all bases match template, XhoI is in italics). Platinum Pfx DNA polymerase (Invitrogen) was used for DNA amplification with a RoboCycler Gradient 96 Temperature Cycler (Stratagene). The PCR products were cleaned using Wizard PCR Preps (Promega). The products were then cut by restriction digest with XhoI and then with BsiWI. After digestion, the fragments were isolated from a 0.8 % NuSieve
GTG agarose (ISC BioExpress) tris-borate-EDTA gel using a Qiagen, Qiaex II Gel Extraction Kit. The fragments were then ligated together using T4 high concentration DNA ligase (New England Biolabs). The ligation reaction was transformed into TOP10 SuperCompetent Cells (Invitrogen) using the “One-Shot Transformation Reaction” procedure in TOPO TA Cloning Kit. Transformants were selected on (Luria-Bertani) LB agar plates containing kanamycin. After selection, plasmid was isolated from LB broth cultures and screened with \( HpaI \). The sequences around \( XhoI \) and \( BsiWI \) were confirmed by sequencing (Lone Star Labs, Inc.), and the plasmid was named pSCP102. The plasmid was then transformed into MC4100 chemically competent cells using the RbCl procedure [65].
pSCP141

Table A.5. Location of regulatory elements on plasmid pSCP141.

3888 total bp

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda PR</td>
<td>7</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>66</td>
<td>848</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>872</td>
<td>1013</td>
<td></td>
</tr>
<tr>
<td>PLlacO-1</td>
<td>1040</td>
<td>1109</td>
<td></td>
</tr>
<tr>
<td>tetR-lite</td>
<td>1141</td>
<td>1797</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1804</td>
<td>1908</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1938</td>
<td>2746</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2883</td>
<td>3677</td>
<td>complement</td>
</tr>
</tbody>
</table>

“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.4. Plasmid map of pSCP141.
Table A.6. Location of regulatory elements on plasmid pSCP151.

<table>
<thead>
<tr>
<th></th>
<th>start</th>
<th>end</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda PR</td>
<td>7</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>66</td>
<td>848</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>872</td>
<td>1013</td>
<td></td>
</tr>
<tr>
<td>*tetR-lite</td>
<td>1053</td>
<td>1676</td>
<td>complement</td>
</tr>
<tr>
<td>T1</td>
<td>1683</td>
<td>1787</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1817</td>
<td>2625</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2762</td>
<td>3556</td>
<td></td>
</tr>
</tbody>
</table>

*partial: missing N-terminal end through start codon
“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]
**pSCP142**

**Table A.7.** Location of regulatory elements on plasmid pSCP142.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLtetO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>lambda PR</td>
<td>1061</td>
<td>1105</td>
<td></td>
</tr>
<tr>
<td>lacI-lite</td>
<td>1148</td>
<td>2272</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2296</td>
<td>2400</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2426</td>
<td>3234</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>3371</td>
<td>4165</td>
<td>complement</td>
</tr>
</tbody>
</table>

“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

![Figure A.6. Plasmid map of pSCP142.](image-url)
<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rPLtetO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>*lacI-lite</td>
<td>1055</td>
<td>2126</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2150</td>
<td>2254</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2282</td>
<td>3090</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>3227</td>
<td>4021</td>
<td>complement</td>
</tr>
</tbody>
</table>

*partial: missing N-terminal end through start codon
“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]
**pSCP113**

**Table A.9.** Location of regulatory elements on plasmid pSCP113.

<table>
<thead>
<tr>
<th></th>
<th>start</th>
<th>end</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLlacO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>PLtetO-1</td>
<td>1039</td>
<td>1108</td>
<td></td>
</tr>
<tr>
<td>cl-lite</td>
<td>1135</td>
<td>1881</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1912</td>
<td>2016</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2024</td>
<td>2832</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2969</td>
<td>3763</td>
<td>complement</td>
</tr>
</tbody>
</table>

"lite" indicates reduced half-life due to addition of degradation tag [39].

gfpaav is GFPmut3 with degradation tag ending with AAV [56].

**Figure A.8.** Plasmid map of pSCP113.
**Table A.10.** Location of regulatory elements on plasmid pSCP123.

3863 total bp
MC4100 glycerol stocks: 6-32-3, GNB11157

<table>
<thead>
<tr>
<th></th>
<th>start</th>
<th>end</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLlacO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>Tl</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>*cl-lite</td>
<td>1039</td>
<td>1770</td>
<td>complement</td>
</tr>
<tr>
<td>Tl</td>
<td>1797</td>
<td>1911</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1913</td>
<td>2721</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2858</td>
<td>3652</td>
<td></td>
</tr>
</tbody>
</table>

*partial: missing N-terminal end through start codon
“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

**Figure A.9.** Plasmid map of pSCP123.
pSCP114

Table A.11. Location of regulatory elements on plasmid pSCP114.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLtetO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>PLtetO-1</td>
<td>1039</td>
<td>1108</td>
<td></td>
</tr>
<tr>
<td>cl-lite</td>
<td>1135</td>
<td>1881</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1912</td>
<td>2016</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2024</td>
<td>2832</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2969</td>
<td>3763</td>
<td>complement</td>
</tr>
</tbody>
</table>

"lite" indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.10. Plasmid map of pSCP114.
Table A.12. Location of regulatory elements on plasmid pSCP124.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLtetO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1030</td>
<td></td>
</tr>
<tr>
<td>*cl-lite</td>
<td>1039</td>
<td>1770</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1797</td>
<td>1911</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1914</td>
<td>2722</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>2859</td>
<td>3653</td>
<td>complement</td>
</tr>
</tbody>
</table>

*p: partial, missing N-terminal end through start codon

"lite" indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.11. Plasmid map of pSCP124.
**Table A.13.** Location of regulatory elements on plasmid pSCP146.

4359 total bp  
MC4100 glycerol stocks: 6-45-2

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda PR</td>
<td>7</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>gfpav</td>
<td>66</td>
<td>848</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>872</td>
<td>1013</td>
<td></td>
</tr>
<tr>
<td>lambda PR</td>
<td>1044</td>
<td>1088</td>
<td></td>
</tr>
<tr>
<td>lacI-lite</td>
<td>1131</td>
<td>2255</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>2279</td>
<td>2383</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2409</td>
<td>3217</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>3354</td>
<td>4148</td>
<td>complement</td>
</tr>
</tbody>
</table>

“lite” indicates reduced half-life due to addition of degradation tag [39]  
gfpav is GFPmut3 with degradation tag ending with AAV [56]

**Figure A.12.** Plasmid map of pSCP146.
**Table A.14.** Location of regulatory elements on plasmid pSCP156.

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda PR</td>
<td>7</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>66</td>
<td>848</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>872</td>
<td>1013</td>
<td></td>
</tr>
<tr>
<td>*lacI-lite</td>
<td>1038</td>
<td>2109</td>
<td>complement</td>
</tr>
<tr>
<td>T1</td>
<td>2133</td>
<td>2237</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>2263</td>
<td>3071</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>3208</td>
<td>4002</td>
<td></td>
</tr>
</tbody>
</table>

*partial: missing N-terminal end through start codon
“lite” indicates reduced half-life due to addition of degradation tag [39]
gfpaav is GFPmut3 with degradation tag ending with AAV [56]

**Figure A.13.** Plasmid map of pSCP156.
Table A.15. Location of regulatory elements on plasmid pSCP110.

2982 total bp
MC4100 glycerol stocks: 5-74-1, GNB11122

<table>
<thead>
<tr>
<th>Element</th>
<th>Start</th>
<th>End</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLlacO-1</td>
<td>7</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>gfpaav</td>
<td>110</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>916</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>1032</td>
<td>3184</td>
<td></td>
</tr>
<tr>
<td>KmR</td>
<td>1977</td>
<td>2771</td>
<td>complement</td>
</tr>
</tbody>
</table>

gfpaav is GFPmut3 with degradation tag ending with AAV [56]

Figure A.14. Plasmid map of pSCP110.
Appendix B

Additional Gene-Switching Network Experiment

Questions that remained after studying the gene-switching system had to do with the stability of the two states.

1. The authors of the toggle paper [37] observed that cultures that became induced into the upper steady state stayed induced after the inducer was removed. However, they used a GFP with a longer than 24 hour half-life, and we wondered if the high stability of the protein created an artifact in the results.

2. After inducing a culture with IPTG, the authors always grew the cultures for a period of time without inducer before raising the temperature to 42°C in order to switch the cells back to the low state. Is that intermediate period of time without inducer necessary to dilute out the amount of IPTG in the cells so that they can switch to domination of free lac repressors without delay?

The answer to both of these questions was found by performing another experiment with the pGNB10030 plasmid, with gfpaav (half-life of 60 min) substituted for GFPmut3 in pTAK117. Instead of using JM2.300, which had been established as a weak strain, we used the healthier MC4100. After transforming the plasmid into the new host, a similar procedure to that used in the oscillatory system experiments (chapter 4), that of continual subculturing, was used in order to keep the cultures in the same growth state while subjecting them to changing conditions.

The cells were grown from the glycerol stock overnight in an orbital shaker for 12
hours in 5 mL LSRB, Ap 100 mg/L, at 32°C. Then they were subcultured to an OD_{600} of 0.0001 and grown again without inducer to get back into exponential phase. Once the culture reached an OD_{600} of 0.01, they were then continuously diluted by 10 in fresh media once they grew back to 0.01. The conditions for these dilution steps are as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 mM</td>
<td>2 mM</td>
<td>2 mM</td>
<td>2 mM</td>
<td>no IPTG</td>
<td>42°C</td>
<td>no IPTG</td>
</tr>
<tr>
<td></td>
<td>no IPTG</td>
<td>no IPTG</td>
<td>no IPTG</td>
<td>no IPTG</td>
<td>42°C</td>
<td>42°C</td>
<td></td>
</tr>
</tbody>
</table>

After the third dilution in 2 mM IPTG, the culture was split into 2 lines of flasks, one continuing with 2 mM IPTG, and one without IPTG (cells were washed first). After two subculturing steps, both cultures were subjected to the same conditions of no IPTG, growing at 42°C for two dilution steps.

The results of the experiment are shown in figure B.1. One can see a sharp rise to high fluorescence when the culture is initially exposed to high IPTG. The high expression level is stable whether, in the 4th and 5th dilution steps, the culture is exposed to IPTG or not. Therefore, the results that Gardner et al. found were true, that the system is stable in the high steady state even without constant inducer. In the last two dilution steps, both cultures experience the same sharp drop in fluorescence as the temperature is raised and \( \lambda \) repressors become unstable. This shows that an intermediate step without IPTG is not necessary to allow the system to switch back to the low state once the temperature is raised.

This experiment also shows that MC4100 is a suitable host for this gene-switching system, one that does not require the use of PI to distinguish live cells from
dead. In addition, this experiment independently demonstrated the usefulness of this procedure, since the single-flask procedure used in the previous experiments with this system would not have allowed the cells to be kept in exponential phase for such an extended period of time.

Figure B.1. Average fluorescence and average forward scatter values over time. Filled in circles and squares are forward scatter values. Open circles and squares are the average fluorescence values. The open triangles along the x-axis mark the times at which a dilution to fresh media was made. The circles follow the culture that was not exposed to IPTG at the time indicated by the arrow near 250 minutes.
Appendix C

Expression of Oscillatory Network in Continuous Culture

After extensive testing with the oscillatory system in shake flask cultures (see chapter 4), we set out to find how the expression in this batch mode compared to expression in controlled continuous culture conditions.

Three inducer conditions were tested: 1 nM, 200 nM, and 1000 nM aTc. Each produced green fluorescence distributions consistently in the low, middle, and high states, respectively, in the shake flask experiments. We wanted to see whether these states would be seen in continuous culture, and whether the distributions would be the same. What we discovered was that the trend of increasing fluorescence with increasing inducer concentration was the same as with the shake flask experiments. Also, the three fluorescence distributions were different from each other, but the differences were muted in comparison to those with the shake flask experiments, and the distributions did not resemble those from the shake flask experiments.

The major procedural differences between the modes, as operated:

<table>
<thead>
<tr>
<th>Shake Flask</th>
<th>Continuous Culture in Bioreactor</th>
</tr>
</thead>
<tbody>
<tr>
<td>growth at maximum µ</td>
<td>growth determined by dilution rate</td>
</tr>
<tr>
<td>pH not controlled</td>
<td>pH controlled</td>
</tr>
<tr>
<td>relatively constant, controlled, low OD&lt;sub&gt;600&lt;/sub&gt; (0.001-0.1)</td>
<td>constant, uncontrolled OD&lt;sub&gt;600&lt;/sub&gt; (near 5)</td>
</tr>
<tr>
<td>inducer kept at 4°C until use (cells diluted to fresh media every 2-3.5 hours)</td>
<td>inducer added to media tank to keep concentration constant, but tank not kept refrigerated</td>
</tr>
<tr>
<td>flasks shaken at 250 rpm</td>
<td>agitation speed set at 800 rpm</td>
</tr>
</tbody>
</table>
Procedure

A 1 L bioreactor (New Brunswick Scientific, Bioflo 110) was used, and the media volume was maintained close to 600 mL for both batch and chemostat stages. First, cells from a glycerol stock (MC4100(pSCP102)) were inoculated in 5 mL LSRB with 50 mg/L Km and grown at 32°C for 10-12 hours. 2.5-5 mL of the overnight culture was concentrated into 1 mL media and used to inoculate the bioreactor. The batch stage lasted 4 hours, and continuous flow stage lasted >20 hours. Illustrated in figure C.1 is a schematic of the continuous culture bioreactor. pH, agitation speed, air flow, media flow rate, and temperature were carefully controlled. \( \Delta \text{O}_2 \) was monitored but not actively controlled. aTc was added in the appropriate concentration to the media tank. aTc was introduced directly to the bioreactor as well (in the appropriate concentration) just before starting the pumps to keep the concentration in the bioreactor constant for the entire continuous stage. Both the bioreactor and the media tank were wrapped in foil to prevent GFP, aTc, and Km degradation.

Generally, the reactor is assumed to be at steady state after 5-8 residence times in continuous flow; we waited a minimum of 12 residence times (which corresponds to 18 generations, assuming a 70 min doubling time, which is close to the number of generations required for the middle steady state to stabilize in shake flask experiments). Samples were taken at a 70 minute interval after 20 hours of continuous flow. Samples were processed for flow cytometry in the same manner as in chapter 4, and the flow cytometer (FACScalibur) settings were the same as well. Different FSC and SSC gates were applied (linear values of 20 and 30, respectively) because the average values for both of these parameters were smaller.
The operating conditions were set as follows:

- 2.0 L/min air flow, 800 rpm agitation speed – to keep $dO_2 > 50$
- 32°C
- 0.6 hr$^{-1}$ dilution rate – corresponds to a 69.3 min doubling time
- pH set to 7; controlled by additions of HNO$_3$ and NaOH
- media – LSRB with 50 mg/L Km, appropriate amount of aTc, 30 µL/L antifoam

![Figure C.1. Schematic of chemostat.](image-url)
Results

Here are comments regarding each run that were of note:

Condition 1: 1 nM aTc (corresponds to low state in shake flask experiments)
- pH of prepared media was set too high (7.2) to maintain a pH of 7.0 during the entire run with the amount of acid prepared - ended up turning off pH control after 24 hours - climbed up to near 8.0
- dilution rate - in actuality closer to 0.59 hr⁻¹
- corrected pressure buildup in reactor after 24 hours - media outflow was not fast enough for inflow, and combined with the fact that the air outflow filter was wet due to a high media level in the reactor, the dO₂ level oscillated in the first 20 hours after continuous mode was started

Condition 2: 200 nM aTc (corresponds to middle state in shake flask experiments)
- dilution rate - in actuality closer to 0.65 hr⁻¹
- pH set to 7.19 (corresponds to pH of 7.05 at 24°C)
- fixed another pressure buildup in reactor after 24 hours

Condition 3: 1000 nM aTc (corresponds to high state in shake flask experiments)
- dilution rate - in actuality closer to 0.595 hr⁻¹
- fixed minor pressure buildup after 7 hours
- took 3 extra samples after inducer was added to get some transient data

Even though there were corrections made during each run, and in the 1 nM run there were more problems, by the time the last samples were taken, the bioreactor parameters were relatively stable.
In figure C.2 is plotted the pH and dissolved oxygen levels for each run over time. By the end of the runs, the pH and dO₂ were stabilized. The dissolved oxygen level stabilized near 80%. In the 1 nM aTc run, the oscillation in dO₂ levels is quite pronounced. The culture did not get to a steady state until the pressure problem was fixed and the pH control was turned off (at 24 hours). pH and dO₂ levels were much more consistent for the 200 nM and 1000 nM runs.

![Figure C.2. pH and dissolved oxygen readings over time.](image)

In figure C.3 is plotted the average fluorescence and light scattering values for each run over time. The 1000 nM curves start much earlier than the 1 nM or 200 nM curves because early samples were taken to get some transient data. What was discovered for the samples taken at earlier time points is that the cultures were highly induced initially, but the level of fluorescence quickly fell and did not recover. In the early samples, the light scattering values were higher than the later samples, reaching values more similar to those in the shake flask experiments.
The ending forward scatter values for each run were quite similar, but they were lower than expected based on the shake flask results (which were close to 600). It seems that the restraint of growth rate based on the dilution rate affected the cellular morphology strongly, with cells at FSC and SSC values more closely associated with late log or stationary phase in the shake flask experiments.

It was also noted that the average fluorescence increased as the level of inducer increased, a trend that was expected. However, the fluorescence levels were much lower than those seen in the shake flask experiments (refer to figure 4.4). Also, it does not appear that the values stabilize as well as in the shake flask experiments (refer to the transient plots in figure 4.6). Nevertheless, we regard these values as stabilized and take the average values and distributions as characteristic for these conditions.

Looking at the distributions of green fluorescence in figure C.4, we see that the distributions are not very distinctive from each other. The distributions start at a similar
point near 1, and extend to higher fluorescence as a function of their inducer concentration. The distributions flatten as the culture extends to higher fluorescence levels. Compared to the distributions for the shake flask experiments (figure 4.5), the differences between the states (if we can consider them different states) are much less pronounced, but the fact that there are differences proves that different aTc concentrations are felt differently. It was also noted that the range of fluorescence at the upper end of each distribution is similar to the values reached in the shake flask experiments, which indicates that a very small percentage of the cells manage to stay induced.

![Green fluorescence distributions at steady state for each condition.](image)

**Figure C.4.** Green fluorescence distributions at steady state for each condition.

To summarize, the chemostat experiments with the oscillatory system yielded the same trends as the shake flask experiments, but the fluorescence and light scattering values were much lower. I believe that the lower values resulted from the restricting control of cellular growth by dilution rate, which prevented cells from reaching the maximum growth rate and changed their morphology. Perhaps if the dilution rate was
increased (carefully to avoid washout of cells), we might see distributions more similar in shape and value to those in the shake flask experiments. We also do not know whether these distributions are characteristic for low, middle, and high states, or whether they are simply characteristic of the specific inducer concentrations and there is a continuum of distributions and averages for different aTc concentrations. More study should be done to answer these questions and get a more complete picture of how the oscillatory system acts in chemostat conditions.
Appendix D

Equations used in Oscillatory System Modeling

The following was written by Sergio Iadevaia as an explanation of the equations involved in modeling the oscillatory system dynamics with the nonspecific interactions\(^1\). Following that is the parameter table used in creating figure 4.9 in chapter 4.

In order to model the dynamics of the repressilator in the presence of nonspecific repressor-promoter interactions, we modify the mathematical model derived by Elowitz and Leibler [39] to explicitly account for a) the ability of each protein to nonspecifically complex with the promoter of the upstream repressor as well as its own promoter (the promoters not containing its operator site), b) the effect of the aTc and IPTG inducers, and c) the dilution effect due to cellular growth [84].

Let \(m_i\) and \(p_i\) (i=\(lac(L)\), \(tet(T)\), \(\lambda(C)\)) denote the concentrations of the mRNA transcripts and repressor proteins respectively transcribed and translated from the LacI, TetR and \(\lambda\)C1 genes. Let, also, \(\alpha_{0i}\) and \(\alpha_i\) represent the synthesis rates of the mRNAs respectively transcribed in presence and absence of saturating amounts of repressors, whereas \(\varepsilon_i\) account for the specific rates (per unit transcript) of protein translation. Moreover, while \(k_{ji}\) (j=\(\lambda(C)\), \(lac(L)\), \(tet(T)\)) are the dissociation constants describing the interaction between promoters and downstream repressors, \(k_{ki}\) (k= \(tet(T)\), \(\lambda(C)\), \(lac(L)\)) are the dissociation constants accounting for the interactions between promoters and upstream repressors, whereas \(k_{ii}\) are the dissociation constants representing the

\(^1\) Part of this chapter will be submitted for publication.
interactions between the repressors and their own promoters. Furthermore, \( \eta_i \) and \( \gamma_i \) respectively denote the degradation rates of the mRNA transcripts and proteins, while \( \mu \) represents cellular growth rate. Finally, aTc and IPTG represent the concentrations of the extracellular inducers, whereas \( K_{\text{aTc}} \) and \( K_{\text{IPTG}} \) are the tet repressor-aTc and lac repressor-IPTG dissociation constants, respectively. The dynamics of the artificial repressilator can then be described by the following differential equations subject to appropriate initial conditions:

\[
\begin{align*}
\frac{d m_L}{dt} &= \alpha_{oL} + \frac{\alpha_L}{1 + p_c^2/k_C + p_T^2/k_{LT} + p_L^2/k_{LL}} - (\eta_L + \mu) m_L \\
\frac{d m_T}{dt} &= \alpha_{oT} + \frac{\alpha_T}{1 + p_L^2/k_{LT} + p_C^2/k_{CT} + p_T^2/k_{TT}} - (\eta_T + \mu) m_T \\
\frac{d m_C}{dt} &= \alpha_{oC} + \frac{\alpha_C}{1 + p_T^2/k_{TC} + p_L^2/k_{LC} + p_C^2/k_{CC}} - (\mu_C + \mu) m_C \\
\frac{dp_L}{dt} &= \varepsilon_L m_L - (\gamma_L + \mu) p_L - 2K_{\text{IPTG}}p_L^2\text{IPTG}^2 \\
\frac{dp_T}{dt} &= \varepsilon_T m_T - (\gamma_T + \mu) p_T - 2K_{\text{aTc}}p_T^2\text{aTc}^2 \\
\frac{dp_C}{dt} &= \varepsilon_C m_C - (\gamma_C + \mu) p_C
\end{align*}
\]

Equations (D.1) state that the rates of mRNA transcript (D.1a)-(D.1c) and protein (D.1d)-(D.1f) accumulation are equal to the difference between the synthesis rates respectively due to the transcription and translation processes and the consumption rates due to degradation and the dilution effect. In equations (D.1d) and (D.1e) there are additional consumption rates, which account for the depletion of free lac and tet proteins due to the inducer-repressor binding. The functional form of the transcript synthesis rates includes the effect of all possible interactions. Moreover, the functional form of the inducer-
repressor binding rates reflects the ability of each repressor dimer to complex with two inducer molecules. The nominal values of the parameters appearing in equations (D.1) are listed in table D.1. In employing equations (D.1), we assume that the λ repressor and GFP expression level dynamics are identical.

Table D.1. Nominal model parameters for the modified repressilator dynamics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{01}$ (nM/min)</td>
<td>$1.5 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\alpha_{0T}$ (nM/min)</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\alpha_{0C}$ (nM/min)</td>
<td>$1.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\alpha_{L}$ (nM/min)</td>
<td>15</td>
</tr>
<tr>
<td>$\alpha_{T}$ (nM/min)</td>
<td>45</td>
</tr>
<tr>
<td>$\alpha_{C}$ (nM/min)</td>
<td>36</td>
</tr>
<tr>
<td>$k_{CL}$ (nM$^2$)</td>
<td>0.9</td>
</tr>
<tr>
<td>$k_{LT}$ (nM$^2$)</td>
<td>0.7</td>
</tr>
<tr>
<td>$k_{TC}$ (nM$^2$)</td>
<td>1.2</td>
</tr>
<tr>
<td>$k_{LC}$ (nM$^2$)</td>
<td>15</td>
</tr>
<tr>
<td>$k_{TL}$ (nM$^2$)</td>
<td>10</td>
</tr>
<tr>
<td>$k_{CT}$ (nM$^2$)</td>
<td>8</td>
</tr>
<tr>
<td>$k_{LL}$ (nM$^2$)</td>
<td>200</td>
</tr>
<tr>
<td>$k_{TT}$ (nM$^2$)</td>
<td>150</td>
</tr>
<tr>
<td>$k_{CC}$ (nM$^2$)</td>
<td>100</td>
</tr>
<tr>
<td>$\varepsilon_L$ (1/min)</td>
<td>20</td>
</tr>
<tr>
<td>$\varepsilon_T$ (1/min)</td>
<td>20</td>
</tr>
<tr>
<td>$\varepsilon_C$ (1/min)</td>
<td>20</td>
</tr>
<tr>
<td>$\eta_L$ (1/min)</td>
<td>$\ln(2)/2$</td>
</tr>
<tr>
<td>$\eta_T$ (1/min)</td>
<td>$\ln(2)/2$</td>
</tr>
<tr>
<td>$\eta_C$ (1/min)</td>
<td>$\ln(2)/2$</td>
</tr>
<tr>
<td>$\gamma_L$ (1/min)</td>
<td>$\ln(2)/10$</td>
</tr>
<tr>
<td>$\gamma_T$ (1/min)</td>
<td>$\ln(2)/10$</td>
</tr>
<tr>
<td>$\gamma_C$ (1/min)</td>
<td>$\ln(2)/10$</td>
</tr>
<tr>
<td>$\mu$ (1/min)</td>
<td>$\ln(2)/40$</td>
</tr>
<tr>
<td>$K_{STe}$ (1/min/ nM$^2$)</td>
<td>$6.9 \times 10^{-9}$</td>
</tr>
</tbody>
</table>
References


53. Tsien, R.Y., *The green fluorescent protein*. Annual Review of Biochemistry,


