The Texas Medical Center Library

DigitalCommons@TMC

The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)

The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

5-2012

DETERMINING THE GENOTYPE-PHENOTYPE CONNECTION IN SYNTHETIC INDUCIBLE GENE EXPRESSION SYSTEMS

Rhys M. Adams

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the Biophysics Commons, and the Systems Biology Commons

Recommended Citation

Adams, Rhys M., "DETERMINING THE GENOTYPE-PHENOTYPE CONNECTION IN SYNTHETIC INDUCIBLE GENE EXPRESSION SYSTEMS" (2012). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 273.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/273

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

DETERMINING THE GENOTYPE-PHENOTYPE CONNECTION IN SYNTHETIC INDUCIBLE GENE EXPRESSION SYSTEMS

By

Rhys Michael Adams B.A., M.S.

APPROVED:

_____
Gábor Balázsi, Ph.D., Supervisory Professor

_____
John H. Byrne, Ph.D.

_____
Oleg Igoshin, Ph.D.

_____
Krešimir Josić, Ph.D.

_____
Prahlad T. Ram, Ph.D.

APPROVED:

_____
Dean,
The University of Texas
Graduate School of Biomedical Sciences at Houston

DETERMINING THE GENOTYPE-PHENOTYPE CONNECTIONIN SYNTHETIC INDUCIBLE GENE EXPRESSION SYSTEMS

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Rhys Michael Adams, B.A., M.S.
Houston, Texas

May 2012

Dedication

To Brian and Cindy Adams.

All of yesterday fell away as we sank into conversation.

I remember feeling this each time we met.

Thank you for your support and good spirits.

Acknowledgments

I thank Gábor Balázsi, who has been a fantastic mentor.

I thank Dmitry Nevozhay and Kevin Murphy for the use of their data and expertise, which have provided a solid foundation to my ideas.

I thank the members of the Gábor Balázsi laboratory- Bhaskar Dhutta, Diogo F.T. Veiga, Jinho Lee, J. Cristian Ray, Victor Shum, Caleb Gonzalez, Lin Chen, Junchen Diao- who have provided much feedback and many interesting conversations.

DETERMINING THE GENOTYPE-PHENOTYPE CONNECTION IN SYNTHETIC INDUCIBLE GENE EXPRESSION SYSTEMS

Publication No._____

Rhys Michael Adams, B.A., M.S.

Supervisory Professor: Gábor Balázsi, Ph.D.

Introduction Gene expression is an important process whereby the genotype controls an individual cell's phenotype. However, even genetically identical cells display a variety of phenotypes, which may be attributed to differences in their environment. Yet, even after controlling for these two factors, individual phenotypes still diverge due to noisy gene expression. Synthetic gene expression systems allow investigators to isolate, control, and measure the effects of noise on cell phenotypes. I used mathematical and computational methods to design, study, and predict the behavior of synthetic gene expression systems in *S. cerevisiae*, which were affected by noise.

Methods I created probabilistic biochemical reaction models from known behaviors of the *tetR* and *rtTA* genes, gene products, and their gene architectures. I then simplified these models to account for essential behaviors of gene expression systems. Finally, I used these models to predict behaviors of modified gene expression systems, which were experimentally verified.

Results Cell growth, which is often ignored when formulating chemical kinetics models, was essential for understanding gene expression behavior. Models incorporating growth effects were used to explain unexpected reductions in gene expression noise, design a set of gene expression systems with "linear" dose-responses, and quantify the speed with which cells explored their fitness landscapes due to noisy gene expression.

Conclusions Models incorporating noisy gene expression and cell division were necessary to design, understand, and predict the behaviors of synthetic gene expression systems. The methods and models developed here will allow investigators to more efficiently design new gene expression systems, and infer gene expression properties of TetR based systems.

List of illustrations

List of Tables

Abbreviations

tetR- tetracycline Repressor

revtetR- reverse tetracycline Repressor

tTA- tetracycline-controlled trans-activator

rtTA- reverse tetracycline-controlled trans-activator

tetO2- binding site for TetR to DNA

yEGFP- yeast Enhanced Green Fluorescence Protein

zeoR- Zeocin Resistance (gene)

ATc- Anhydrotetracycline

Dox- Doxycycline

CV- Coefficient of variation

FACS- Fluorescence-activated cell sorting

Mathematical notation

X- the state of the system defined by numbers of molecules

V- The "volume" of a system. This parameter is used to relate the number of species to the concentration of species

x- the state of the system defined by concentrations of molecules (**X**/ *V*)

p- the probability distribution function

π- the stationary probability distribution function (i.e., *p* when *p* does not change over time)

P- probability. *P* can be obtained from the integral of the probability distribution function. In sections 1.2.1 and 1.3.4, *P* is used to denote the probability of being in a discrete state

N_T- the total number of cells in a population

N- the number of cells in a sub-population ($N_T P$)

I- Cellular current (number of cells crossing a boundary per unit time)

C- inducer influx, directly proportional to ATc concentration in media

ν- total transcription factor concentration

w- inactive transcription factor concentration

x- active transcription factor concentration

y- ATc concentration within the cell

z- reporter (yEGFP) concentration

t- time

Z- Zeocin concentrations

φ- a boundary separating two cell states

γ- Instantaneous fitness reduction. This is the reduction in cell fitness due to its internal chemical state.

Γ- Instantaneous fitness. This is the fitness of a cell for a specific cell state..

g- The average fitness over an interval of cell states, *Γ*/*N*, or $g(\varphi_1,\varphi_2,t)=\frac{\int_{\varphi_1}^{\varphi_2}\Gamma(x)p(x,t)dx}{\int_{\varphi_1}^{\varphi_2}p(x,t)dx}$. This rate is assumed to be equal to the rate of cell growth and is used as a dilution rate. In general I studied stationary distributions of cells, so that the average fitness becomes constant over time.

H- cell state characterized by high levels of yEGFP (i.e., cells with $F \geq \varphi$)

L- cell state characterized by low levels of yEGFP (i.e., cells with $F < \varphi$)

F- fluorescence

μ- mean

σ**-** Standard deviation

σ- Covariance matrix

η - Noise, measured as the coefficient of variation, $\eta = \sigma / \mu$

I used dot notation to denote the time derivative (e.g., $\dot{x} = \frac{\partial x}{\partial t}$), except in section 5.1.2 where the time derivative was applied to the product of two variables. Molecular species and propensity rates used in Gillespie simulations may overlap with previous mathematical notation, but should be clear from context.

Chapter 1 Introduction

1.1 Background

Understanding how an organism's genotype affects its phenotype is a fundamental goal of genetics, with broad implications for medicine (1), agriculture (2), and industry (3,4). The central dogma of biology connects gene sequence to phenotype by "gene expression" by stating that information within a gene (DNA) is copied to RNA molecules during transcription, whose information may then be translated into a sequence of amino acids that form a protein (5). RNA and protein molecules expressing the information within a gene can subsequently control the behavior of a cell. Cells may respond to new environments by changing their gene expression, and consequently their phenotype. Thus, genotype and environment can interact to create many distinct phenotypes in living organisms (6).

However, accumulating evidence has demonstrated that organisms with identical genomes, grown in identical environments, can have different levels of gene expression (7-11). Although this can be partially attributed to subtle "extrinsic" differences (e.g., cell cycle, cell age, and micro-environment), part of this phenotypic divergence is attributed to random events "intrinsic" to gene expression (12). Thus gene expression is subject to deterministic (environment and genotype) and intrinsically noisy (random) components (13-16). The deterministic component is often measured as mean gene expression while noise is commonly measured as the coefficient of variation (or standard deviation divided by the mean, $CV = \sigma/\mu$) of a cell population's gene expression (17).

Certain biological processes are only possible when noisy gene expression is taken into account. For example, bimodal gene expression can occur in monostable populations (18,19). Additionally, sub-populations with drug resistant phenotypes may allow cells to resist drug treatment and repopulate their environment without genomic mutations or drug sensing proteins (20,21).

By viewing gene expression through the lens of noise, new insights about cell phenotypes have come into focus; for example, stem cell differentiation is affected by noise (22), clonal bacterial populations employ noise-induced behaviors in times of stress (23), and critical HIV genes are activated by noise that is further amplified by a positive feedback loop (24).

If stochastic gene expression changes are slow over time compared to a cell's division time, then gene expression differences may become heritable. The time that a cell stays in a particular state before changing is called "cellular memory" and can significantly affect cell phenotypes and fitness (15,25,26). For example, a population of cells that noisily express a drug resistance gene may be able to survive prolonged exposure to this drug if their cellular memories are very high.

To study the aforementioned effects of gene expression noise, appropriate measurements of mRNA and protein levels are necessary. Western blots and real-time polymerase chain reactions (RT-PCR) are both common methods for inferring two components of gene expression (i.e., proteins and RNA, respectively) from a population of cells. However, these methods rely on lysing millions of cells, eliminating information about the variability (or noise) of gene expression between cells. In contrast, fluorescence based measurement techniques can be used to infer the gene expression noise of a cell population. These measurement techniques include the use of fluorescently labeled antibodies that target specific proteins, and the fusion of the gene encoding the green fluorescence protein (GFP) with another gene of interest (27). Fluorescence intensities may then be measured within individual cells by flow cytometry or by microscopy. However, even with measurements of gene expression noise, it is difficult to determine the extent to which noise is intrinsic to a gene, or whether it is responding *deterministically* to extrinsic factors such as transcriptional regulators, gene duplication, or environmental effects (12,28).

In order to control and measure the noise-dependent effects for study, synthetic gene expression systems are often used. Synthetic biology studies the design of biological parts, devices, and systems, for useful purposes (29). Often these parts are genes and genetic regulatory sequences (such as DNA recognition sequences for transcriptional regulators) which are assembled to into gene circuits. In order to minimize the number of interactions with host cells, the genes are often non-native to their host organisms. By isolating gene expression control from the host cell's native genes, it is possible to accurately study the underlying mechanisms controlling noise.

The property of a sub-system being isolated from other parts of the host organism is called modularity, and is an important aspect of synthetic biology (30). For example, a gene circuit that is only affected by a single inducer (e.g., molecule, protein or light) and has no effect on other parts of the cell would be considered a module. This property means that synthetic gene circuit "modules" will retain their intrinsic behaviors, independent of organism or the presence of other synthetic gene circuits (31). This property is used to connect multiple synthetic gene circuits to obtain more complicated behaviors than would be possible from any single component. However, in order to rationally design complex gene circuits, the modular components must be well understood.

Two particularly important families of genes that can serve as components of gene circuit models are those encoding for the green fluorescence protein (*GFP*) (32) and the Tetracycline Repressor (*tetR*) (33). Members of the *GFP* family are important "output" genes, since they emit specific wavelengths of light when excited by lasers. This light emission is used to infer promoter activity within a cell, either by microscopy or flow cytometry measurements. This ability to measure gene expression in individual cells allows investigators to measure gene expression noise. On the other

hand, the transcriptional regulator *tetR* is used by investigators to "tune" the input to a system. By adding an analogue of tetracycline into the medium, investigators are able to control how effectively TetR binds to DNA. By changing the binding affinity of TetR to DNA, genes under TetR's control can be expressed strongly or weakly - allowing investigators to observe the relationship between gene expression and cell phenotype. However, while many of the basic reactions of TetR-based systems are known, a quantitative understanding of these systems as a whole is lacking.

In order to understand how genotype affects cell phenotype, I used mathematical modeling to describe and predict the behaviors of synthetic gene expression systems based on members of the *tetR* and *GFP* gene families (Chapter 2, Chapter 3, and Chapter 4). These mathematical models were used to understand how noise propagates through simple synthetic gene networks (Chapter 2), develop synthetic gene circuits with linear responses to inducer (Chapter 3), and predict how noisy gene expression allows cell populations to maximize their fitness in stressful environments (Chapter 4).

In the following sections I review the mathematical models used to understand noisy gene expression (section 1.2), cellular memory (section 1.3), and previous studies of *tetR*-based gene expression systems (section 1.4). Noting that previous models had inconsistent predictions of gene expression, I hypothesized that incorporating the effects of cell growth into chemical kinetics models would be necessary to understand certain behaviors in *tetR*0-based gene expression systems (section 1.5).

1.2 Modeling noisy gene expression

From whence comes noise? Noise originates from the thermal motion of small particles like atoms and molecules. This was observed when Robert Brown observed "oscillations" (Brownian motion) of pollen grains (34). If we repeated his observations with grains suspended in fluid we would notice that the motions are not reproducible. Even after placing the grains in exactly the same positions, the pollen motions would not be reproduced. This is because when we make these observations, we do not reproduce the exact positions and velocities of the fluid molecules colliding with the pollen grains (35).

Pollen grains may collide with each other at various times. Although pollen grains will simply bounce off of each other when they collide, molecules may undergo reactions when they collide with other reactive molecules. Thus molecules in a solution, subject to thermal motions, will also have variable collision and reaction times (Fig 1.1). Thermal fluctuations apply to more than just molecular collisions and reactions – they also affect the breaking of chemical bonds and molecular degradation.

Random reaction times imply that the numbers of a molecular species can fluctuate over time. For example, if the reaction time for producing a molecule is unusually short, there may be a temporary

excess of that molecule. When there are many molecules of the same type undergoing chemical reactions, the reaction times tend to average out such that the numbers of molecules are subject to little noise. However, when there are only a few reactive molecules of a certain type within a cell, the number of molecules may have high noise (36).

Intracellular protein motions are also subject to thermal fluctuations (37). This means that in the absence of a guiding force, proteins that control gene expression move randomly to find the promoters genes they regulate. Because small numbers of molecules are associated with high noise levels (38), and cells often contain only 1 or 2 copies of a particular gene, DNA based reactions are often implicated as a major source of noise (39-42). However, any chemical reaction involved in gene expression can contribute to noise. Thus, to the extent that the reaction and collisions of regulatory proteins, DNA, and other intermediate molecules are subject to thermal fluctuations, gene expression will be noisy.

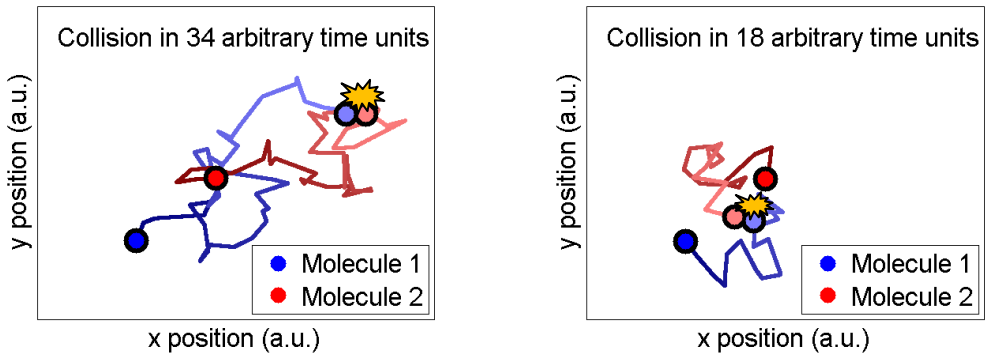


Figure 1.1. Hypothetical molecular collisions.

Two molecules were simulated according to Brownian dynamics to show how thermal fluctuations can result in "random" chemical reaction times. Simulations were stopped when the molecules were within 1 a.u. distance from each other. Starting from identical positions, the two molecules collided in A) 34 and B) 18 time units, illustrating how reaction times can vary. The time courses of the molecules are shown as gradually brightening colors. Final molecular positions are shown with the two molecules touching each other.

1.2.1 The Chemical Master Equation

Modeling thermal fluctuations for each molecule is computationally demanding. Instead, the times taken by the many non-reactive collisions between reactive collisions are coalesced into random reaction times. This is done by modeling noisy chemical reactions as a Markov process; referred to as the Chemical Master Equation (see Fig. 1.2).

The Chemical Master Equation assumes that a mixture of reacting chemicals with n molecular species, $\{S_1,...,S_n\}$, is completely defined by the number of each species, given by the $n\times1$ vector $\mathbf{X}$. The state of the system can randomly transition to other states through a set of m reactions, $\{R_1,...,R_m\}$. The vector $\boldsymbol{\nu}_\mathbf{k}$ denotes the numbers of molecular species that change in the k^{th} reaction (R_k) (i.e., $\mathbf{X} \leftarrow \mathbf{X} + \boldsymbol{\nu}_\mathbf{k}$). The likelihood of transitioning to another state is defined by the probability of the k^{th} reaction occurring within a small span of time (dt), and is denoted $a_k(\mathbf{X})$. Because the time between reactions is random, the time evolution of the system is defined as the probability of being in the state $\mathbf{X}$ at a time t, given an initial state ($\mathbf{X}_0$) and an initial time (t_0), or $P(\mathbf{X},t|\mathbf{X}_0,t_0)$.

The probability of being in a state $\mathbf{X}$ changes because of two major factors- entering the state and exiting the state via a reaction. This formalism gives the Chemical Master Equation (43),

$$\frac{dP(\mathbf{X},t\mid\mathbf{X}_0,t_0)}{dt} = \underbrace{-\sum_{k=1}^{m} a_k(\mathbf{X})P(\mathbf{X},t\mid\mathbf{X}_0,t_0)}_{\text{Exit}} + \underbrace{\sum_{k=1}^{m} a_k(\mathbf{X}-\boldsymbol{\nu}_k)P(\mathbf{X}-\boldsymbol{\nu}_k,t\mid\mathbf{X}_0,t_0)}_{\text{Enter}}. \qquad [1]$$

The reaction propensities (a_k) are generally modeled as proportional to the number of potential of molecular interactions, or

$$a_k(\mathbf{X}) = c_k \prod_{i=1}^{n} \frac{X_i!}{(X_i - r_{ki})!}, \qquad [2]$$

where c_k is a rate constant, and r_{ki} is the number of X_i reactants required for the k^{th} reaction (see Fig. 1.2A). The reaction propensity is 0 if any of the molecule numbers can transition to less than 0 in the reaction. Reactions are often limited to a maximum of two reactants, since reactions with more than two reactants can be separated into sub-reactions.

1.2.2 The Gillespie algorithm

In general, it is not practical to solve the Chemical Master Equation [1]. Instead the Gillespie algorithm is used to numerically simulate chemical trajectories defined by the Chemical Master Equation (44). The Gillespie algorithm can be implemented by the following steps:

0) Set initial conditions, $\mathbf{X}(t=0)=\mathbf{X_0}$.

1) For each possible reaction, calculate a potential time to the next reaction as $\Delta t_k = -\frac{\ln[U(0,1)]}{a_k(\mathbf{X}(t))}$, where U is a uniformly distributed random number with values between 0 and 1, and a_k is the propensity function ([2], for example). For each Δt_k, $U(0,1)$ is a unique random number.

2) Choose the reaction with the smallest change in time, and update the number of molecules and time;

$$\mathbf{X}(t+\Delta t_\lambda)=\mathbf{X}(t)+\mathbf{v}_\lambda$$
$$t \leftarrow t+\Delta t_\lambda$$

where

$$\Delta t_\lambda = \min(\Delta t_k)$$

.

3) Repeat steps 1 and 2 until the time of the trajectory exceeds a maximum time value (i.e., until $t > t_{max}$).

An example of the Gillespie Algorithm is shown in Fig. 1.2C and D.

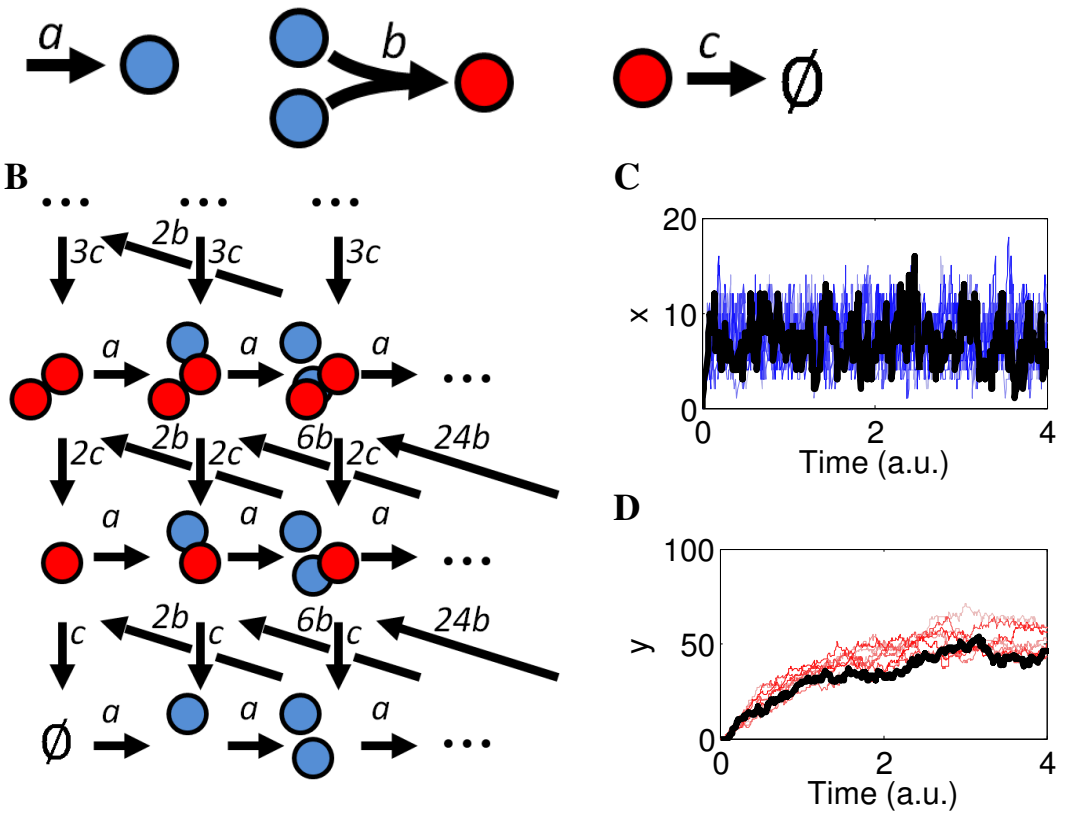


Figure 1.2. Chemical Master Equation example

A) Three mass action chemical reactions with "blue" (x) and "red" molecules (y) are described as an example. Reaction rate constants are shown above the arrows. B) A grid of molecular states is shown for 0 to 2 red and blue molecules. The state of the system changes by the three chemical reactions. The Chemical Master Equation defines the probability of being in any state defined by molecular numbers. Ten Gillespie simulations were performed based on this reaction scheme for the blue (C) and red (D) molecules, with a = 100, b = 1, and c = 1. A single representative timecourse is displayed in black. Blue molecules are denoted 'x', while red molecules are denoted 'y'.

1.2.3 Approximating the Chemical Master Equation

It is often desirable to approximate the Chemical Master Equation [1] by the Fokker-Planck equation (45):

$$\dot{p} \approx -\sum_{i=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial X_i}\left[v_{ik}a_k p\right]+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\frac{\partial^2}{\partial X_i \partial X_j}\left[v_{ik}v_{jk}a_k p\right], \qquad [3]$$

where p is the probability distribution function $p(\mathbf{X},t|\mathbf{X}_0,t_0)$, a_k is the propensity function (as defined by [2], for example), and v_{ik} is the number of S_i molecules that change in the k^{th} reaction. In this

formulation, the number of molecules can be any real non-negative number (hence the transition from $P \rightarrow p$). Correspondingly, the propensity function and probabilities exist at states with non-integer molecules. This approximation assumes that reaction rates are not sensitive to changes in molecule numbers, and that molecules are continuous (45). For example, active genes that switch states from 0 to 1 in a cell will drastically change the rates at which proteins are produced, which would violate the first assumption, and might not be reasonably thought of as being continuous, which would violate the second assumption. As a practical rule, promoter dynamics are often severely distorted by this approximation.

In order to simplify the mathematical notation, I define the deterministic "drift" matrix as:

$$A_i = \sum_{k=1}^{m} \nu_{ik} a_k \qquad [4]$$

and the stochastic "diffusion" matrix as:

$$B_{ij} = \sum_{k=1}^{m} \nu_{ik} \nu_{jk} a_k \; . \qquad [5]$$

The effects of the drift and diffusion matrices have analogies to the behaviors of dye dropped into water. Dye concentrations will "diffuse" out from where the dye was dropped even if the water is still. Similarly, the diffusion matrix describes how the distribution of probabilities "diffuses" outward over time. This expansion of probability distributions corresponds to random changes in molecule numbers.

If there is a current to water, as in a stream, dyes will "drift" with the current in addition to diffusing outward. Similarly, the drift matrix describes the general trends for how chemical concentrations change over time. If the initial number of molecules is known, and the system of molecules studied is only described by the drift equation, then the probability distribution for chemical states will have no variance - it will only move by drift. Thus the drift term corresponds to deterministic changes in the chemical state.

In terms of the drift and diffusion matrices, the Fokker-Planck equation is written as

$$\dot{p} \approx \underbrace{-\sum_{i=1}^{n} \frac{\partial}{\partial X_i} \left[A_i p \right]}_{\text{Deterministic}} + \underbrace{\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial X_i \partial X_j} \left[B_{ij} p \right]}_{\text{Stochastic}}. \qquad [6]$$

1.2.4 The Fokker-Planck equation for chemical concentrations

One method to justify treating the number molecular species as a continuous number is to model the evolution of molecular concentrations instead of molecular numbers (45). This can be done by transforming molecule numbers to molecular concentrations by the relationship

$$x_i = \frac{X_i}{V}, \quad [7]$$

where V is the volume of a system (e.g., cell volume). The reaction propensities [2] are transformed to

$$\tilde{a}_k(\mathbf{x}) = h_k \prod_{i=1}^{n} x_i^{r_{ki}}, \quad [8]$$

where h is the reaction rate, and r_{ki} is the number of x_i reactants required for the k^{th} reaction. Rewriting the drift matrix as

$$\tilde{A}_i = \sum_{k=1}^{m} \nu_{ik} \tilde{a}_k, \quad [9]$$

and the diffusion matrix as

$$\tilde{B}_i = \frac{1}{V} \sum_{k=1}^{m} \nu_{ik} \nu_{jk} \tilde{a}_k, \quad [10]$$

the Fokker-Planck equation may be written as

$$\dot{\tilde{p}} \approx -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[\tilde{A}_i \tilde{p}\right] + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} \left[\tilde{B}_{ij} \tilde{p}\right], \quad [11]$$

where $\tilde{p}$ is the probability distribution function as a function of molecular concentrations, $\tilde{p}(\mathbf{x}, t \mid \mathbf{x_0}, t_0)$ (45).

1.2.5 The Linear Noise Approximation

Obtaining analytical solutions of the Fokker Planck equation can be difficult. For this reason, an even more reductive method based on the Linear Noise Approximation is used to describe the noise of a system (43,46).

The Linear Noise Approximation makes the 3 assumptions that:

1) Concentrations are centered around an attracting stable steady state, $(\mathbf{X} \to \mathbf{X^*}; \mathbf{A}(\mathbf{X^*}) = \mathbf{0})$.

2) The drift matrix is approximated by the first order Taylor expansion

$$\mathbf{A}(\mathbf{X}) \approx \underbrace{\mathbf{A}(\mathbf{X^*})}_{=\mathbf{0}} + \mathbf{J}(\mathbf{X^*})(\mathbf{X} - \mathbf{X^*}) + \ldots \quad [12]$$

where **J** is the Jacobian of the drift matrix,

$$J_{ij} = \frac{\partial}{\partial X_j} A_i, \quad [13]$$

evaluated at the steady stable state **X*.**

3) The diffusion matrix is approximated by its evaluation at the stable steady state, or **B**(**X***).

Applying these 3 assumptions to the Fokker Planck equation defines an Ornstein–Uhlenbeck process (47,48). Ornstein–Uhlenbeck processes have analytical solutions given by multivariate normal distributions with means **X*** and covariance matrices **σ**. The covariance matrix (**σ**) of an Ornstein–Uhlenbeck process can be found by solving the equation:

$$\dot{\boldsymbol{\sigma}} = \mathbf{J}(\mathbf{X}^*)\boldsymbol{\sigma} + \boldsymbol{\sigma}\mathbf{J}^{\mathrm{T}}(\mathbf{X}^*) + \mathbf{B}(\mathbf{X}^*). \quad [14]$$

It is usually more informative to solve the steady state solution of the covariance matrix. The steady state covariance matrix is given by the continuous Lyapunov equation:

$$\mathbf{0} = \mathbf{J}(\mathbf{X}^*)\boldsymbol{\sigma} + \boldsymbol{\sigma}\mathbf{J}^{\mathrm{T}}(\mathbf{X}^*) + \mathbf{B}(\mathbf{X}^*). \quad [15]$$

In general, I will be solving for the time invariant steady state solution of **σ** as a function of **X***. The Linear Noise Approximation may also be used for systems with dynamical limit cycles (49). Unfortunately, great care must be used when studying limit cycles, since **σ** may not be convergent and may asymptotically increase over time (50). Other behaviors such as multistability and bifurcations require more sophisticated methods than the naïve application of the linear noise approximation (49).

1.2.6 Estimating extrinsic noise for reaction cascades

The Linear Noise Approximation [14] can be used to quantify the effect of extrinsic noise (or noise propagation) through a network of reactions (51). Consider two molecules - X_1, and X_2, subject to the reactions

$$\begin{aligned} R_1 &: X_1 \xrightarrow{a_1(X_1)} \varnothing \\ R_2 &: \xrightarrow{a_2(X_1)} X_1 \\ R_3 &: X_2 \xrightarrow{a_3(X_1, X_2)} \varnothing \\ R_4 &: \xrightarrow{a_4(X_1, X_2)} X_2 \end{aligned} \quad . \quad [16]$$

where a_1, a_2, a_3, and a_4 are reaction propensity functions (e.g., [2]). In this case, X_1 affects X_2, but X_2 does not affect X_1. Reactions analogous to [16] will have diffusion [5] and Jacobian [13] matrices with the general forms:

$$\mathbf{B} = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix}$$
$$\mathbf{J} = \begin{bmatrix} J_{11} & 0 \\ J_{21} & J_{22} \end{bmatrix}, \qquad [17]$$

with solutions for X_1 and X_2 variances (σ_1^2 and σ_2^2) from the Linear Noise Approximation [14]:

$$\sigma_1^2 = -\frac{B_{11}}{2J_{11}}$$
$$\sigma_2^2 = \underbrace{-\frac{B_{22}}{2J_{22}}}_{\text{"intrinsic" variance}} \underbrace{-\frac{\overbrace{B_{11}}^{\sigma_1^2}}{2J_{11}} J_{21}^2 \frac{1}{J_{22}\left(J_{11}+J_{22}\right)}}_{\text{"extrinsic" variance}}. \qquad [18]$$

where J_{11} and J_{22} are negative if the system has relaxed to a stable steady state. The corresponding CVs, or noise terms $\left(\eta_1, \text{and } \eta_2\right)$ are given as:

$$\eta_1^2 = -\frac{B_{11}}{2J_{11}} \frac{1}{X_1^2}$$
$$\eta_2^2 = \underbrace{-\frac{B_{22}}{2J_{22}} \frac{1}{X_2^2}}_{\text{"intrinsic" noise}} + \underbrace{\eta_1^2 \frac{X_1^2}{X_2^2} J_{21}^2 \frac{1}{J_{22}\left(J_{11}+J_{22}\right)}}_{\text{"extrinsic" noise}}. \qquad [19]$$

When derived from the Chemical Master Equation, the diffusion and Jacobian matrices are:

$$\mathbf{B} = \begin{bmatrix} \nu_{11}^2 a_1\left(X_1\right) + \nu_{12}^2 a_2\left(X_1\right) & 0 \\ 0 & \nu_{23}^2 a_3\left(X_1, X_2\right) + \nu_{24}^2 a_4\left(X_1, X_2\right) \end{bmatrix} \qquad [20]$$

and

$$\mathbf{J} = \begin{bmatrix} \frac{\partial}{\partial X_1}\left[\nu_{11} a_1\left(X_1\right) + \nu_{12} a_2\left(X_1\right)\right] & 0 \\ \frac{\partial}{\partial X_1}\left[\nu_{23} a_3\left(X_1, X_2\right) + \nu_{24} a_4\left(X_1, X_2\right)\right] & \frac{\partial}{\partial X_2}\left[\nu_{23} a_3\left(X_1, X_2\right) + \nu_{24} a_4\left(X_1, X_2\right)\right] \end{bmatrix}. \qquad [21]$$

However, due to the complexity of cell processes, the diffusion matrix is sometimes inferred from observed noise instead of being derived from the Chemical Master Equation (51,52).

These equations are used to quantify the propagation of noise through cascading networks (i.e., DNA→RNA→Protein, or cascades of transcriptional regulators). Equation [19] is used to justify the assertion that the more sensitive X_2 is to X_1, the more noise from X_1 will propagate to X_2 through the J_{21} (Jacobian/sensitivity) term.

1.2.7 The deterministic chemical kinetic formulation

The simplest approximation of the Chemical Master Equation describes chemical concentrations by the deterministic drift equation [4] (53) as

$$\dot{\mathbf{X}} = \mathbf{A}(\mathbf{X}). \qquad [22]$$

This approximation completely ignores stochastic events. It is most appropriate for monostable systems with large numbers of molecules (45), since gene expression noise is generally inversely proportional to the number of mRNA transcript and protein numbers in a cell (54). These guidelines provide a general rule of thumb about when it is most appropriate to use the deterministic formulation. However, probabilistic models of cells (or of biochemical reactions) are capable of capturing behaviors seen in deterministic models, while deterministic models do not capture all of the behaviors in probabilistic models (44).

1.2.8 Growth affects chemical kinetics

It is often convenient mathematically to assume a constant cell volume when modeling chemical kinetics (45,55). However, for many living organisms, cell volume is constantly changing. Practically, this has the effect of changing the concentrations of proteins over time (56). In the absence of protein production, cells that only experience growth will dilute proteins until there is practically no concentration left within a cell.

One approach for quantifying the effects of cell growth is to model chemical concentrations instead of molecule numbers (i.e., $x = X/\mathrm{V}$, where V is cell volume) and subtract a dilution term from the rate equations, since

$$\dot{x} = \frac{\dot{X}V - X\dot{V}}{V^2} = \frac{\dot{X}}{V} - \underbrace{gx}_{\text{dilution}}, \qquad [23]$$

where g is the relative growth rate of cell volume $\left(g = \dot{V}/V\right)$. The remaining volume ($V$) term on the right hand side is absorbed into the reaction rates and chemical species so that models are defined by chemical concentrations.

A second approach has been used whereby individual cells are simulated, and divide depending on their internal chemical states (57-59). Upon division, molecules are re-apportioned to daughter cells by a binomial distribution (60,61). Although this approach accounts for the effects of cell division, these simulations are generally more complex and demanding to analyze.

1.3 Cellular memory

The history of a cell can affect how it responds to its environment (62,63). For example, *Bacillus subtilis* cells temporarily exposed to a starvation environment, and then moved back to a high nutrient environment may not commit to long-term phenotypic changes. In contrast, after a prolonged period of starvation these cells may commit to sporulation (64). This is an example of cellular memory, which may be defined as the extent to which a cell's history determines its phenotype. A broader definition is sometimes defined to include the extent to which a cell's *lineage* determines its present phenotype (65).

Several measurements have been used to quantify cellular memory. These measurements include the presence of hysteresis (66-69), autocorrelation of gene expression (70), mutual information between cell states and previous environments (64), and the escape rates of cells out of a phenotype (25,71).

One important distinction between these different measurements is whether they can be used in a changing environment, or in a constant environment. In a changing environment, a cell may be responding deterministically to changes in the environment, it may be randomly transitioning to different phenotypes, OR it may be acting due to a mixture of these two processes. Although I briefly describe four types of memory, I use escape rate based measurements of memory in subsequent chapters. This is because the escape rate based measurements describe cell memory in a constant environment, and relate memory to the time cells display a phenotype, which could significantly affect cell fitness.

1.3.1 Hysteresis as a measure of cellular memory

Hysteresis arises from the existence of multiple chemical steady states (72), and affects how a cell responds to a changing environment based on its current state (66-69). For inducible gene expression systems, hysteresis can be inferred by the temporary addition (or removal) of inducer to a cell's environment. If a cell does not retain its original phenotype in the presence of its initial induction, then the cell displays hysteresis.

This measure of cellular memory implies that stochastic transitions between cell states are not important. For most reaction schemes, every reaction is either reversible (e.g., binding and unbinding of two molecules) or has a complementary reaction (e.g., degradation and production of a molecule). Because of this, ofttimes every chemical state within a cell is accessible. Given enough time, both chemical steady states will be visited if they are both accessible. Therefore, stochastic transitions may only be ignored when the time scales being studied are small. In this situation, hysteresis can be used to determine whether a cell has memory or not.

1.3.2 Autocorrelation as a measure of cellular memory

A second measurement of memory is based on autocorrelation (70). For a cell with a phenotype measured over time (e.g., protein concentrations), autocorrelation is the correlation between the cell's phenotype and its past phenotype, or

$$m = \rho\left(x\left(t-\Delta t\right), x\left(t\right)\right) \qquad [24]$$

where ρ is the correlation of the phenotype x (at time t) with itself in the past (at time $t-\Delta t$), and m is the metric of memory.

Intuitively, this measurement asks how uncertain a cell's history is given its present state in a constant environment. A cell's recent past should be more certain than in its distant past, which should result in a higher correlation. However this may not be true when the phenotype oscillates. In this case, cellular memory may increase with increasing Δt as it approaches x's period of oscillation.

1.3.3 Mutual information as a measure of cellular memory

The mutual information between a cell population's response and its environmental history proposed as a measure of memory (64). Mutual information is similar to correlation, and measures how much is known about a random variable X, given a random variable Y. In this case, memory was defined as how much was known about a cell population's environmental history (X), given its gene expression over time (Y).

The environments that cells were treated with were defined by the experimentalists, which could bias the randomness of the environment. Additionally, gene expression time courses over time were clustered into several classes of time course. This means that memory was effectively measured as how much was known about the experimentalists choice of environment, given how cell populations' gene expression over time was classified. In this definition, there must be a distinct change in environmental conditions to a common environment.

1.3.4 Escape rate out of a cell state as a measure of cellular memory

Chemical kinetics models can be used to study the time it takes gene expression levels to pass a boundary. The average time it takes for a cell to cross a boundary is a fourth measurement of memory (25,71).

For a system that executes a continuous-time random walk over discrete states with only one chemical species (i.e., X=0, 1, 2, …), the average time to transition from α to α +1 molecules is

$$\tau(\alpha,\alpha+1)=\frac{\sum_{X=0}^{\alpha}P(X)}{k_{\rightarrow}(\alpha)P(\alpha)}\ , \qquad [25]$$

where P is the stationary probability distribution of molecules, and $k_{\rightarrow}$ is the rate at which molecules are produced (73).

The average time to transition from α to β molecules ($\alpha < \beta$) can be described by the recursive function

$$\tau(\alpha,\beta)=\tau(\alpha,\alpha+1)+\tau(\alpha+1,\beta), \qquad [26]$$

with the stopping condition, $\tau(\beta,\beta)=0$. This equation has also been described as (73,74):

$$\tau(\alpha,\beta)=\sum_{n=\alpha}^{\beta-1}\left[\frac{\sum_{X=0}^{n}P(X)}{k_{\rightarrow}(X)P(X)}\right]. \qquad [27]$$

For continuous probability distributions, the mean average time within a state can be described as:

$$\tau=\frac{\int_{0}^{\beta}\pi(X)dX}{I_{\rightarrow}}, \qquad [28]$$

where π is the stationary probability distribution, and the average time in a state (τ) from 0 to β is the ratio of the probability of being in that state divided by the probability current out of that state, I (75,76). The probability current is the rate at which cells leave their state times the probability distribution of cells at the boundary, or

$$I_{\rightarrow}=k_{\rightarrow}(\beta)\pi(\beta), \qquad [29]$$

and is equivalent to [27] when $\alpha = \beta$-1.

Because cells form a stationary probability distribution, this implies that cells must remain in a constant environment for this definition of memory. For cells undergoing random differentiation, escape times can be used to estimate the time a cell takes to differentiate. Recent evidence has suggested that in fact, a probabilistic model of differentiation may be necessary to understand cell differentiation (22,77).

1.3.5 Bet-hedging survival strategies

Cells may or may not be capable of responding in a controlled manner to stressful environments. In some cases, cells sense the presence of toxic molecules, and correspondingly express a gene to nullify their effects (78-80). In cases where sensing the environment is fast, accurate, and has low fitness costs, this may be the best strategy (20,21,81,82). However, when a cell's response to a toxic environment is slow, it may be more beneficial for the cell population to diversify its phenotypes.

The strategy of diversification is called bet-hedging. In this case, cells are unable to sense their environments. Instead, they transition to multiple phenotypes stochastically. Cells are predicted to be most fit when their switching rates between phenotypes follow the relationship

$$H_{ij}(\text{optimal}) = b_{ij} / T_j$$

where H_{ij} is the switching rate from phenotype i to j, b_{ij} is the switching rate from environment i to j, and T_j is the average time cells spend in environment j. Phenotype j is defined to be the most fit phenotype in environment j (20).

This result means that for bet hedging to be a good strategy, when the environment switches quickly, stochastic cell switching should be fast (low cellular memory). If the environment changes slowly, stochastic cell switching should be correspondingly slow (high cellular memory).

1.4 TetR based synthetic gene expression systems

The family of transcription factors (both repressors and activators) derived from the Tetracycline Repressor (*tetR*) are often used in eukaryotic synthetic gene expression systems. This is because 1) their transcriptional activity can be easily controlled by members of the tetracycline family of molecules, 2) their DNA binding regions are non-native to eukaryotes, and 3) because only a single *tetR*-based gene is required to regulate downstream activity. Furthermore, *tetR* based systems can be induced to intermediate levels of activity - from negligible expression to fully induced gene expression - by progressively increasing tetracycline concentrations. Doxycycline (Dox) and Anhydrotetracycline (ATc) are the most commonly used tetracycline inducers because they have high affinities to TetR with almost no toxic effects on eukaryotic cells (83). Depending on what gene expression characteristics are desired, the transcriptional activators "tetracycline-controlled trans-activator" (*tTA*) and the "reverse tetracycline-controlled trans-activator" (*rtTA*) are sometimes used instead of the transcriptional repressor, *tetR*.

The TetR transcription factor originates from prokaryotic organisms, and acts to sense and sequester members of the tetracycline family. In the absence of tetracyclines, TetR binds to the two operators *tetO1* and *tetO2* to block transcription of the target gene (84). TetR will bind to

tetracyclines, at which point its affinity for the *tetO* sites is greatly reduced (83,85). Thus when bound to tetracyclines, TetR is no longer able to act as a transcriptional repressor.

It was found that multiple point mutations of the *tetR* gene could reverse TetR sensitivity to tetracycline analogs (86). This mutant, called reverse TetR (*revtetR*), preferentially binds to *tetO* sites in the *presence* of tetracycline analogues.

The *tTA* and *rtTA* genes were created by fusing the VP16 transcriptional activator from the Herpes simplex virus to the *tetR* and *revtetR* genes, respectively (87,88). By placing the *tetO2* binding sites upstream of the TATA box of target genes, the repressory activity of the *tetR* and *revtetR* domains is minimized while the VP16 domain is able to activate target gene expression. Thus the tTA protein activates transcription in the absence of tetracyclines, while rtTA activates transcription in the presence of tetracyclines.

The VP16 domain has been implicated in sequestering (squelching) various transcription factors and preventing them from performing their normal functions (89). Although squelching has been implicated to reduce cell growth (90), the exact mechanism by which this happens is unknown.

1.4.1 Modeling TetR and rtTA systems

Constitutively active *tetR* genes cause noisy gene expression of their targets at intermediate levels of induction. These targets can display extremely sensitive, sigmoidal dose-responses of mean downstream activity as tetracycline analog concentrations are increased. Corresponding to this increased sensitivity, downstream noise peaks (see Fig. 1.3) (91-93). Because TetR is a popular and efficient transcriptional repressor, many models of constitutive gene expression systems have been created to explain its behavior (52,91,92,94-97).

The simplest models describe TetR repression as a Hill function with TetR binding reprieved by the addition of the tetracycline analog anhydrotetracycline (ATc) as

$$\dot{z} = a\frac{ATc^n}{ATc^n + TetR^n} + l - dz \ , \qquad [30]$$

where TetR is constant, ATc tunes repression, and z is downstream gene expression (often measured by yEGFP) (94,97). The parameters l and a represent gene expression of fully repressed and fully unrepressed downstream gene expression, n represents the cooperativity or Hill constant of TetR binding to its DNA binding sites, and d is the rate at which the downstream gene product is lost, either through degradation or dilution. When the Hill constant is greater than 1 ($n > 1$), the dose-response is said to be ultrasensitive (98).

Surprisingly, after varying the number of *tetO2* binding sites over a small range (1 to 3) and fitting a Hill function to the resulting dose-response, the Hill constant (n) varied from ~3 (less sensitive) to ~10 (extreme sensitivity) (92). Because the Hill constant can be interpreted as the number of molecules that must bind at once to affect a molecule, the highest Hill constant (n = 10) was unexpected since it exceeded the theoretical number of ATc molecules that could block TetR activity (in this case, 2 ATc binding sites × 2 *tetO2* sites implied that $n \leq 4$). For this reason, the authors argued that a chemical kinetic scheme explicitly accounting for intermediate reactions was necessary to model chemical kinetics instead of a phenomenological Hill function. However, the dissociation rate of TetR to DNA due to ATc was still assumed to be extremely cooperative.

The sensitivity of downstream gene expression to TetR levels has strong implications for gene expression noise, since more sensitive downstream promoters have been shown to amplify upstream TetR noise (51,52,96). Unfortunately, none of these models explain why downstream sensitivity can fluctuate so wildly with such similar genetic architectures.

Distinct, potentially useful behaviors have been studied by changing the regulatory architecture of *tetR* or its variants. *tetR* negative autoregulation has been shown to have potentially useful properties for reducing noise, and decreasing response times to induction (26,94,95,97,99). Although these properties have been justified using simple models- these results depend upon the strength and severity of TetR repressory activity. Because this strength and severity (i.e., sensitivity) does not appear to be well understood, the proposed mechanisms by which negative autoregulation decreases noise and decreases response times may be questionable.

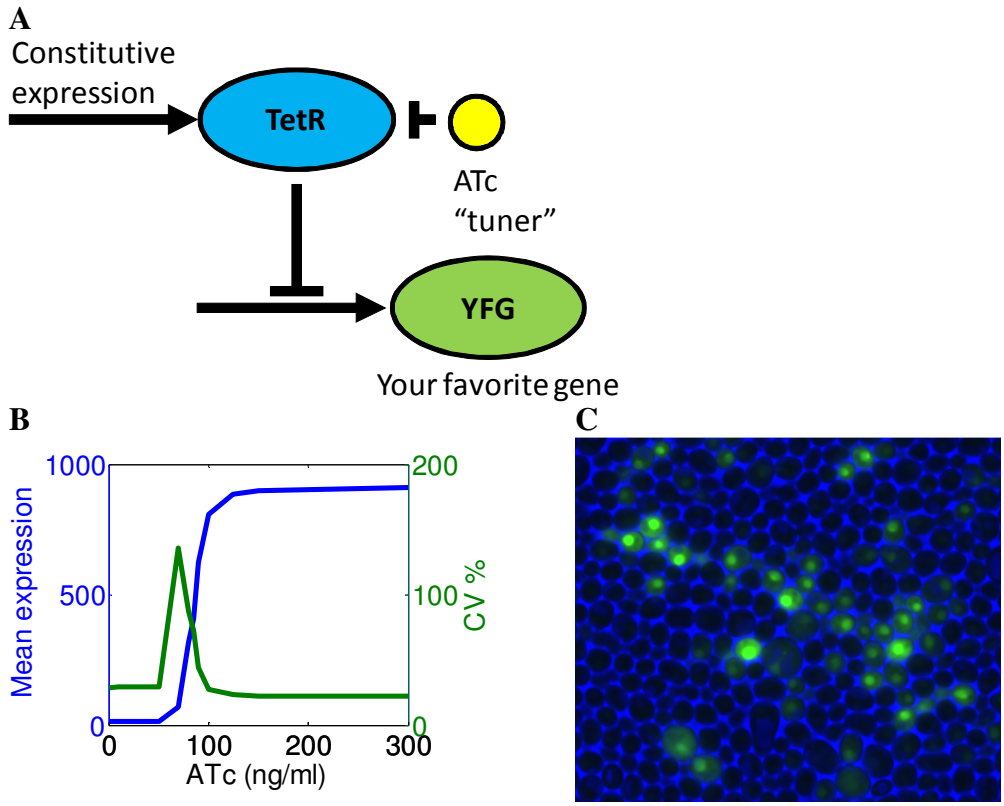


Figure 1.3. Constitutively expressed tetR dose-response.

A) Template for constitutively expressed tetR gene expression systems. B) The green fluorescence protein (GFP) reporter was expressed from a GAL1 promoter containing TetR binding sites in S. cerevisiae. tetR was constitutively expressed from the GAL1 promoter. ATc was added to the cell culture media to block TetR repression, resulting in a sigmoidal dose-response (blue). The Downstream gene expression coefficient of variation (CV) peaks when most, but not all, of TetR's repressory activity is blocked (green). The CV is used to measure gene expression stochasticity or "noise" over an entire population at a single time point. C) An example of noisy gene expression in yeast cells is shown in this picture. Gene expression, indicated in green, varies strongly from cell to cell (image provided by Elizabeth Van Itallie and Mathew Bennett).

Conversely, *rtTA* positive autoregulation can result in slow, noisy, and bimodal gene expression properties (94,100-104). Models have been created to describe rtTA based positive feedback loops (101), but they have ignored the growth effects of rtTA (90). Studies of rtTA autoregulatory loops have hinted at growth retardation by showing that cells expressing high levels of rtTA appeared to have significantly fewer offspring over 22 hours than cells with low rtTA grown over 16 hours (100). This effect may be important, since cell growth acts strongly on the chemical kinetics of cells (see section 1.2.8).

1.5 Hypothesis

Chemical kinetics models have been used in a limited manner to understand the gene expression properties of tetR and rtTA gene expression systems, but have ignored the effects of cell division and cell growth. Excluding the properties of growth could mask hidden behaviors. Taking these factors into consideration, I asked; how do cell division and growth affect *tetR*-based gene expression systems? I hypothesized that:

Cellular growth and division affects chemical kinetics resulting in

1) sequestration of TetR molecules by inducer,
2) disequilibrium of inducer concentrations internal and external to cells,
3) and an overrepresentation of quickly growing cells in the population.

1.6 Specific Aims

In order to understand how noise and memory were controlled in *tetR* based gene expression systems, I created the three aims:

Aim 1 Determine the origin of ultrasensitivity and noise in constitutive TetR systems.

Aim 2 Determine the sensitivity of *tetR* autoregulatory systems to promoter architecture.

Aim 3 Develop a method for modifying *rtTA* autoregulatory system noise and memory.

My efforts to pursue these three aims are described in the next three chapters.

Chapter 2 The effects of molecular sequestration on noise propagation

2.1 Objectives

Inhibiting the activity of constitutively expressed TetR by increasing anhydrotetracycline (ATc) concentrations results in a surprisingly sharp, abrupt increase of target gene expression. This sharp increase in downstream target gene expression is accompanied by a spike in gene expression noise. This noise "spike" has been used to compare phenotypic benefits from noisy downstream gene expression to gene expression with low noise levels (91). This extremely sensitive response has been attributed to cooperative binding affinities for TetR to its *TetO* recognition sites. However, this cooperativity is not justified biologically. My objective is to determine whether ATc-TetR sequestration provides a plausible mechanism of ultrasensitivity (105,106) and noise amplification for constitutively expressed *tetR* gene expression systems. This section closely follows the work done by Adams et al 2010 (107). All experimental work in this chapter was performed by Kevin Murphy.

2.2 Methods and Materials

2.2.1 Cells analyzed

I analyzed yeast strains constructed by Kevin Murphy (107). Briefly, these cells had genes under the control of chromosomally integrated GAL10-GAL1 promoters. *tetR* was placed under the control of the GAL10 promoter. Four TATA box mutations were introduced to the GAL10 promoter, modifying *tetR* expression. These mutations were:

GAL-10 WT	TATAA
GAL-10 int1	TATA<u>T</u>
GAL-10 int2	<u>A</u>ATAA
GAL-10 sev1	TAT<u>T</u>A
GAL-10 sev2	<u>GC</u>TAA

yEGFP was placed under the control of the GAL1 promoter. Three *tetO2* binding sites were placed between the GAL1 TATA box and the *yEGFP* gene (T123 promoter). The GAL10 and GAL1 promoters were constitutively activated by incubating cells in 2% galactose. TetR repression was weakened by adding ATc to the medium.

In addition to these five strains, five additional control strains were constructed where *yEGFP* was placed directly under the control of the previously described GAL-10 mutants.

2.2.2 Flow cytometry analysis

yEGFP fluorescence in cells was measured by flow cytometry. Cells were gated by their forward and side scatter values. The gating used was a rectangle. At least 5,000 cells were retained from 50,000 total cells.

2.2.3 The template model

A model was created with all major kinetic reactions associated with constitutive TetR expression. *yEGFP* gene expression mean and noise were simulated using the Gillespie algorithm, and solved with the Linear Noise Approximation [14]. Because the Gillespie simulations were ergodic, sample statistics at each ATc concentration were taken from a single time course for 5000 simulated hours.

The template model incorporates promoter binding and unbinding dynamics to take into account the effect of GAL10 mutations and GAL1 promoter noise. Gene expression is simulated by transcription of active promoters, and translation of mRNA intermediates. mRNA degradation rates are assumed to be much higher than cell growth (no dilution). mRNA is translated to yEGFP monomers and TetR dimers. TetR and yEGFP do not degrade, but dilute out of the cell due to growth as "stable" proteins (see section 6.14) (108). ATc can diffuse into and out of the cell, or it can bind irreversibly to one of TetR's two binding sites. TetR was allowed to bind to one of 3 *tetO2* binding sites for the yEGFP promoter, and thus inactivate yEGFP expression. TetR bound to DNA hindered further TetR binding to the promoter consistent with previous results (92). This model is described by Dizzy code (sections 6.1 and 6.2) (109) corresponding to Fig. 2.1, with parameters as described in section 6.2.

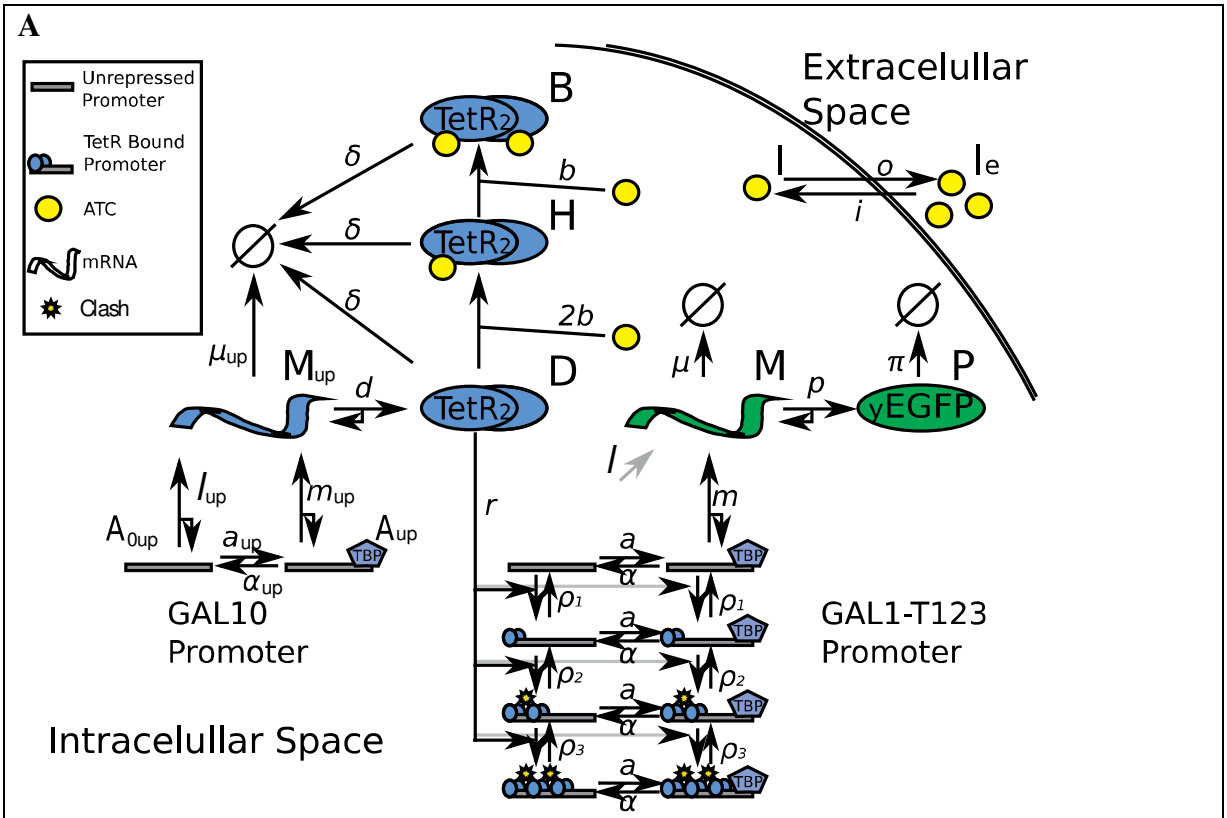


Figure 2.1. Schematic of the template model.

This schematic describes the molecular model used to simulate GAL10-T123 mutants. The chemical kinetic reactions and parameters are given in dizzy code found in section 6.1, with parameters described in section 6.2

2.2.4 The promoter model

I approximated the promoter model from the reaction scheme of the template model as

$$\begin{aligned}\text{TetR production:}&\xrightarrow{a_x} x\\ \text{TetR dilution:}&\; x\xrightarrow{g}\varnothing\\ \text{yEGFPproduction:}&\xrightarrow{a_z F(x,C)+l_z} z\\ \text{yEGFPdilution:}&\; z\xrightarrow{g}\varnothing\end{aligned}\quad , \qquad [31]$$

where x is the concentration of TetR and z is the downstream yEGFP concentration. The rate constants correspond to *tetR* expression (a_x), *yEGFP* expression (a_z), dilution out of the cell due to growth (g), and basal *yEGFP* expression (l_z). TetR repression of the *yEGFP* promoter is modeled by the ODE

$$\dot{F} = -rx^n F + \rho I^n (1-F), \qquad [32]$$

where F is the free downstream promoter (0 corresponding to fully repressed, 1 corresponding to fully active), r is the binding rate of TetR (x) to F, ρ is the binding rate of ATc to TetR, I is the ATc concentration, and n is the cooperativity of TetR binding to DNA and ATc binding to TetR. Assuming fast promoter dynamics, activation of the target gene is described by the Hill function

$$F(x,C)=\frac{C^n}{C^n+x^n}, \qquad [33]$$

where

$$C=\left(\frac{\rho}{r}\right)^{\frac{1}{n}} I. \qquad [34]$$

The promoter model is commonly used to describe TetR repression (52,94,96,97). The drift, diffusion, Jacobian, and steady state matrices and parameters used to obtain the Linear Noise Approximation are given in section 6.4.

2.2.5 The sequestration model

I approximated the "sequestration" model from the template model (107,110) as an alternative to the promoter model. In the sequestration model, ATc sequesters TetR in a practically irreversible manner and explicitly takes into account ATc influx and outflux. The reactions are defined as

$$\begin{aligned}
\text{TetR production:} &\xrightarrow{a_x} x\\
\text{TetR degradation/dilution:}\ x &\xrightarrow{g} \varnothing\\
\text{ATc influx:} &\xrightarrow{C} y\\
\text{ATc outflux/degradation:}\ y &\xrightarrow{h} \varnothing\\
\text{TetR ATc binding:}\ x+y &\xrightarrow{b} \varnothing\\
\text{yEGFP production:} &\xrightarrow{a_z F(x)+l_z} z\\
\text{yEGFP degradation/dilution:}\ z &\xrightarrow{g} \varnothing
\end{aligned}, \qquad [35]$$

where x is the concentration of TetR, y is the ATc concentration within cells, z is the downstream yEGFP concentration, and C is the rate of ATc diffusion into the cell directly proportional to ATc concentrations in media. The rate constants correspond to *tetR* expression (a_x), *yEGFP* expression (a_z), TetR-ATc binding (b),dilution out of the cell due to growth (g), diffusion of ATc out of the cell (h), and basal *yEGFP* expression (l_z). The function for promoter activation is

$$F(x)=\frac{\theta^n}{\theta^n+x^n},$$

where θ is the dissociation constant of TetR to DNA, and n is the degree of cooperativity. I assume $n = 1$.

Growth was implicitly incorporated into the model with the irreversible binding of ATc to TetR, since ATc binds to TetR longer (38 hours) (111) than cell division times (2~3 hours). The drift, diffusion, Jacobian, and steady state matrices and the parameters used to obtain the Linear Noise Approximation are given in section 6.5. The resulting correlation matrix between chemical species is given in section 6.6.

2.3 Results

2.3.1 Yeast strains

Five different yeast strains were developed by Kevin Murphy to examine how TetR's gene expression affects its downstream targets. The bidirectional *GAL10-GAL1* promoter was used to constitutively express the transcriptional repressor TetR and the fluorescent yEGFP reporter in galactose containing media from the *GAL10* and *GAL1* promoters, respectively. Three TetR binding sites (*tetO2* sites) were inserted into the *GAL1* promoter (renamed T123 promoter) so that TetR repressed *yEGFP* expression (107). TetR repression was blocked by adding the molecule ATc to the growth media. ATc diffuses into cells and binds to TetR, preventing TetR from binding to *tetO2* sites. These promoters were chromosomally integrated into *S. cerevisiae* (see Fig. 2.2A).

Four yeast strains with mutations in the TATA box of the *GAL10* promoter controlling *tetR* expression were created. Point mutations were introduced into the TATA box of the *GAL10* promoters, reducing gene expression. These *GAL10* promoter mutants were ranked and then named according to their level of gene expression. The promoters by descending gene expression are *GAL10* wild-type (*GAL10WT*), intermediate 1 (*GAL10int1*), intermediate 2 (*GAL10int2*), severe 1 (*GAL10sev1*), and severe 2 (*GAL10sev2*). Because these mutants only refer to the *GAL10* promoter, I omit the *GAL10* prefix (e.g., *GAL10int1*→*int1*). The specific mutation for each TATA box is shown in Fig. 2.2B.

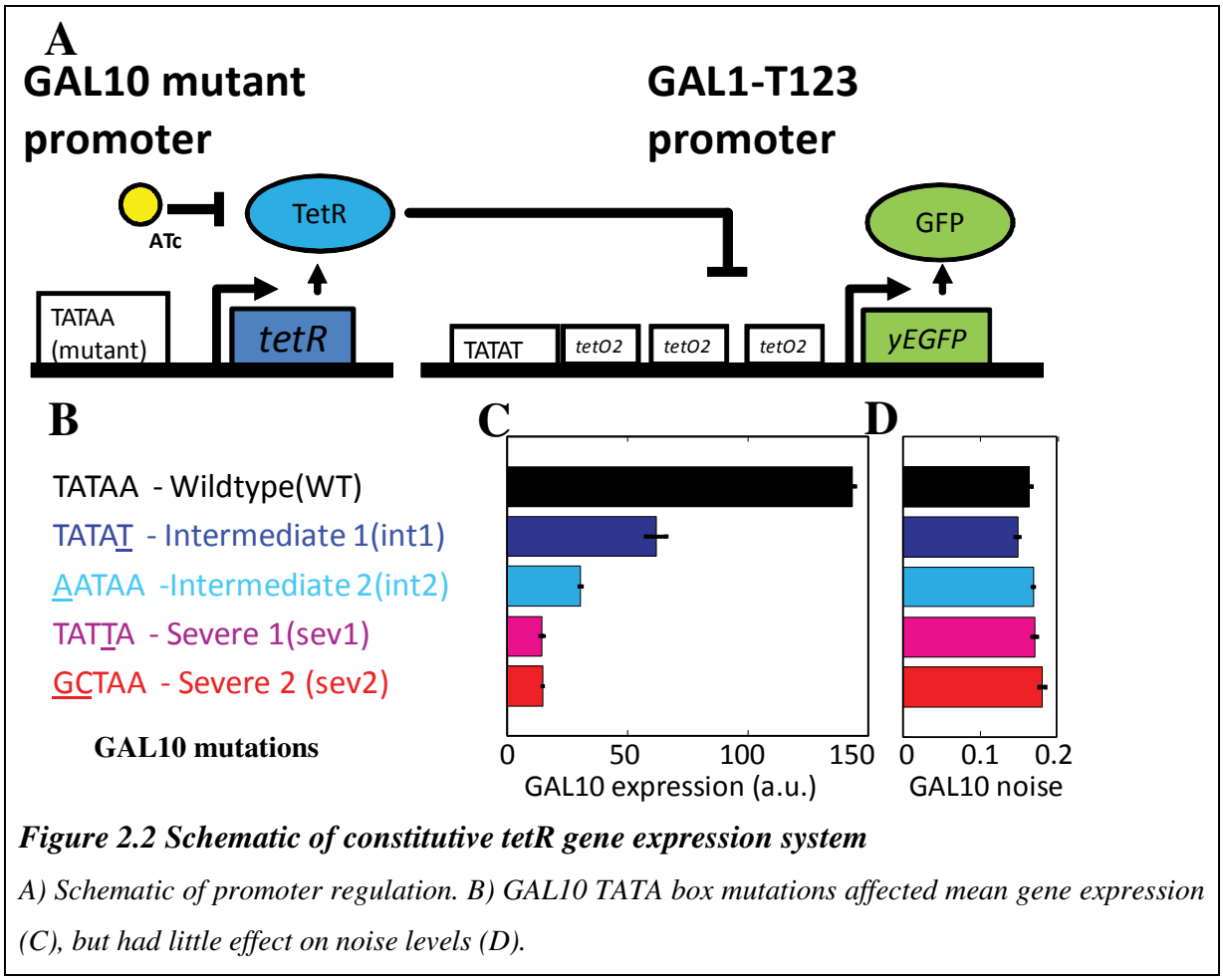


Figure 2.2 Schematic of constitutive tetR gene expression system

A) Schematic of promoter regulation. B) GAL10 TATA box mutations affected mean gene expression (C), but had little effect on noise levels (D).

2.3.2 Measurement of expression from upstream GAL10 promoters

The fluorescent reporter *yEGFP* was used to measure the gene expression of the five *GAL10* promoters. Each GAL10 promoter was measured in triplicate by flow cytometry in *S. cerevisiae* strains (see Fig. 2.2C) at fully inducing 2% galactose. The mutations in the *GAL10* TATA box reduced the mean gene expression of the wild-type variant by up to 90% (see Fig. 2.2C, Table 2.1).

Slight, but statistically significant differences were measured between the CVs (noise) of the 5 GAL10 promoter mutants ($p < 0.01$ using Statistica's 'multiple comparisons of mean ranks for all groups' nonparametric test, see Fig. 2.2D). This nearly constant level of GAL10 noise was unexpected since mass action kinetics of creation/degradation chemical reactions predict noise will increase mean gene expression decreases ($\eta \propto \sqrt{1/\mu}$, [19]) (54). Thus, the 90% reduction in observed gene expression was expected to correspond to a ~3-fold increase in gene expression noise. In contrast, gene expression noise barely changed even after a 90% reduction in observed gene expression.

	GAL 10 Mean Expression (a.u.)	GAL 10 Expression CV	Fit TBP binding rate (a, per hour)	Fit TBP unbinding rate (α, per hour)
WT	143.43 ± 6.80	0.162 ± 0.007	2.79	1.05
int1	61.78 ± 4.02	0.147 ± 0.005	6.82	16.22
int2	30.41 ± 1.01	0.168 ± 0.004	4.19	41.56
sev1	14.37 ± 0.83	0.169 ± 0.006	0.20	24.84
sev2	14.66 ± 0.33	0.179 ± 0.004	0.20	22.68

Table 2.1 Gene expression of GAL10 mutants

Gene expression mean and CV for mutated GAL10 promoters expressing GFP.

2.3.3 The effect of GAL10 TATA box mutations on T123 dose-response

Dose-responses were measured for each GAL10-T123 mutant where yEGFP was measured as a function of the inducer ATc (see Fig. 2.3). Mutations had little effect on the dynamic range of dose-responses – only a slight increase in yEGFP was observed in uninduced cells with severe GAL10 mutants, and no change was observable in fully induced cells (see Table 2.2). However, as mutations became more severe TetR repression was relieved at lower ATc concentrations. I defined a metric of ATc concentrations where yEGFP was 50% of its maximum value (the **induction threshold**) to quantify shifts in the dose-response. The induction thresholds were calculated from linear interpolation of the T123 dose-response curves. The resulting values had a nearly linear dependence on upstream promoter strength, with a correlation coefficient of $r = 0.9986$ (see Tables 6.1 and 2.2).

	T123 basal expression (a.u.)	T123 maximum expression (a.u.)	Induction threshold (ng/ml)
WT	1.75 ± 0.03	665.38 ± 7.07	39.6
int1	2.22 ± 0.07	656.49 ±17.77	17.2
int2	2.94 ± 0.03	696.23 ±30.43	6.5
sev1	3.73 ± 0.05	686.48 ± 9.98	4.1
sev2	5.02 ± 0.14	676.46 ±18.18	3.4

Table 2.2 T123 gene expression due to mutated tetR promoters

Minimum and maximum gene expression of GFP expressed from the T123 promoter controlled by TetR expressed from a mutated GAL10 promoter. ATc concentrations where GFP expression was half of their maximum levels is shown (induction threshold).

Next, yEGFP noise was also calculated for the GAL10-T123 mutants. As TetR repression is relieved by ATc, yEGFP noise peaks. This is consistent with the idea that sensitivity to upstream chemical species can amplify upstream noise [19] (51,52). As the mutations become more severe the noise peaks decrease (see Fig. 2.3B).

2.3.4 Template model simulations of T123 strains

Gillespie simulations of T123 strains by the template model (see Fig. 2.1, sections 6.1 and 6.2) were performed with appropriate rate constant modifications corresponding to the GAL10 mutation and ATc concentrations. I used the linear noise approximation [14] to fit parameters corresponding to *GAL10* expression mean and noise. The fit parameters correspond to the upstream GAL10 promoter activation/inactivation rates, and basal expression (a_{up}, and α_{up}, l_{up}, see Table 6.1).

The simulated *yEGFP* gene expression mean and noise mirrored experimental values (see Fig. 2.3). The induction threshold and downstream gene expression noise increased with GAL10 expression.

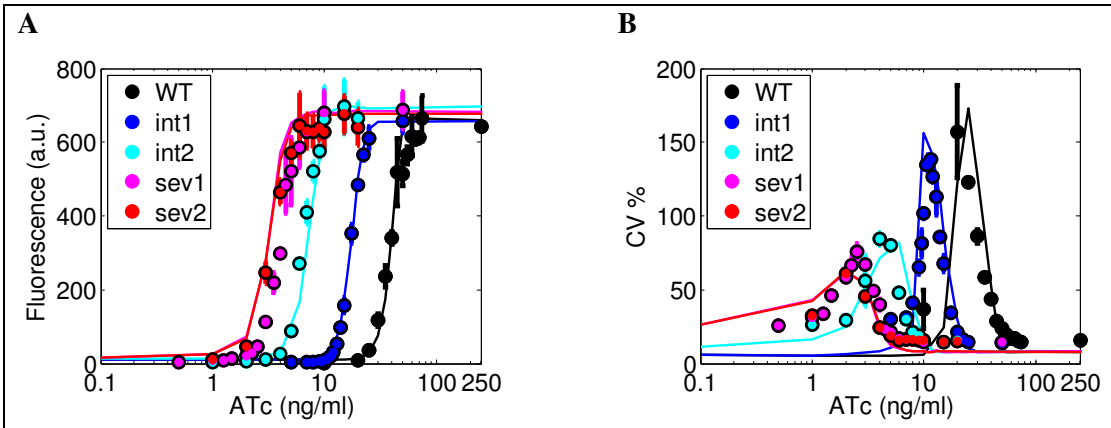


Figure 2.3. Upstream GAL10 TATA mutations change ATc dose-response

A) Mean fluorescence values measured for yeast cells with mutated GAL10 promoters expressing tetR, and downstream yEGFP promoters controlled by T123. Expression is shifted to the left as GAL10 mutations become more severe. B) The CV peaks occur at lower ATc concentrations and decrease with the severity of mutations. Experimental measurements are shown by colored circles. Gillespie simulations of corresponding Template models are shown by solid lines (see section 6.1 and 6.2).

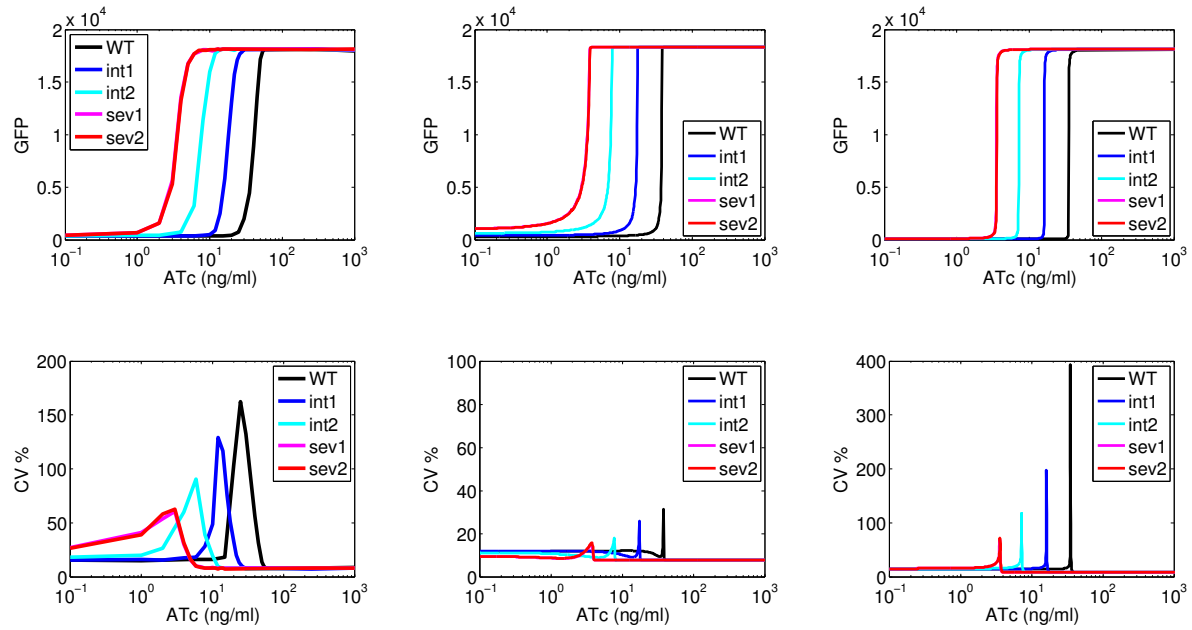


Figure 2.4. Comparing two forms of the LNA to Gillespie simulations

Simulations were performed for each GAL10-T123 mutant A) using the Gillespie algorithm for the Template model (see sections 6.1 and 6.2), B) the Linear Noise Approximation of the Template model using identical parameters(see section 6.3), and C) the Linear Noise Approximation of the Template model with promoter dynamics defined by a Hill function with Hill coefficient = 1 (see section 6.3). The deterministic means are sharper than Gillespie simulation means. The relative relationships are retained, with dose-responses shifting to the left and having decreasing noise peaks as mutations become more severe, but promoter dynamics are poorly approximated by the Linear Noise Approximation. The general reaction scheme is shown in Fig. 2.1, while reaction rates are given in section 6.2.

In order to simplify the Template model so that analysis could be performed, I asked how the Linear Noise Approximation could distort the dose-response. I compared the Linear Noise Approximations [14] of the template model to their Gillespie simulation counterparts (see Fig. 2.4A, B, and section 6.3 for Linear Noise Approximation equations). The Linear Noise Approximation had an overly sharp dose-response and muted noise levels.

Because average gene expression estimates in the Linear Noise Approximation are based on deterministic kinetics, statistical estimates will be severely distorted for single molecules such as gene promoters. Thus I modified the promoter kinetics to follow a Hill function with Hill coefficients and dissociation constants equal to 1 (see section 6.3). The behavior of the dose-response improved after this modification, although the Linear Noise Approximation retained its sharp dose-response (see Fig. 2.4C).

The sharp dose-response of the Linear Noise Approximation is due to the fact that it does not account for skewed yEGFP distributions. Experimentally, the fluorescence distribution became highly skewed when repression was almost relieved by ATc. A small population of cells had high gene expression, while the majority remained completely repressed. Because the Linear Noise Approximation ignores the effects of skewness, the mean values will appear to change more sharply than is seen using Gillespie simulations.

2.3.5 Comparing noise from the promoter model to the sequestration model

In order to better understand the origin of ultrasensitivity and noise in TetR cascades, I compared two models describing ATc's effects on TetR repression. The promoter model [31] was taken from previous models where ATc implicitly acts as a dissociation constant of TetR to DNA (see section 2.2.4 for derivation) (52,91,92,94,96) as supported by previous studies (112). I created an alternate hypothesis from the sequestration model [35], where ATc instead acts to irreversibly sequester TetR. The sequestration model was also supported by experimental measurements; ATc has delayed influx into cells due to membrane diffusion (113) , an unbinding half time of ~38 hours with TetR (111), and TetR appears to be a stable protein without any targeted degradation (114) so that ATc-TetR complexes leave the cell through dilution.

Most rate constants were retained from the template model. The WT-T123 dose-response was used to fit the Hill constant (n) for TetR-DNA binding for the promoter model, while the TetR ATc binding rate (b) for the sequestration model was reduced from the template model to soften the dose-response. Upstream TetR noise was modeled by changing a noise constant term (s) for TetR production in the diffusion matrix so that theoretical noise was identical to the measured noise from the GAL10 mutation. The relative strength of TetR expression (a_x) corresponded to measured mean GAL10 expression (parameter values are shown in sections 6.4 and 6.5).

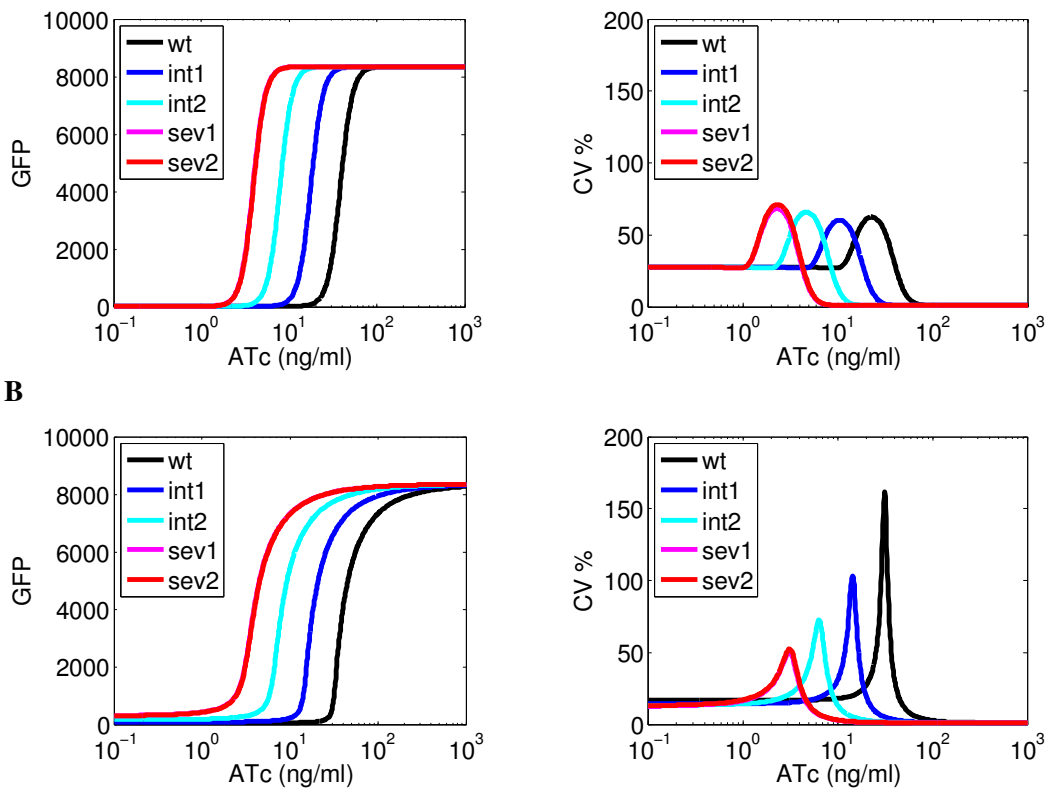


Figure 2.5. Comparison of the promoter and sequestration models

Mean and predicted CVs for the promoter model (A) and sequestration model (B). The promoter model has high cooperativity constants (n = 6 compared to n = 1 for the sequestration model) and downstream noise that is proportional to upstream noise. Parameters for the promoter model are given in section 6.4, while parameters for the sequestration model are given in section 6.5. Both models have comparable mean expression levels to experimental data. The promoter model's predicted noise levels do not decrease with the severity of GAL10 promoter mutations. In contrast, the sequestration model's peak noise decreases with severity of GAL10 promoter mutations, consistent with experimental results.

I compared the dose-response of the promoter model to the simplified model (see Fig. 2.5). Both the promoter [31] and sequestration models [35] show a linear relationship between the induction threshold and GAL10 promoter strength, consistent with experimental results (see Fig. 2.6A). However, noise behaves differently between these two models. The promoter model has noise peaks directly proportional to GAL10 noise. In contrast, the sequestration model displays lower noise peaks as the mutations become more severe, consistent with experimental observations (see Fig. 2.5).

These differences in behavior are explained by free TetR noise. In contrast to the promoter model, free TetR in the sequestration model decreases due to ATc sequestration, causing free TetR noise to generally increase with ATc (see Fig. 2.6B).

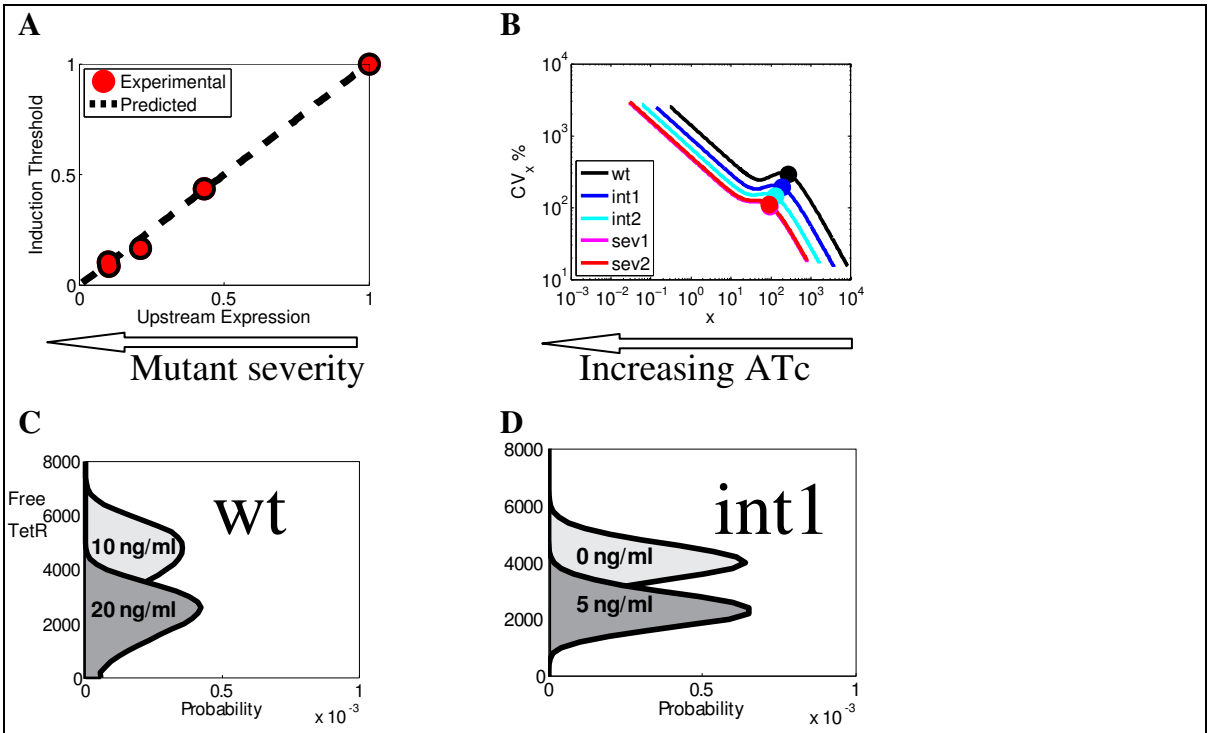


Figure 2.6. Implications of sequestration model

A) Both the promoter and sequestration models predict a linear relationship between TetR mean expression levels and induction thresholds (the ATc concentration where yEGFP expression is half of its maximum value). B) The sequestration model says that as free TetR (x) is depleted by ATc, free TetR noise will generally increase. A decrease in noise occurs when ATc saturates the cell and free TetR loses sensitivity to ATc concentrations. This increase in noise was also observed in Gillespie simulations of the template model. Solid circles correspond to peak yEGFP noise. WT C) and int1 D) promoter simulations are shown at similar means to illustrate this increase in free TetR noise.

Although all 5 strains had similar upstream noise levels, more ATc was required to eliminate repression of the stronger promoters (Fig. 2.6C,D). The increased noise associated with higher induction thresholds (based on the sequestration model) explains why noise increased with TetR promoter strength.

Although these two models make different predictions, they do not imply incompatible biological properties. Instead, these results suggest that whereas ATc may act on promoter bound TetR, this reaction is less significant to downstream regulation than the binding of free floating TetR to ATc.

2.4 Discussion

Previous work has modeled TetR as binding in an extremely cooperative manner to DNA (52,91,92,94,96,97). This cooperativity resulted in a steep downstream response to TetR, and explained corresponding noise peaks. The "promoter model", which also assumed cooperative TetR-DNA binding, predicted that downstream noise would be proportional to TetR noise as long as long as the downstream promoter did not change.

I tested an alternative hypothesis, based on the idea that protein sequestration could cause a genetic network to display an ultrasensitive response to an inducer (105,106). In this "sequestration model", I assumed that the extreme sensitivity of downstream expression was due to sequestration of active TetR by ATc. In this model, only a small number of TetR molecules were needed to repress downstream gene expression. As active TetR became depleted by ATc, repression was quickly and suddenly lifted, resulting in a steep dose-response. A more reductive version of the sequestration model was also tested, but had extreme noise values precluding it from serious consideration (see section 6.7).

Both the promoter and sequestration models predict a linear relationship between TetR expression and the ATc concentration where repression is relieved (see Fig. 2.6A and see Fig. 2.5 for the dose-responses). However these two models give very different predictions for gene expression noise (see Fig. 2.5). The promoter model assumes TetR noise remains constant regardless of ATc concentrations. In contrast, the sequestration model says that *free* TetR noise generally increases as TetR binds to ATc (since fewer molecules imply higher noise, see Fig. 2.6B). The stronger the TetR promoter, the more ATc molecules are required to relieve repression, and the more noise will increase. This increase is shown for the Gillespie simulations of WT-T123 and int1-T123 cells (see Fig. 2.6C and D).

The differences between the two models arise from the effects of cell growth on thermodynamic equilibrium. The promoter model assumes that ATc concentrations within a cell have reached thermodynamic equilibrium with the environment so that their concentrations are identical (see Fig. 2.7A). On the other hand, the sequestration model assumes that ATc concentrations within a cell are in thermodynamic non-equilibrium (115) with the environment due to cell growth (see Hypothesis, section 1.5). The effect of this thermodynamic non-equilibrium is that ATc mostly diffuses *into* cells, binds to TetR molecules, and is then diluted out due to cell growth and cell division (see Fig. 2.7B). This continuous flux of ATc from the environment into daughter cells forms an open system. As long as ATc is sequestered and diluted out of the cell, TetR repression remains very high. As ATc influx overwhelms TetR sequestration, and TetR repression of *yEGFP* expression will become extremely sensitive to changes in ATc.

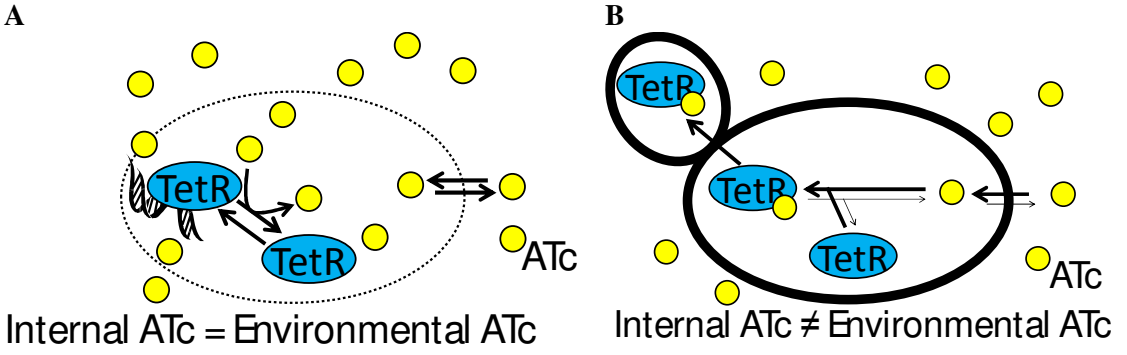


Figure 2.7. Cell growth and TetR sequestration prevent ATc from reaching equilibrium

A) The promoter model assumes that free ATc concentrations within a cell are identical to environmental concentrations. B) The sequestration model assumes that ATc concentrations should be treated as an open system due to cell growth. ATc tends to diffuse into the cell, where it is bound to TetR and diluted out of the cell. As ATc concentrations become similar to TetR concentrations, downstream repression is quickly relieved and yEGFP is expressed.

These results have important implications for working with and designing new gene expression systems. Because constitutive *tetR* expression systems are widely used to study genes of interest, as well create synthetic gene expression systems, understanding *tetR* dynamics have broad implications. Understanding how mean TetR gene expression and TetR noise affect downstream noise could allow new gene expression circuits to be designed in a modular and robust manner (as shown in subsequent chapters). Knowing whether TetR sensitivity originates from ATc sequestration or cooperative promoter binding could allow larger gene networks to be built with a better understanding of how noise propagates through them.

Chapter 3 *tetR* autoregulatory sensitivity to promoter architecture

3.1 Objectives

The negative autoregulation motif can significantly affect gene expression, resulting in homeostatic gene expression and faster responsiveness. These properties could be very useful in synthetic gene expression systems, allowing investigators to very precisely measure the effects of gene expression on a cell's phenotype. In this section I investigate how placing the tetracycline repressor *tetR* under its own control changes the overall behavior of gene expression. My objective was to identify how modifications to the *tetR* and target gene promoters could affect gene expression properties. This section focuses on the computational work done in Adams et al. 2009 (110). All experimental work in this chapter was performed by Dmitry Nevozhay.

3.2 Methods

3.2.1 Yeast strains analyzed

I analyzed yeast strains constructed by Dmitry Nevozhay (110). Negative Regulation (NR) cells were created where *tetR* was constitutively expressed by the GAL1 promoter, and *yEGFP* was expressed by the variant of the GAL1 promoter D12 (92). The D12 promoter had two *tetO2* binding sites placed between the TATA box and the *yEGFP* gene, allowing *tetR* to repress *yEGFP*.

A second Negative Feedback (NF) strain was created, where both *tetR* and *yEGFP* were placed under the control of the D12 promoter. Additional negative feedback strains were created by changing the number of *tetO2* binding sites between the TATA box and the promoter. These variant promoters were S1 (1 *tetO2* binding site) and T123 (3 *tetO2* binding site). These variants were named by the convention <tetR promoter> - <yEGFP promoter>. The 5 variants created are S1-S1, D12-S1, D12-D12 (also referred to as NF), D12-T123, and T123-T123. All genes were chromosomally integrated.

These promoters were activated by incubating cells in galactose containing media. TetR repression was weakened by adding ATc to the medium.

3.2.2 Flow cytometry analysis

yEGFP fluorescence in cells was measured by flow cytometry. Cells were filtered by their forward and side scatter values using a rectangular gate. At least 5,000 cells were retained from 50,000 cells after gating.

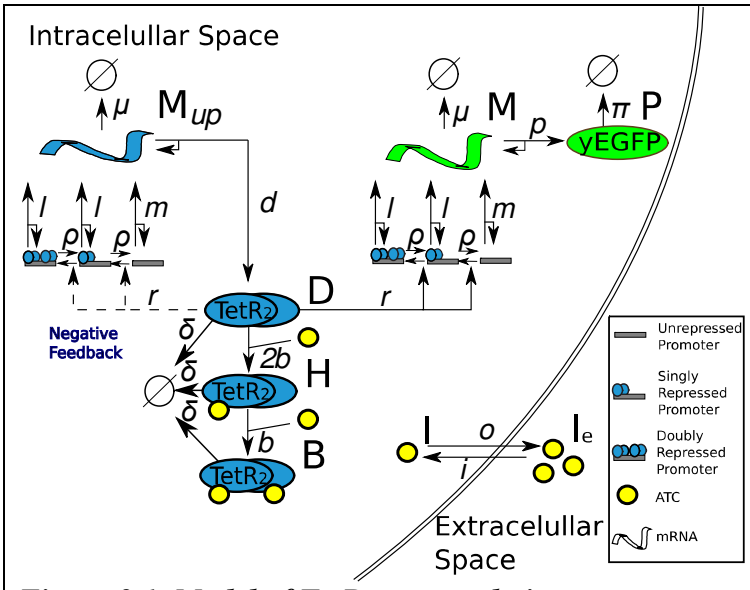


Figure 3.1. Model of TetR autoregulation

The template model of TetR autoregulation. This schematic is similar to the one used for constitutive TetR expression (see Fig. 2.1), except that TetR is capable of blocking its own expression, and promoters have only two tetO2 binding sites.

3.2.3 Gillespie simulation of TetR models

I modified the previously defined Template model for constitutively expressed *tetR* gene expression systems (see Fig. 2.1) to have two *tetO2* binding sites controlling the *yEGFP* promoter for NR cells, and identical up- and down-stream promoters for NF cells (see Fig. 3.1.). Gillespie simulations of the negative feedback Template model were performed from t = 0 to 10,000 hours at various levels of ATc induction (Dizzy code is given in sections 6.8 and 6.9, parameters are described in section 6.10). The first 10 hours were excluded from the calculation of mean and standard deviation of yEGFP to allow the distributions to relax from their starting conditions.

3.2.4 The simplified negative feedback model

I simplified the Template model to a set of reactions (similar to the previously defined sequestration model, [35] where TetR was capable of repressing both *tetR* and *yEGFP* promoters. The reaction scheme is:

$$\begin{aligned}
\text{TetR production:} &\xrightarrow{a_x F_x(x)+l_x} x \\
\text{TetR dilution:} \; x &\xrightarrow{g} \varnothing \\
\text{ATc influx:} &\xrightarrow{C} y \\
\text{ATc outflux/degradation:} \; y &\xrightarrow{h} \varnothing \\
\text{TetR ATc binding:} \; x + y &\xrightarrow{b} \varnothing \\
\text{yEGFP production:} &\xrightarrow{a_z F_z(x)+l_z} z \\
\text{yEGFP dilution:} \; z &\xrightarrow{g} \varnothing
\end{aligned} , \qquad [36]$$

where the molecules x, y, and z correspond to free intracellular repressor, inducer, and reporter concentration, respectively, and C is the experimentally controlled parameter proportional to extracellular inducer concentration. The rate constants are a_x (TetR synthesis rate), a_z (yEGFP synthesis rate), b (ATc–TetR association rate), g (dilution rate due to cell growth), h (combined rate of inducer dilution, outflux, and degradation), l_x (TetR basal synthesis rate), and l_z (yEGFP basal synthesis rate). The functions F_x and F_z describe the relationship between free repressor concentrations and protein synthesis for the upstream and downstream promoters, respectively.

I used a phenomenological Hill function to approximate the relationship between repressor concentration and promoter activation as:

$$F(x) = \frac{\theta^n}{\theta^n + x^n}, \qquad [37]$$

where θ is the dissociation constant of TetR to *tetO2* sites, and n is the Hill coefficient. For identical upstream promoters $F_x = F_z$, for constitutively expressed upstream promoters $F_x = 1$. The drift, diffusion, Jacobian, and steady state matrices and the parameters used to obtain the Linear Noise Approximation are given in section 6.11.

3.2.5 Measuring distances between promoters and dose-responses

I used the L_1 norm to measure the distance between two dose-responses as:

$$L_1(F, G) = \int_0^{\varphi} \left| F(x(C)) - G(x(C)) \right| dC, \qquad [38]$$

where F and G are the promoter responses to free TetR as a function of ATc, from $C = 0$ to the cutoff value φ.

The L_1 norm was used to measure the differences between promoter responses and to measure the linearity of dose-responses.

3.3 Results

3.3.1 Construction of negative regulatory and negative autoregulatory strains

A yeast strain with a negative-regulatory (NR) gene circuit consisting of chromosomally integrated, separate regulator and reporter parts (see Fig. 3.2.) was built by Dmitry Nevozhay. The regulator part of this cascade, the TetR repressor, was expressed constitutively from the native yeast *GAL1* promoter (P_{GAL1}) when grown in galactose media. The repressory DNA-binding activity of TetR was controlled by modifying the extracellular concentration of anhydrotetracycline (ATc) which binds to TetR, and subsequently prevents TetR repressing its target promoters. The reporter part of the NR cascade consisted of the *yEGFP* gene under the control of the $P_{GAL1\text{-}D12}$ promoter, which is repressed as TetR molecules bind to 2 *tetO2* sites downstream from the TATA box.

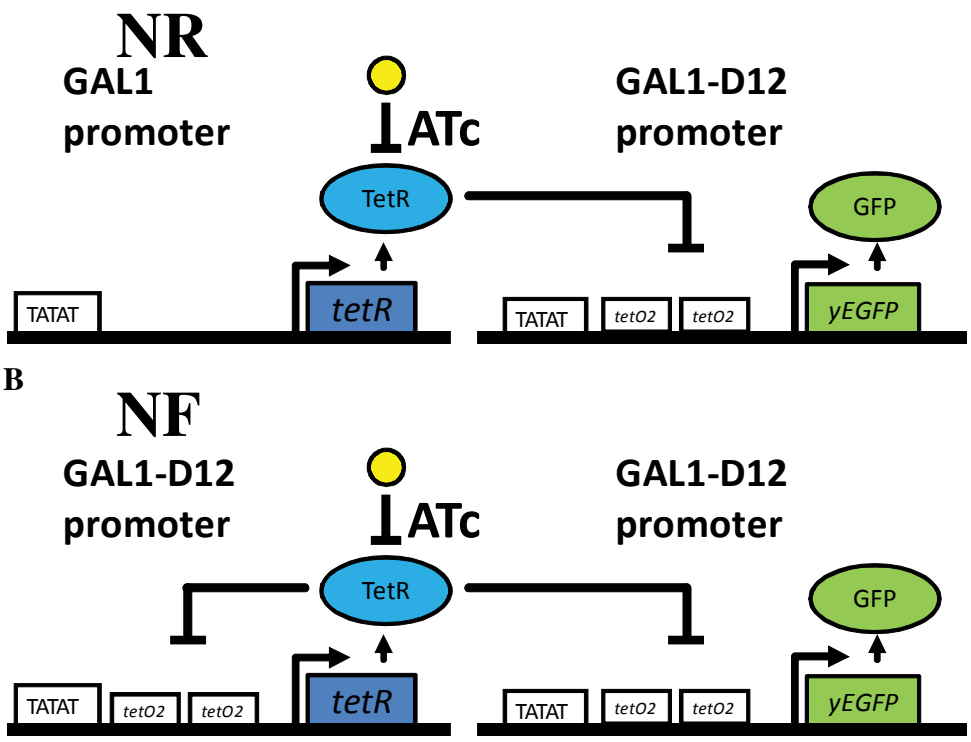


Figure 3.2. Experimental design

Promoter schematic of cell strains being studied. A) NR cells expressed tetR constitutively from the GAL1 promoter, binding to tetO2 sites in front of the yEGFP gene, blocking gene expression. B) NF cells are similar to NR cells; with the addition that tetR binds to tetO2 sites in its own promoter, blocking its own expression levels.

NR strains had similar behavior to previously developed strains (i.e., GAL10-T123 described in Chapter 2), with an ultrasensitive response to ATc (the Hill coefficient is $n \approx 11$), when the dose-response is fit to a Hill function), and a peak in gene expression noise occurring at the point where yEGFP repression was slightly relieved.

The NR dose-response was modeled by the simplified Template model (see section 3.2.3). The ultrasensitive response was explained by the nearly irreversible binding rates of TetR to ATc which then diluted out of the cell through growth. These behaviors, along with reaction rates, are supported by the experimental literature and provide an alternative explanation to ultrasensitivity from extremely cooperative DNA binding (see Fig. 3.3.A,B) (105,106).

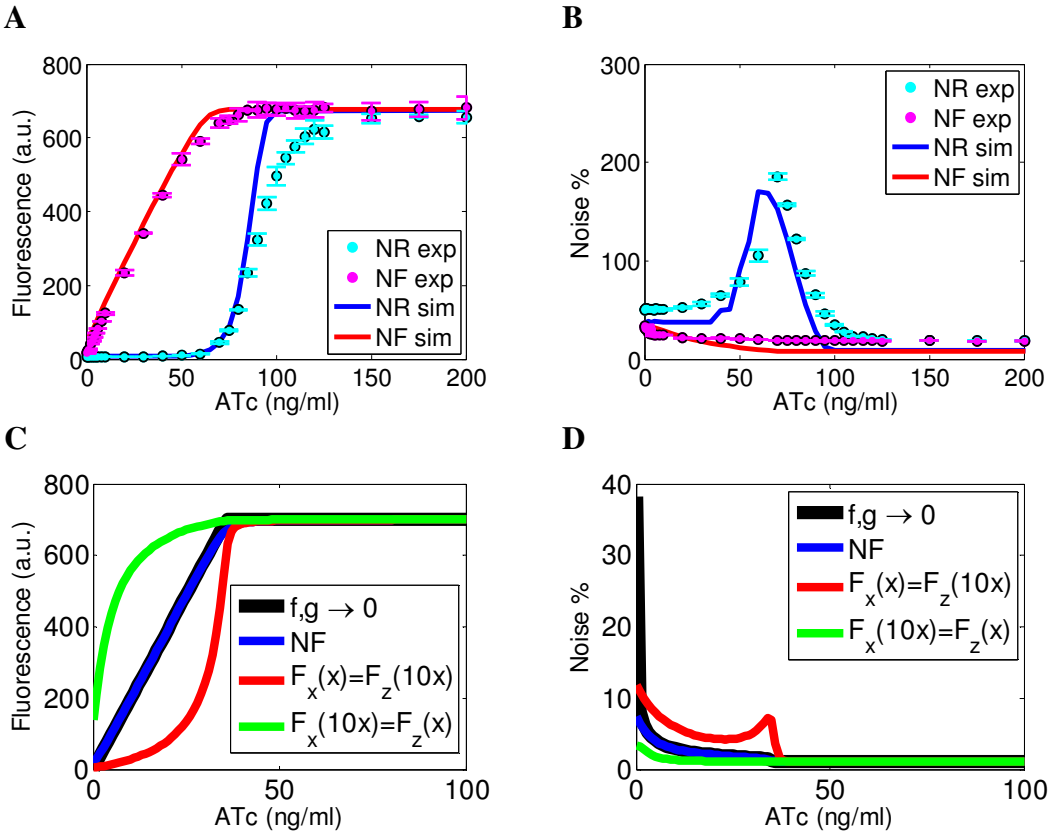


Figure 3.3. Negative autoregulation linearizes the dose-response and reduces noise

Dose-responses of NR and NF strain fluorescence means (A) and noise levels (B). Experimental results measured after overnight incubation are shown by circles, Gillespie simulations of the template model are shown by solid lines. Based on the simplified model, NF cells will display a linear dose-response up to saturation when TetR degradation is low, ATc outflux is low, and downstream promoters are identical to upstream promoters (C). Along with linearization, sensitivity to ATc is decreased, eliminating the characteristic noise peak found in NR cells (D). Subsequent measurements show that the dose-response becomes slightly steeper after 24 hours. Dizzy code for the template models may be found in sections 6.8 and 6.9 for NR and NF strains, respectively. Parameters for the template model are given in section 6.10. Parameters for the sequestration model are given in section 6.11.

Negative feedback (NF) strains were also created by placing the *tetR* gene under the control of $P_{GAL1\text{-}D12}$ promoter. Consistent with Template models, NF showed a linear dose-response to ATc, up to 90% saturation, demonstrating a novel property of negative feedback gene expression systems (see Fig. 3.3A).

3.3.2 Linear dose-response of GFP to ATc

The deterministic formulation and Linear Noise Approximation of the simplified model [35], see section 3.2.4) were used to analyze the NF dose-response (see Fig. 3.3C, D, section 6.11 for parameters). The deterministic formulation of the simplified model is

$$\begin{aligned}\dot{x} &= a_x F_x(x) + l_x - bxy - gx \\ \dot{y} &= C - bxy - hy \\ \dot{z} &= a_z F_z(x) + l_z - gz\end{aligned} \quad . \qquad [39]$$

Under certain conditions, the steady state solution of [39] converges to a linear relationship between *yEGFP* expression and ATc induction. If ATc is unable to leave the cell, either due to slow diffusion or strong TetR binding ($hy << bxy$), and if unbound TetR has a higher affinity to "leave" the cell through ATc sequestration than dilution ($gx << bxy$), then yEGFP expression will respond to ATc concentrations in a linear manner (see Fig. 3.3.C, D).

As $gx \rightarrow 0$, and $hy \rightarrow 0$ then the relationship between free TetR and inducer becomes

$$x = F_x^{-1}\left(\frac{C}{a}\right). \qquad [40]$$

Here F_x^{-1} is the inverse function of F_x, suggesting that free TetR (x) is a (pre-)distorted function of ATc (C). The steady state solution for yEGFP is then

$$z = \frac{a_z}{g} F_z(x) = \frac{a_z}{g} F_z\left(F_x^{-1}\left(\frac{C}{a_x}\right)\right), \qquad [41]$$

where yEGFP (z) is a distorted function of a predistorted function of ATc (C) concentrations. Therefore, when F_x and F_z are related by the linear transformation

$$F_z[x(C)] = sF_x[x(C+\varphi)] + l, \qquad [42]$$

the dose-response converges to the linear relationship

$$z = \frac{s}{g}C + \frac{s\varphi + l}{g}, \qquad [43]$$

where the factors s, l, and φ stretch and translate F_x vertically and horizontally (for identical promoters, $s = a_z/a_x \rightarrow 1$, $l \rightarrow 0$, and $\varphi \rightarrow 0$, resulting in a linear dose-response).

This result [43] predicts that if the upstream and downstream TetR controlled promoters responses are identical, *or* related by a linear transformation [42], then yEGFP will increase linearly with ATc induction.

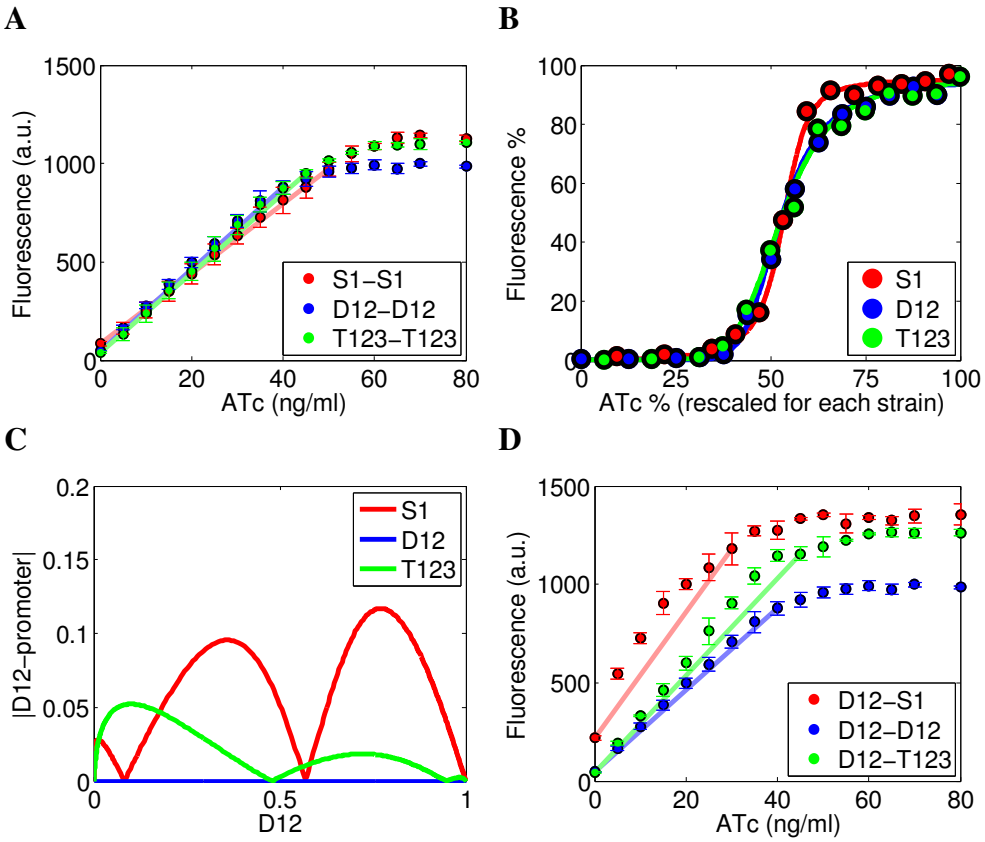


Figure 3.4. Linear transformations of promoters imply linear dose-responses

A) Cells with identical tetR and yEGFP promoters displayed linear dose-responses up to 90% saturation. B) In order to determine how much different promoter combinations would distort the linearizer dose-response, I performed linear transformations of GAL10-S1 and GAL10-T123 cells in such a way that their L_1 norm distances to the GAL10-D12 dose-response were minimized. C) The areas under the curves visually represent the L_1 norm distances between the transformed dose-responses and the D12 promoter. Cells with the T123 promoter were closer to the D12 promoter than cells with the S1 promoter. D) Cells with non-identical promoters lost their linear dose-response, with the severity of this loss corresponding to the L_1 norm distances. Light straight lines are from 0 ATc to 90% of maximum fluorescence.

3.3.3 The simplified model predicts linearity

In order to test the prediction that cells with identical upstream and downstream promoters (from [42], when $s \to 1$, $\varphi \to 0$, $l \to 0$) would have a linear dose-response, Dmitry Nevozhay created additional negative feedback strains by changing the number of *tetO2* binding sites of up- and down-stream promoters. These strains were created by using the previously developed promoters; S1 (1 *tetO2* site), D12 (2 *tetO2* sites), and T123 (3 *tetO2* sites) (92). Cell strains were named according to the convention <upstream promoter>-<downstream promoter>. For example D12-S1 refers to strains with TetR controlled by the D12 promoter, and yEGFP controlled by the S1 promoter.

The constructed strains S1-S1, D12-D12, and T123-T123 confirmed my predictions, displaying linearity (see Fig. 3.4.A). These identical promoter strains demonstrated a linear dose-response to ATc with identical slopes ($L_1 = 2.1 \times 10^{-2}$, 5.7×10^{-2}, and 3.0×10^{-2}, respectively).

To determine whether non-identical promoters would be approximately linear transformations of each other, I used the L_1-norm to measure the difference between two promoters. I scaled and transformed the dose-responses of NR analogues containing the S1, D12, and T123 downstream promoters such that their L_1-norm distances were minimized (Fig. 3.4.B). These dose-responses were taken from previously measured data (92), and had GAL10 promoters expressing TetR instead of the GAL1 promoter used in NR (denoted P_{GAL10}-S1, P_{GAL10}-D12, P_{GAL10}-T123).

The rescaled dose-responses had a distance of $L_1 = 0.0632$ between P_{GAL10}-S1 and P_{GAL10}-D12 cells, and a distance of $L_1 = 0.0209$ between P_{GAL10}-T123 and P_{GAL10}-D12 cells. Thus, the D12-S1 circuit was expected to have a more severe reduction in linearity than the D12-T123 circuit (Fig. 3.4.C). Additionally, the slope of the D12-T123 dose-response is determined by the ratio of upstream and downstream promoter strengths (i.e., $s = a_z/a_x$ [41] and [43]). Because the T123 promoter was previously measured to have higher fully induced expression than D12, the D12-T123 dose-response was predicted to be 1.28× steeper than D12-D12.

Consistent with predictions, D12-S1 cells had high curvature ($L_1 = 14.7 \times 10^{-2}$), while the linear dose-response of D12-T123 was less distorted (L_1-norm $= 7.9 \times 10^{-2}$) with a steeper dose-response of 1.32 times the D12-D12 dose-response, similar to the predicted increase of 1.28.

3.4 Discussion

The addition of a negative feedback loop to a previously used *tetR* based gene expression system resulted in a dramatically different dose-response. Whereas evidence had accumulated that negative feedback could minimize the effects of external noise sources (95), with delays negative feedback can result in a destabilizing oscillatory behavior (116,117), and can increase noise (118). Thus, it was not entirely obvious if negative feedback would act to reduce or increase noise (119). The mathematical model showed that the noise decrease arose in tandem with the reduced sensitivity of the downstream

gene expression to induction. Thus reduced downstream sensitivity insulated NF cells from upstream noise (i.e., from TetR and ATc noise).

Although a previous gene regulatory model predicted a linearization of the dose-response with negative autoregulation, this model had different behaviors from those observed experimentally. It was predicted that the dose-response with negative autoregulation would retain a sigmoidal shape, but the dose-response would become more linear when repression was relieved (120). This dose-response was predicted for systems when environmental inducer was equal to internal inducer levels. Previous work does not support this assumption (see Chapter 2), and the experimental dose-response did not retain its sigmoidal dose-response with the addition of negative autoregulation.

Importantly, computational models were used to design new linearizer gene expression systems. This behavior was unprecedented in synthetic eukaryotic gene expression systems, and could allow investigators to design powerful new gene expression systems. Although to the best of my knowledge, no systematic study of linearization of gene expression had been undertaken, examples have been known and applied in electronics, control theory, and neuroscience. For example, negative feedback can be used to linearly amplify electronic amplifiers, and may convert nonlinear instantaneous firing rate responses to linear responses at steady state (121).

A set of gene constructs analogous to the NR and NF strains were studied in *Escherichia coli*, with the NF analogue also reducing noise over a broad range of inducer concentrations (95). This reduction in noise was explained by negative feedback acting to reduce the extrinsic noise from plasmid copy variation. However, the yeast strains modeled here had single-copy, chromosomally integrated constructs, eliminating this source of extrinsic noise. Although some of the reduction of noise in NF cells may be due to reduced extrinsic noise, much of the noise reduction observed in NF cells can be attributed to the reduced sensitivity of NF to ATc induction (as opposed to NR's extreme response to ATc).

Chapter 4 *rtTA* autoregulatory system noise and memory

4.1 Objectives

Under certain situations, gene expression noise may be advantageous. A population of cells with high gene expression noise will have a wide range of phenotypes, which can potentially allow the population to survive in multiple environments. In a slowly changing (or static) environment, slowly changing gene expression can allow sub-populations of cells to display fit phenotypes long enough to survive an otherwise toxic environment. My objective in this section is to understand the connection between cell fitness and the temporal effects of noise. To do this, I ask how noise shapes the population distribution, and how both temporal noise properties and individual cell fitness interact to determine an isogenic population's overall fitness. This chapter largely follows the work done in Adams et al 2012 (122). All experimental work in this chapter was performed by Dmitry Nevozhay.

4.2 Methods

4.2.1 Yeast strains analyzed

I analyzed cells carrying a positive feedback gene circuit (PF) created by Dmitry Nevozhay. In PF cells, a modified version of the rtTA activator (rtTA-MF) capable of activating its own expression was chromosomally integrated into *S. cerevisiae* (123). Concurrently, a fusion of the fluorescent reporter *yEGFP* and antibiotic resistance gene *zeoR* (*yEGFP::zeoR*) was placed under the control of the rtTA activator. Both *rtTA* and *yEGFP::zeoR* promoters were activated by rtTA binding to the synthetic P_{TETREG} promoters (Fig. 2A) (100). rtTA promoter activation was increased by adding anhydrotetracycline (ATc) to the cell growth medium.

4.2.2 Flow cytometry analysis

yEGFP::ZeoR fluorescence in PF cells was measured by flow cytometry. Cells were loosely gated by their forward and side scatter values to eliminate debris. The gating used was an ellipse. At least 25,000 cells were retained, and usually 30,000 cells were retained from 50,000 total cells.

I classified a cell population as having bimodal expression if the fluorescence histogram had a local minimum between two peaks. Because variations in the numbers of cells classified within a bin are subject to noise, I used a statistical test to determine whether a local minimum was significant.

Gated fluorescence intensity histograms were first smoothed with a 32-point moving average to estimate the expected number of events for each fluorescence bin. Because the number of events for a bin is approximately Poissonian (124,125), I expected the variance for the number of events within each bin to be equal to the average number of events. Furthermore, since the *average* number of events is obtained by smoothing, the Central Limit Theorem states that this smoothed data should

tend towards a normal distribution whose standard deviation is equal to the number of cells detected, divided by the square root of the number of points used for smoothing (i.e., $\sqrt{32}$).

Given two local maxima of heights h_L and h_R, located to the (*L*)eft and (*R*)ight of a local (*M*)inimum of height h_M, I rejected the null hypothesis that the distribution was unimodal if h_M was at least 4 standard deviations lower than both h_L and h_R,

$$\begin{aligned} &\left|h_M - h_L\right| > 4\sigma_L = 4\sqrt{h_L/32} \Leftrightarrow z_L > 4 \\ &AND \\ &\left|h_M - h_R\right| > 4\sigma_R = 4\sqrt{h_R/32} \Leftrightarrow z_R > 4 \end{aligned} \quad . \qquad [44]$$

If multiple sets of local minimum /maximum pairs were found, I chose the pairs that maximized the relationship:

$$D = \ln\left|h_M - h_L\right| + \ln\left|h_M - h_R\right| . \qquad [45]$$

4.2.3 Measuring cell division rates

Cells were resuspended in fresh media containing 2% galactose, and various ATc and Zeocin concentrations every 12 hours. Cells were counted by the NexCelom Cellometer T4 cell counter. Because cell numbers increased exponentially, division rates were calculated from the slope of log-transformed cell counts with respect to time. I used bootstrapping to minimize the effect of outliers on cell division rates. Linear regression was performed on all possible combinations of 3 of the 6 time points, and averaged to give the overall cell division rate.

4.2.4 Modeling overall cell population fitness

I modeled *overall cell population fitness (g_T)* according to the relationship

$$g_T = g_0 \int_0^\infty \gamma(F) p(F,t) dF , \qquad [46]$$

where g_0 is the maximum division rate of PF cells, $\gamma(F)$ is the instantaneous fitness reduction of being in the F state, and $p(F,t)$ is the probability distribution function of cells in the state F at time t. In general, I only analyzed cell populations with stationary probability distributions $\left(p(F,t) \to \pi(F)\right)$, so that overall cell population fitness did not change over time. The product

$$\Gamma(F) = g_0 \gamma(F), \qquad [47]$$

is the *instantaneous fitness* and is assumed to be independent of time. The overall cell population fitness (g_T) describes the change in the total number of cells (N_T) as:

$$\dot{N}_T = g_T N_T . \quad [48]$$

The average fitness over an interval of cell with stationary distributions is denoted

$$g(\varphi_1,\varphi_2,t) = g_0 \frac{\int_{\varphi_1}^{\varphi_2} \gamma(F)p(F,t)dF}{\int_{\varphi_1}^{\varphi_2} p(F,t)dF} . \quad [49]$$

Again, since I studied cell populations with stationary distributions, the average fitness of cells over an interval was generally assumed to be constant over time.

4.2.5 Modeling the cost of rtTA expression

Consistent with previous observations, rtTA reduces cell division rates. Previous papers have suggested that toxicity occurs when rtTA sequesters important transcription factors, preventing them from performing their usual activities (126,127). Based on this, I created the model of rtTA kinetics with the following reactions:

$$\begin{aligned} \text{TF binding} &: x + S \longrightarrow xS \\ \text{TF unbinding} &: xS \longrightarrow x + S \\ \text{rtTA activation} &: w + C \longrightarrow x \\ \text{rtTA inactivation} &: x \longrightarrow w + C \end{aligned} . \quad [50]$$

Here, w and x are inactive (no ATc bound) and active (ATc bound) rtTA, S is a transcription factor directly or indirectly responsible for cell division, C is ATc in the media, and F is the fluorescence of the cell from yEGFP::ZeoR. Because *rtTA* and yEGFP::ZeoR have identical promoters, total rtTA is assumed to be correlated to fluorescence,

$$F \propto x + w . \quad [51]$$

Cell division rates were assumed to be proportional to the transcription factor being squelched. Using basic mass action kinetics at steady state, I created the equation

$$\gamma_x \propto S = \frac{\alpha}{\alpha + x} \rightarrow \gamma_x \propto \frac{\alpha}{\alpha + F\dfrac{C}{C+\beta}} \quad [52]$$

where γ_x is the reduction in cell division due to active rtTA activity. I fit the parameters as α = 936 M, β = 5.8 (ng/ml) h^{-1}, C = [ATc] h^{-1}, F = (fluorescence) × M $(a.u.)^{-1}$.

4.2.6 Modeling drug resistance

Zeocin interferes with cell division by causing double stranded breaks along DNA. Consequently, I modeled the effects of Zeocin by a two state model of DNA where undamaged DNA (D_u) transitions to a damaged state (D_d) due to Zeocin (Z_i), and damaged DNA is constantly repaired. This is given by the scheme:

$$\begin{aligned} &\text{DNA damage} : D_u \xrightarrow{f(Z_i)} D_d \\ &\text{DNA repair} : D_d \xrightarrow{r} D_u \end{aligned} . \qquad [53]$$

I assumed that DNA damage repair occurred at a constant, faster rate than cell division time, so that the instantaneous fitness reduction was proportional to the fraction of time that DNA remained undamaged. The instantaneous fitness reduction due to DNA damage was defined by the first order Hill function:

$$\gamma_Z \propto D_u = \frac{\chi}{Z_i + \chi}, \qquad [54]$$

where Z_i is the intracellular Zeocin concentration dependent on the extracellular Zeocin concentration (Z) and χ is the rate of DNA repair divided by the rate of Zeocin-induced damage accumulation.

yEGFP::ZeoR, a fusion of the yEGFP protein to the ZeoR protein, provides protection from Zeocin by binding to it and sequestering it (128,129). In order to infer Z_i, Zeocin was assumed to diffuse into and out of the cell, and become harmless when bound to yEGFP::ZeoR. Intracellular Zeocin concentration was modeled as

$$\begin{aligned} &\text{Zeocin binding} : F_u + Z_i \xrightarrow{s} F_b \\ &\text{Zeocin influx} : \xrightarrow{Z} Z_i \\ &\text{Zeocin outflux} : Z_i \xrightarrow{h_z} \varnothing \end{aligned} \qquad [55]$$

yielding the deterministic chemical equations:

$$\begin{aligned} \dot{Z}_i &= Z - h_Z Z_i - s F_u Z_i \\ \dot{F}_b &= s F_u Z_i - d F_b \end{aligned}, \qquad [56]$$

where Z_i and Z are internal and external Zeocin, and F_b and F_u are yEGFP::ZeoR bound and unbound to Zeocin. The total fluorescence (F) imposes the constraint $F = F_u + F_b$. The rate constants are h_Z (Zeocin diffusion out of the cell membrane), s (yEGFP::ZeoR binding rate for Zeocin), and d (yEGFP::ZeoR degradation/dilution rate, assumed to be constant for simplicity). Assuming that free Zeocin equilibrates quickly, I obtained the steady state solution for intracellular Zeocin concentration from [56] as

$$Z_i = \frac{sZ - dh_Z - dsF + \sqrt{(sZ - dh_Z - dsF)^2 + 4h_Z sdZ}}{2h_Z s}\ , \qquad [57]$$

which was used in equation [54] to estimate growth retardation due to Zeocin concentration and yEGFP::ZeoR expression. I used the rate parameters d = 0.25 h^{-1}, h_z = 0.5 h^{-1},
s = 1.3×106 (mg/ml)$^{-1}$ h^{-1}, χ = 6.5×10^{-7} (mg/ml), Z = 1.182 × [Zeocin] h^{-1},
F = fluorescence × (mg/ml) (a.u.)$^{-1}$.

4.2.7 Estimating cellular memory

For cell states defined by fluorescence (or yEGFP) concentrations, the change in the number of cells over an interval of fluorescence concentrations (φ_1 to φ_2) can be summarized by the relationship:

$$\dot{N}(\varphi_1,\varphi_2,t) = \underbrace{I_{\rightarrow}(\varphi_1,t) - I_{\rightarrow}(\varphi_2,t)}_{\text{Increasing fluorescence}} - \underbrace{I_{\leftarrow}(\varphi_1,t) + I_{\leftarrow}(\varphi_2,t)}_{\text{Decreasing fluorescence}} + \underbrace{N(\varphi_1,\varphi_2,t)g(\varphi_1,\varphi_2,t)}_{\text{Fitness}}, \qquad [58]$$

where $N(\varphi_1,\varphi_2,t)$ is the number of cells with fluorescence φ_1 to φ_2 at time t, $I_{\rightarrow}$ and $I_{\leftarrow}$ are the currents of cells transitioning across a boundary due to increasing and decreasing fluorescence values respectively, and g is the average division rate of cells within the region (see [49]). The argument for modeling cellular current in this form is given in section 6.16. The number of cells is defined by the total number of cells (N_T) times the probability of cells being within the interval, [φ_1, φ_2], or

$$N(\varphi_1,\varphi_2,t) = N_T(t)\int_{\varphi_1}^{\varphi_2} p(F,t)dF\ . \qquad [59]$$

Because no cells can transition into (or out of) a negative fluorescence state, the cellular current at negative fluorescence is defined to be 0, ($I_{\rightarrow}(\varphi < 0,t) = I_{\leftarrow}(\varphi < 0,t) = 0$). Thus, the change in cells from 0 to φ is

$$\dot{N}_L(t) = \underbrace{-I_{\rightarrow}(\varphi,t)}_{\text{Increasing fluorescence}} + \underbrace{I_{\leftarrow}(\varphi,t)}_{\text{Decreasing fluorescence}} + \underbrace{N_L(t)g_L(t)}_{\text{Fitness}}, \qquad [60]$$

where $g_L(t) = g(0,\varphi,t)$, $N_L(t) = N(0,\varphi,t)$, and L denotes cells with low fluorescence concentrations.

The total number of cells increased in an approximately exponential rate since they were resuspended every 12 hours so that

$$N_T(t) = N_0 e^{g_T t}\ , \qquad [61]$$

where N_0 is the initial number of cells, and g_T is overall cell population fitness. Additionally, after ~2 days cells had stationary fluorescence distributions, implying (from [59] and [61]) that

$$\dot{N}_L(t) = \frac{\partial}{\partial t}\left[N_T(t) \int_0^{\varphi} \pi(F) dF \right] = g_T N_L(t). \quad [62]$$

where π is the stationary probability distribution of cells, independent of time.

The decreasing cellular current was defined as the rate at which fluorescence concentrations were lost times the density of cells ($\tilde{u}$) at the fluorescence boundary (φ), or

$$I_{\leftarrow}(\varphi,t) = \varphi\kappa(\varphi)\tilde{u}(\varphi,t) \quad [63]$$

where the growth rate of cells (κ) was assumed to be invariant over time, and equal to the division rates of cells, or $\kappa(\varphi) \to \Gamma(\varphi)$. The justification for describing the decreasing cellular current this way may be found in section 6.16. The density of cells was approximated from flow cytometry measurements as

$$\tilde{u}\left(\frac{F_1 + F_2}{2}, t \right) \approx \frac{N_B(F_1, F_2, t)}{|F_1 - F_2|} \quad [64]$$

where N_B is the number of cells measured in a flow cytometry bin over the interval F_1 to F_2 at time t.

Determining the increasing cellular current is more complicated, since yEGFP production is controlled by transcriptional activators, promoter states, transcriptional reactions, and translational reactions. However, assuming growth rates are invariant due to the stationary distributions of cells, substituting [62] into [60] and rearranging gives the solution to the increasing cell current as:

$$I_{\rightarrow}(\varphi,t) = g_L N_L(t) - g_T N_L(t) + I_{\leftarrow}(\varphi,t). \quad [65]$$

Thus, knowing the instantaneous fitness functions and stationary distributions allows both increasing and decreasing cellular currents to be determined.

The cellular memory was determined for low expressing cells as

$$\tau_L(\varphi) = \ln(2) \frac{N_L(t)}{I_{\rightarrow}(\varphi,t)} = \frac{\ln(2)}{r(\varphi)} \quad [66]$$

and high expressing cells as

$$\tau_H(\varphi) = \ln(2) \frac{N_H(t)}{I_{\leftarrow}(\varphi,t)} = \frac{\ln(2)}{f(\varphi)}, \quad [67]$$

where r is the fraction of cells that rise out of the (L)ow state per unit time, f is the fraction of cells that fall out of the (H)igh state per unit time, and N_H is the number of high expressor cells, $N_H = N(\varphi, \infty)$. Numerical simulations of the cellular current are given in section 6.17.

4.2.8 The 2-state model of population growth and phenotypic switching

The cellular current equation [60] may be rewritten so that it is analogous to the previously described two state model of population dynamics (21,25,26),

$$\begin{aligned} \dot{N}_L &= (g_L - r)N_L + fN_H \\ \dot{N}_H &= rN_L + (g_H - f)N_H \end{aligned}, \qquad [68]$$

where N_L and N_H are the number of (L)ow and (H)igh expressor cells, r is the rate that low expressors rise to the high expressor state, f is the rate that high expressors fall to the low expressor state, and g_L and g_H are the average division rates of the low and high expressor states. For cells sorted into distinct sub-populations, the distributions of the cell sub-populations were assumed not to change, so that g_L and g_H are constant over time.

This set of differential equations has the analytical ratio of low to high expressor cells given by

$$R(t) = \frac{\left(N_{L0} - N_{H0}\frac{f}{g_L - r - a_2}\right)\frac{f}{g_L - r - a_1}e^{a_1 t} + \left(N_{H0}\frac{f}{g_L - r - a_1} - N_{L0}\right)\frac{f}{g_L - r - a_2}e^{a_2 t}}{\left(N_{L0} - N_{H0}\frac{f}{g_L - r - a_2}\right)e^{a_1 t} + \left(N_{H0}\frac{f}{g_L - r - a_1} - N_{L0}\right)e^{a_2 t}}. \qquad [69]$$

with the corresponding eigenvalues

$$\begin{aligned} a_1 &= \frac{(g_L + g_H - r - f) + \sqrt{(g_L + g_H - r - f)^2 - 4((g_L - r)(g_H - f) - rf)}}{2} \\ a_2 &= \frac{(g_L + g_H - r - f) - \sqrt{(g_L + g_H - r - f)^2 - 4((g_L - r)(g_H - f) - rf)}}{2} \end{aligned}. \qquad [70]$$

The resulting log transformed ratio $\ln[R(t)] = \ln[N_L(t)/N_H(t)]$ was fit to experimentally observed log-ratios of low to high expressor subpopulations (see Fig. 4.4D and section 6.16). The predicted overall cell division rate after long periods of time (asymptotic cell division rate) is $g_T = a_1$. This overall cell division rate was used to predict cell fitness in different ATc and Zeocin environments.

4.2.9 Fitting phenomenological sub-population switching and growth rates

In order to determine how mutually dependent switching rates and growth rates affected population fitness in a changing environment, I created a set of phenomenological equations to

describe these rates as a function of induction. For the switching rates, r and f, I used the Hill functions

$$r = r_{\max} \frac{C^n}{\rho^n + C^n}, \quad [71]$$

and

$$f = f_{\max} \frac{\theta^m}{\theta^m + C^m} + f_{\min}, \quad [72]$$

where C is the ATc concentration in the medium.

The fit parameters for the rise and fall functions were $r_{\max} = 0.097$ h^{-1}, $\rho = 11$ ng ml^{-1}, $n = 2.1$, $f_{\max} = 0.029$ h^{-1}, $\theta = 0.39$ ng ml^{-1}, $m = 1.4$, and $f_{\min} = 0.0014$ h^{-1}.

The growth rates were modeled by the exponential function,

$$g = a + be^{-cC}. \quad [73]$$

The parameters for the growth function are given in table 4.1.

	a (h^{-1})	b (h^{-1})	c (ml ng^{-1})
Low expressors, Z = 0	0.2246	0.0026	0.1237
High expressors, Z = 0	0.1357	0.090	0.1686
Low expressors, Z = 2	0.0791	-0.028	0.1201
High expressors, Z = 0	0.1356	0.090	0.1738

Table 4.1 *Growth function parameters*

4.3 Results

4.3.1 A gene circuit with positive feedback displays bimodal gene expression

In order to better understand the temporal aspects of noisy gene expression, and their relationship to cell fitness, I studied cells with gene architectures consisting of a positive feedback loop controlling a drug resistance protein. Positive feedback loops tend to have slower dynamics, can exaggerate the temporal effects of noise, and are often involved in long-term cell decision making and differentiation (94,100,101). Thus, I studied a strain of cells constructed by Dmitry Nevozhay with a positive feedback (PF) architecture.

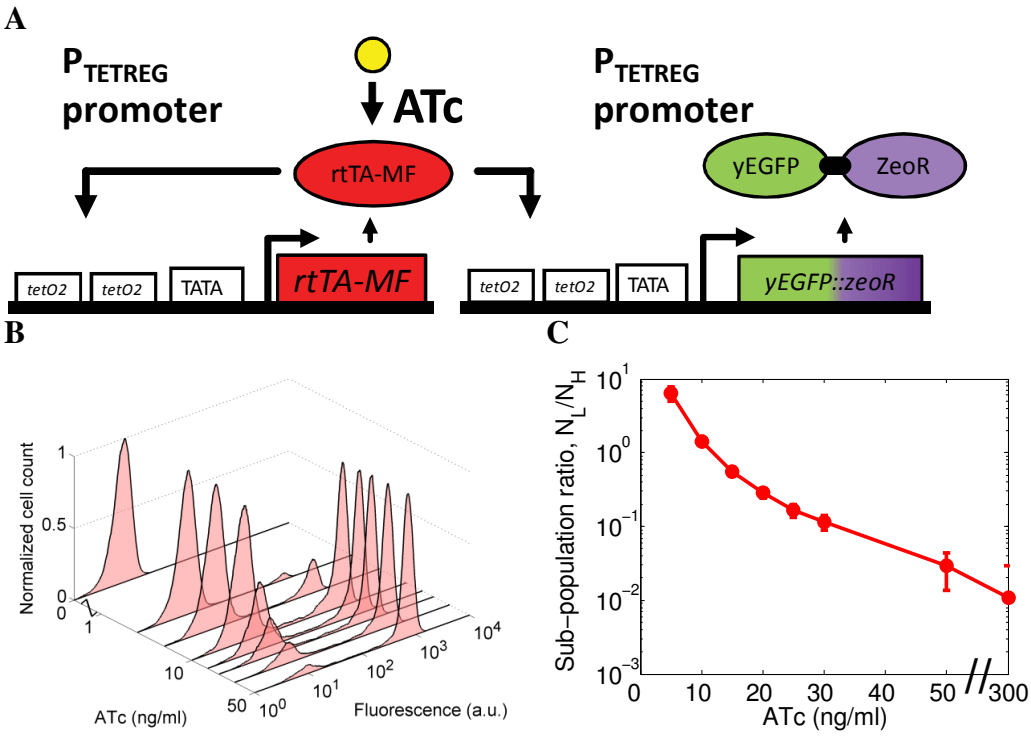


Figure 4.1. Schematic of rtTA autoregulation

A) Promoter schematic of PF cells. The fluorescence distributions of PF cells (B) and the ratio of low to high expressors (C) are plotted as a function of ATc.

In PF cells, a modified version (rtTA-MF) of the rtTA activator (see section 1.4) capable of activating its own expression was chromosomally integrated into yeast (123). Concurrently, a bifunctional fluorescent reporter and antibiotic resistance gene, *yEGFP::zeoR* was also placed under the control of the rtTA activator (see Fig. 4.1A). Both *rtTA* and *yEGFP::ZeoR* promoters were activated by rtTA binding to the synthetic P_{TETREG} promoters (Fig. 2A) (100). rtTA promoter activation was increased by adding anhydrotetracycline (ATc) to the cell growth media.

PF cells were grown in media containing increasing ATc concentrations, and resuspended in fresh media every 12 hours to maintain exponential cell division. *yEGFP*::*zeoR* expression of PF cells, measured by flow cytometry, was observed to stabilize after 2 days. After 2 days, stable dose-response distributions were established for yEGFP::ZeoR distributions (see Fig. 4.1B).

As previously observed for cells with rtTA autoregulatory loops, PF cells had unimodal distributions at low inducer concentrations, and bimodal gene expression at higher inducer concentrations (100). Using a custom algorithm (see section 4.2.2) to identify when and where bimodality occurred, cells were classified into low and high expressor sub-populations and their ratios were plotted as a function of ATc (see Fig. 4.1C). As induction was increased, the percentage of cells with high *yEGFP*::*zeoR* expression increased.

4.3.2 Inferring switching rates from probability distributions.

Systems capable of transitioning between low and high expressor states can be thought of as following the simplest isomerization reaction (75),

$$L \underset{f}{\overset{r}{\rightleftarrows}} H \,, \qquad [74]$$

where *L* and *H* are cells that have been classified as low or high expressors respectively, *r* is the rate at which *L* cells rise to the *H* state, and *f* is the rate that *H* cells fall to the *L* state. For this two state model, the ratio of cells with a stationary distribution determines the ratio of escape rates out of the states as

$$\frac{N_H}{N_L} = \frac{r}{f} \,, \qquad [75]$$

where N_H and N_L are the numbers of high and low expressor cells, and the escape rates *r* and *f* are the rates at which cells (*r*)ise out of the low state into the high state, and (*f*)all from the high state to the low state respectively (71,75).

PF cells grown at 10 ng/ml displayed a ~1:1 ratio of low to high expressor cells, implying that cells transitioned at equal rates between the two states ($f = r$). Thus, I hypothesized that cells separated into low and high populations would relax to a ~1:1 distribution at equal times. This was tested by Dmitry Nevozhay by growing PF cells in 10 ng/ml ATc until the population relaxed to a 1:1 ratio. After relaxation, these cells were separated into two clearly distinct populations using Fluorescence-activated cell sorting (FACS), and resuspended every 12 hours in media containing 10 ng/ml ATc. Surprisingly, these cells did not relax back to their original populations at equal rates. Instead, low expressor cells relaxed to their stationary distributions at ~45 hours, while high expressor cells did not completely relax to their stationary distributions even at 122 hours (see Fig. 4.2.).

These results support the hypothesis that other factors besides chemical kinetics are controlling the distribution of fluorescence. In order to account for this discrepancy, I hypothesized that different growth rates associated with cell phenotypes may be confounding chemical kinetic predictions.

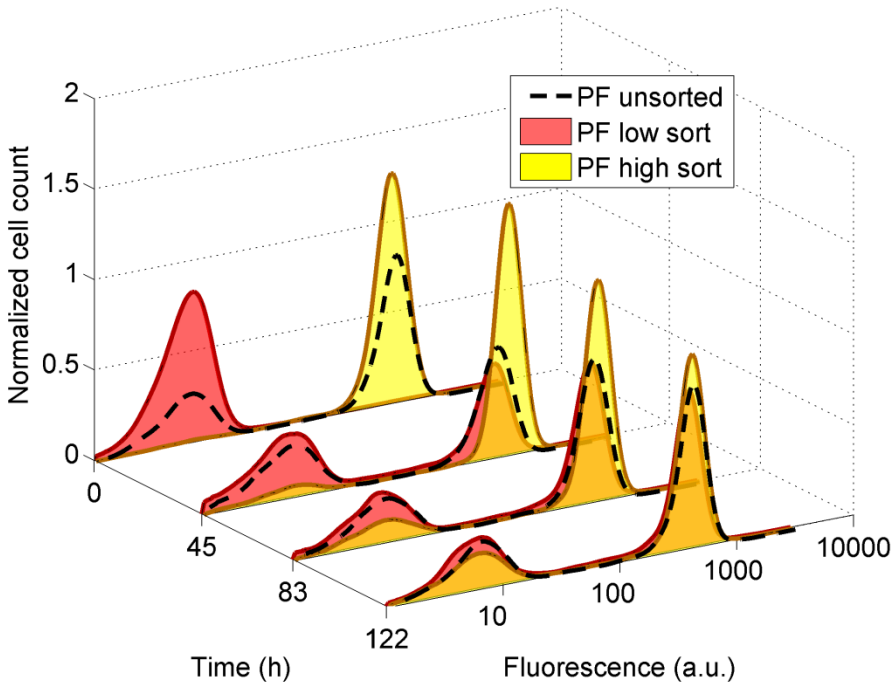


Figure 4.2. PF cell sorted distributions relax asymmetrically

Distributions of PF cells grown in 10 ng/ml ATc (unsorted cells), as well as PF cells sorted into low and high sub-populations over 122 hours at 5 representative time points. Low sorted cells relax to unsorted distributions after 45 hours, high sorted cells relax after 122 hours.

4.3.3 Introducing two different types of fitness.

Fitness (130) is a central concept in evolutionary theory that attempts to measure the contribution of a genotype to future generations. Because direct measurements are often difficult, the fitness of a genotype is often inferred from the fraction of the population containing the genotype (131). However, fitness can also be defined directly by the number of offspring that a population produces (130,132-134).

These definitions of fitness assume that differences are genetic – possibly changing phenotype and thus reproductive potential. However, PF cells are genetically identical, but *phenotypically* diverse. Non-genetic fitness can potentially change very quickly, highlighting an important distinction between genetic fitness and phenotypic fitness. Typically, genotypes can be classified into a definite number of states (e.g., dominant and recessive). However, gene expression levels are more accurately described by a continuum of chemical concentrations. In the case of *yEGFP*::*zeoR* expression, this continuum implies a gradient of drug resistance and fitness. To account for this difference, I introduced the concept of *instantaneous fitness* as the typical rate of cell division for cells within an

infinitesimally small chemical state(135,136). The fitness of the entire population of cells is the *overall cell population fitness*, obtained as the average instantaneous fitness (see section 4.2.4 for mathematical definitions) (133).

4.3.4 Defining the fitness of PF cells in the absence of antibiotic.

I asked whether there was a fitness cost associated with being in the high expressor state. This concern is justified by previous studies stating that rtTA retards cell growth, possibly by sequestering (squelching) transcription factors within the cell (127).

I created a dose-response curve for growth as a function of ATc (in the absence of Zeocin) based on cell count measurements obtained by Dmitry Nevozhay. Growth rates were inversely correlated with ATc induction and fluorescence, supporting the idea that high rtTA and/or yEGFP::ZeoR expression had reduced growth. Alternative gene expression systems based on TetR (NR, see Fig. 3.2.A) controlling yEGFP::ZeoR expression showed negligible growth retardation, supporting the hypothesis that *rtTA* expression was primarily responsible for reducing PF fitness (see Fig. 6.4., section 6.13).

Because growth rates decline as ATc is added to media, even when most cells are in the high state, I asked whether active rtTA was more toxic than inactive rtTA. Previous literature supports this (126). To further clarify this, Dmitry Nevozhay created an alternative strain with constitutively expressed rtTA (PR) and compared its growth rates in no induction and in high levels of induction. PR cells in higher levels of induction experienced lower growth rates than those in no induction, supporting the hypothesis that active rtTA may be more toxic than inactive rtTA (see Fig. 6.4.B).

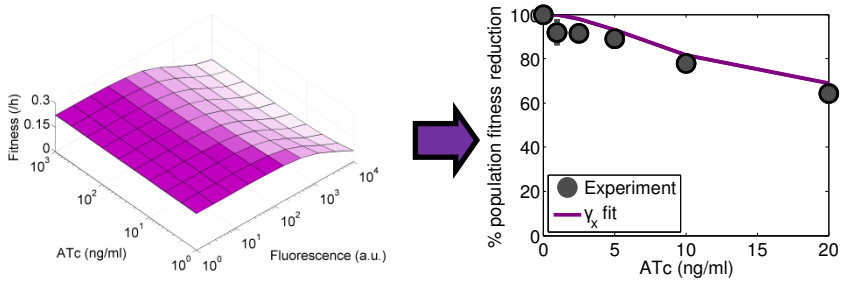


Figure 4.3. Creating the rtTA fitness function

A)Fitness landscapes were created for cells in increasing ATc (inducer) concentrations based on cell yEGFP:ZeoR phenotypic expression levels. B)Using the probability distributions from experimental results, the overall cell population fitness landscapes were calculated, and parameters fit to minimize distance between theoretical and experimental growth reduction. The parameters for the instantaneous fitness reduction function, [76], are $\alpha = 936$ *M,* $\beta = 5.8$ *(ng/ml)* h^{-1}*,* $C = [ATc]$ h^{-1}*,* $F =$ *(fluorescence)* $\times$ M *(a.u.)*$^{-1}$.

I created a semi-phenomenological biochemical reaction model for active rtTA based on its hypothetical squelching effect. The instantaneous fitness reduction of cells was assumed to be correlated to the number of squelched proteins, given by the model

$$\gamma_x = \frac{\alpha}{\alpha + x(F, C)} \qquad [76]$$

where x is active rtTA as a function of fluorescence (F, used as a proxy for rtTA) and ATc (C) which activates rtTA (see section 4.2.5). The probability distribution of cells was estimated from flow cytometry measurements. I used basal fitness rates and the probability distribution function to estimate the overall cell population fitness (g_T). The parameters for the instantaneous fitness reduction were fit by minimizing the Euclidean distance between overall cell population fitness (g_T, see section 4.2.4) with experimental fitness levels (g_E) (see Fig. 4.3.).

4.3.5 Estimates of cellular transition rates based on cellular current.

Probability currents have been used to estimate how quickly particles transition from state to state (75,76). I modified this concept to account for the effects of cell growth. These cellular currents define the flow of cells across an arbitrary boundary (φ). In this case, φ was chosen to separate cells with (*H*)igh *yEGFP*::*ZeoR* expression and (*L*)ow *yEGFP*::*ZeoR* expression. The net cellular current (I) can be split into two components describing increasing fluorescence ($I_{\rightarrow}$) and decreasing fluorescence ($I_{\leftarrow}$).

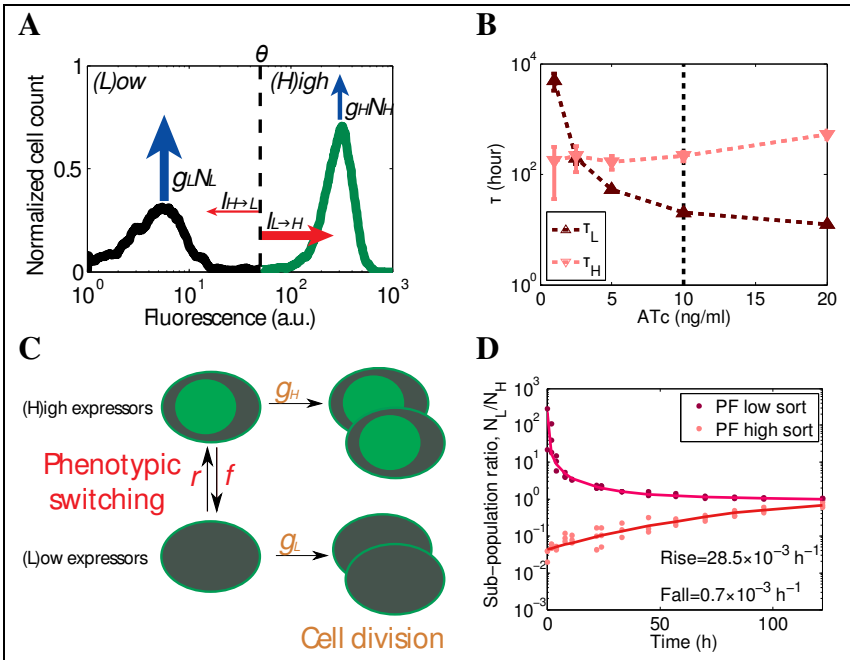


Figure 4.4. Estimating cell transition rates

A) Schematic representation of the relationship between cellular current and cell fitness. Although cells may preferentially switch to a high expressor state, instantaneous fitness can cause an overrepresentation of low expressorcells. B) Cell currents were used to estimate the switching rates of cells based on their probability distributions and instantaneous fitness estimates. The dashed line marks the ATc concentration at which memory verification was performed by cell sorting. C) A two state model using these switching fitness rates was used to predict the time course of sorted cells grown in 10 ng/ml ATc (D). The parameters fit for the two state model are $g_L = 0.32\ h^{-1}$, $g_H = 0.27\ h^{-1}$, $r = 28.5 \pm 6.6 \times 10^{-3}\ h^{-1}$, *and* $f = 0.7 \pm 0.4 \times 10^{-3}\ h^{-1}$.

For a population of cells switching between the *L* and *H* state, the rate at which cells rise out of the low expressor state is given by $r = I_{\rightarrow}/N_L$, while the rate at which cells fall from the high to low expressor state is $f = I_{\leftarrow}/N_H$. The cellular memory of the *L* and *H* states was reported as the inverse of the rise and fall rates as $\tau_L = \ln(2)/r$ and $\tau_H = \ln(2)/f$.

The decreasing cellular current quantifies the number of cells transitioning states due to yEGFP::ZeoR degradation and dilution. Because yEGFP::ZeoR is a stable protein, I assumed that cells lost their fluorescence at the same rate as their cell division times. Thus, the current of cells transitioning to the low state is

$$I_{\leftarrow}(\varphi) \Rightarrow I_{H\rightarrow L}(\varphi) \approx [\varphi \times g_0 \times \gamma_x(\varphi)] \times [\pi(\varphi) \times N_T], \qquad [77]$$

where g_0 is the maximal growth rate of cells, $\gamma_x(\varphi)$ is the reduction in growth rates due to rtTA at the boundary φ, $\pi(\varphi)$ is the probability distribution function of cells at the boundary, and N_T is the total

number of cells in the population. The product of the growth terms times the fluorescence boundary ($\varphi \times g_0 \times \gamma_x(\varphi)$) describes the dilution rate of fluorescent molecules out of cells due to growth.

For stationary probability distributions, the increasing current is equal and opposite to the decreasing current, insuring that probabilities neither increase nor decrease. However unequal fitness rates may cause a disequilibrium between the two currents. Adding a fitness correction term to the high to low current yields (see section 6.16):

$$I_{\rightarrow}(\varphi) \Rightarrow I_{L\rightarrow H}(\varphi) = \underbrace{\varphi g_0 \gamma_x(\varphi)\pi(\varphi)N_T}_{I_{H\rightarrow L}} + \underbrace{\int_0^{\varphi}[g(F) - g_T]\pi(F)dF}_{\text{Selection}}. \qquad [78]$$

These "cellular currents" were numerically estimated by applying the previously obtained growth reduction function, [76], to the number of cells obtained from flow cytometry.

These currents were used to estimate the switching rates of cells over a variety of ATc concentrations (see Fig. 4.4B). I used the previously described two state model to verify cellular memory (see section 4.2.8, Fig. 4.4C) (21,25,26), as well as direct microscopy measurements (see section 6.23). The switching rates *r* and *f* were fit using Matlab's fminsearch, so that the Euclidean distance between simulated and experimental ratios of low to high expressor cells were minimized. Cellular current estimates of memory were verified by fitting the two state model to four sorting experiments from three distinct cell strains (see section 6.16). The predicted ratios of *L* to *H* cells over time were confirmed by experimental sorting ratios, suggesting that growth was indeed responsible for the discrepancies between chemical kinetic models and experimental data.

Dmitry Nevozhay tested the hypothesis that no switching occurred from high to low expressor cells by performing a serial dilution experiment (7). In this experiment, cells were incubated in 10 ng/ml ATc to maintain a 1:1 proportion of cell phenotypes, which were then diluted into concentrations of ~0.1 per tube, and allowed to grow for 10 days. Two tubes had cells that were 99% high expressors after 4 days, which became more evenly distributed over the next 6 days. Five other tubes were composed of approximately equal parts low and high expressor after 4 days, supporting the idea that high expressors switch to the low state slowly, while low expressors switch to the high state quickly (see section 6.15).

4.3.6 Defining the fitness of uninduced PF cells at various levels of Zeocin.

In order to study the interaction between fitness and noise, I used data collected from Zeocin treatment of uninduced (0 ATc) PF cells. I inferred the growth rates of individual cell states from the experimental overall cell population growth rates (g_E) and their fluorescence distributions from cells grown in various Zeocin concentrations. Cells were maintained in the exponential growth phase by

resuspending them every 12 hours in fresh Zeocin containing media. Growth rates were obtained from the slope of a linear fit to log-transformed cell counts as a function of time.

The instantaneous fitness due to Zeocin toxicity was assumed to be proportional to the time DNA remained undamaged. I used simple biochemical assumptions to describe the instantaneous fitness reduction as:

$$\gamma_Z(F,Z) = \frac{\chi}{Z_i(F,Z) + \chi}, \qquad [79]$$

where χ is the ratio of DNA repair rates to Zeocin DNA damage rates, and the concentration of free Zeocin within a cell (Z_i) depends on *yEGFP::zeoR* expression (F) and extracellular Zeocin concentration (Z) (see section 4.2.6).

The parameters were obtained by minimizing the difference between the predicted overall cell population fitness $g_Z(F,Z)$ (see section 4.2.4) and the experimental cell population fitness values g_E. The resulting fitness landscape is shown in Fig. 4.5.

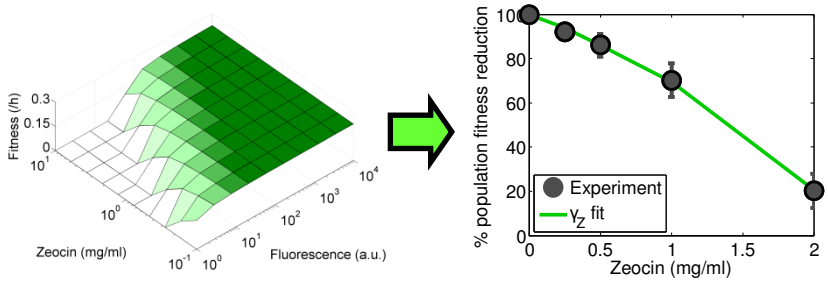


Figure 4.5. Creating the Zeocin fitness function

Fitness landscapes were created for cells in increasing Zeocin stress environments based on cell yEGFP:ZeoR phenotypic expression levels. Using the probability distributions form experimental results, the average fitness landscapes were calculated, and parameters fit to minimize distance between theoretical and experimental growth reduction in the absence of ATc. The parameters used for the instantaneous fitness reduction function, [79]*, are* $d = 0.25\ h^{-1}$, $h_z = 0.5\ h^{-1}$, $s = 1.3\times106\ (mg/ml)^{-1}\ h^{-1}$, $\chi = 6.5\times10^{-7}\ (mg/ml)$, $Z = 1.182 \times [Zeocin]\ h^{-1}$, $F = fluorescence \times (mg/ml)\ (a.u.)^{-1}$.

4.3.7 Defining the fitness landscape for PF cells in novel environments.

Because it is thought that Zeocin and rtTA toxicity occur through two independent pathways (double stranded DNA breaks versus sequestration of transcription factors), I assumed that their toxic effects were also independent. Assuming Bliss independence (137,138), I multiplied the two fitness reductions together to predict the instantaneous fitness reduction of induced cells grown in Zeocin as,

$$\gamma(F,C,Z) = \gamma_Z(F,Z) \times \gamma_x(F,C) \ . \qquad [80]$$

The total cell population growth rate was defined as the maximal division rate times the average reduction in cell fitness;

$$g_t(C,Z) = g_0 \int \gamma(F,C,Z) dF \ . \qquad [81]$$

Loewe additivity was also used to model possible alternative interactions between the two toxic molecules (see section 6.21).

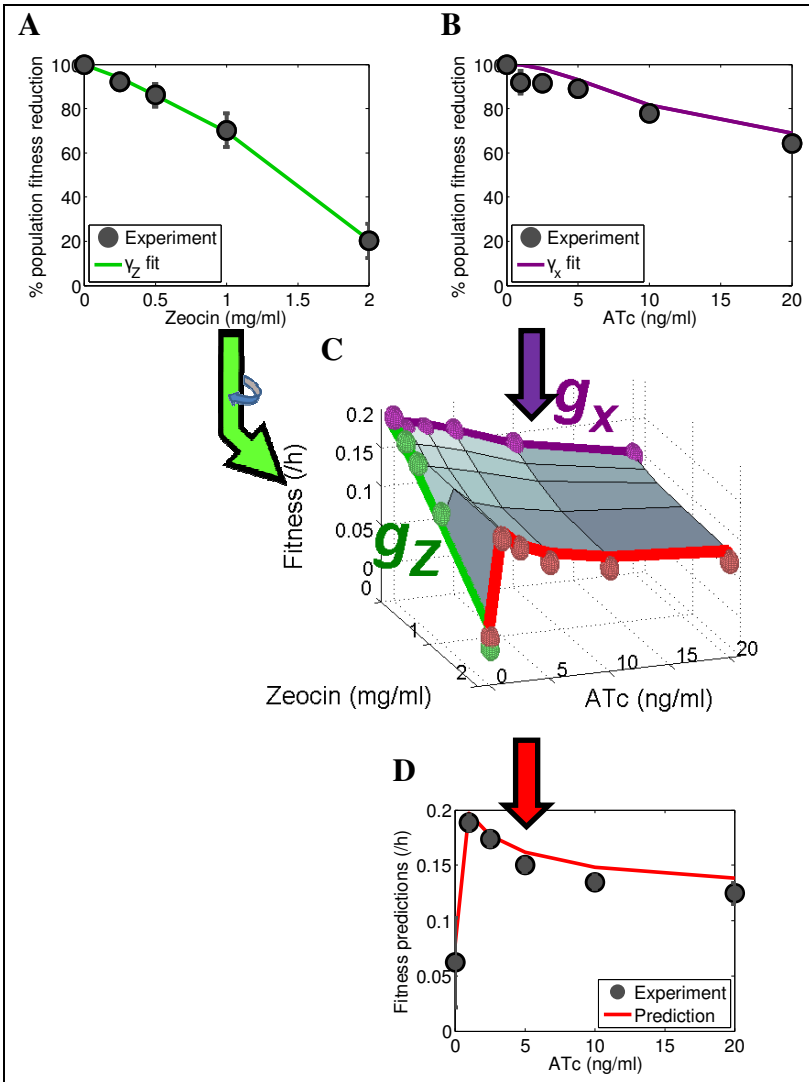
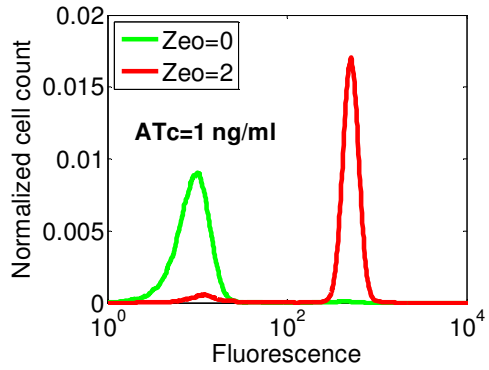


Figure 4.6. Predicting cell fitness in Zeocin environments

The instantaneous fitness function for Zeocin toxicity (A) and rtTA toxicity (B) were used to estimate sub-population growth rates and switching rates. Using these sub-population growth rates, a fitness landscape was predicted for PF cells at many Zeocin and ATc concentrations (C). Experimental points are shown as spheres. These predictions were tested experimentally at 2 mg/ml Zeocin (D).

4.3.8 Predicting overall cell fitness in Zeocin.

Assuming negligible changes to cellular currents in different Zeocin environments, I made predictions of overall cell population fitness (see Fig. 4.6.C). For the most toxic environments, at 2 mg/ml Zeocin, overall cell population fitness was predicted to peak at 1 ng/ml ATc, and then decrease as ATc was increased.



B

Cellular memory (1 ng/ml ATc)		Doubling times	
		0 mg/ml Zeocin	2 mg/ml Zeocin
High expressor	175 h	3.4 h	3.4 h
Low expressor	4970 h	3.0 h	9.0 h

Figure 4.7. The behavior of the sweet spot of drug resistance

(A)Experimental yEGFP::zeoR fluorescence histograms at 0 and 2 mg/ml Zeocin at 1 ng/ml ATc. Cell distributions change from almost 100% low expressors to almost 100% high expressors in the presence of Zeocin due to the resistance conferred by yEGFP::zeoR and memory of the high state. B) Even though cells strongly tend to the low state, cell memory is much longer than cell division rates, allowing the population to dramatically change gene expression in the presence of Zeocin. Doubling times for sub-populations of are given for cells grown in 2 mg/ml Zeocin, 1 ng/ml ATc.

As ATc induction increased, cells were predicted to favor the drug resistant state more strongly. Although this preference for the high state conferred a fitness benefit to cells, the fitness reduction due to more active rtTA negated this benefit. These predictions were confirmed experimentally (see Fig. 4.6.C and D). The "sweet spot" where fitness peaked (at 1 ng/ml ATc) was predicted despite the rise and fall rates strongly favoring the low state. This is because the fall rate was slow enough for cells to reproduce under high Zeocin concentrations (see section 6.22, Fig. 4.7).

Although Bliss independence was assumed to be the most reasonable assumption for the toxic interactions of Zeocin and rtTA, models based on Loewe additivity, synergism, and antagonism were also tested, and did not significantly affect results (see Fig. 6.11.).

4.3.9 Mutual dependence of switching rates and fitness affects survival optimization

For cells with stochastically switching phenotypes, Kussell and Leibler found that cells would have ideal fitness when their switching rates between phenotypes followed the relationship

$$H_{ij}(\text{optimal}) = b_{ij}/T_j \,, \qquad [82]$$

where H_{ij} is the switching rate from phenotype i to j, b_{ij} is the switching rate from environment i to j, and T_j is the average time cells spend in environment j. Phenotype j is defined to be the most fit

phenotype in environment j (20). This result holds as long as the phenotypic switching rate does not affect the fitness of a phenotype.

For PF cells, high expressors are most fit in a high Zeocin environment, and low expressors are most fit in the normal environment. The corresponding ideal memories are predicted to be

$$\tau_H(\text{optimal}) = T_{toxic}^2 \quad -\text{and}- \quad \tau_L(\text{optimal}) = T_{normal}^2 \quad [83]$$

where τ_L and τ_H are the memories of the low and high state cells, and T_{toxic} and T_{normal} are the average duration of the toxic and normal environments, respectively.

For static, toxic environments, cell fitness should increase as memory of the high expressor state increases, corresponding to very high ATc concentrations. However the existence of the sweet spot at 1 ng/ml ATc suggests that these ideal switching rates do not apply when there is a fitness cost to high memory.

Does the ideal switching rate change when phenotypic switching rates and fitness are mutually dependent? In order to answer this question, I created a set of functions that reasonably approximated experimental rates at 0 and 2 mg/ml Zeocin (see Fig. 4.8.A and B). These rates followed experimental rates to maintain biological plausibility, *but were not meant to be predictive*.

I simulated cells switching in oscillating and randomly fluctuating environments (from normal to Zeocin environments). I performed simulations where cells were predicted to have optimal memories at 10 ng/ml ATc (see Fig. 4.8.C) and 100 ng/ml ATc (see Fig. 4.8.D). However simulations showed that ideal switching rates as a function of ATc induction were most ideal at ~1 ng/ml. Next, I simulated environmental switching rates with long intervals in the toxic environment and short intervals in the normal environment (cells would be predicted to be most fit incubated in saturating ATc, see Fig. 4.8.E). Finally, cells were simulated in a purely toxic environment (see Fig. 4.8.F). These simulations suggest that cells with a fitness cost to the rate of switching will benefit from limiting their switching rates instead of mirroring environmental switching rates.

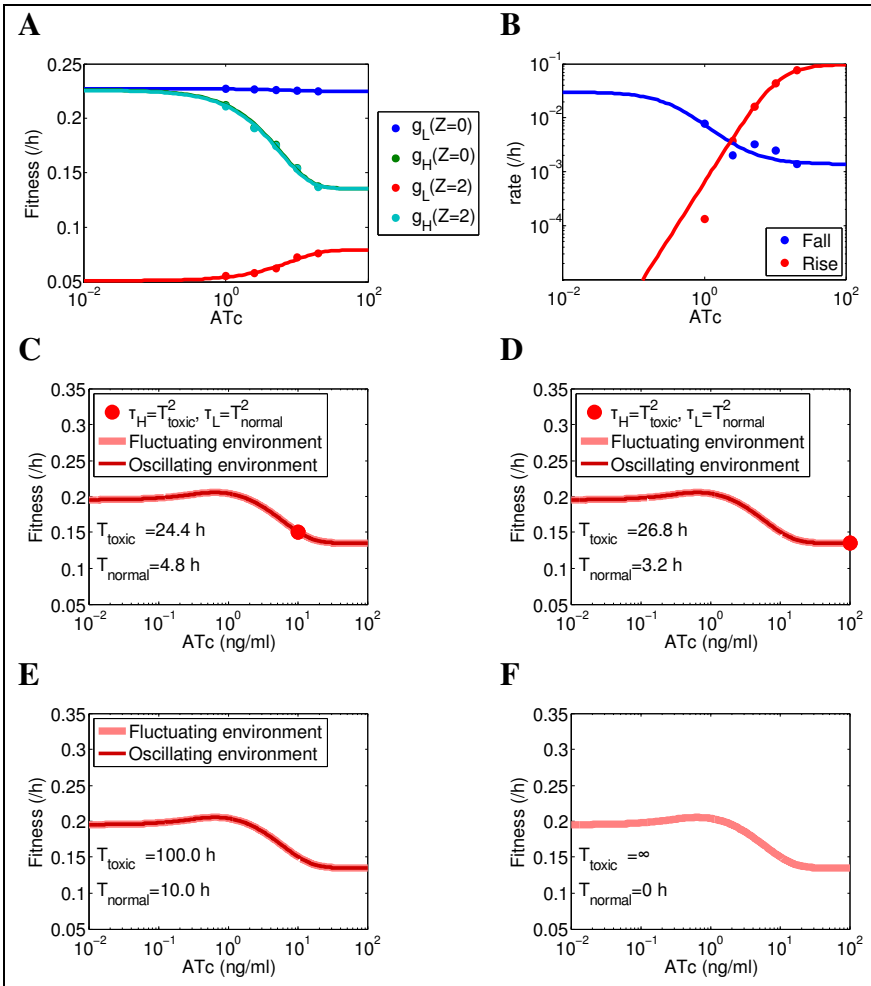


Figure 4.8. Cell fitness in switching environments

Functions were fit to experimental growth at 0 and 2 mg/ml Zeocin (A) and switching (B) rates. These fit functions were used to simulate cell growth when the environment switched between 0 and 2 mg/ml Zeocin. I simulated environments with predicted ideal phenotypic switching rates at 10 ng/ml ATc (C), 100 ng/ml ATc (D), high environmental times (E), and a purely toxic environment (F). Predicted ideal switching rates do not agree with simulated ideal switching rates. The parameters and functions used to fit the growth and switching rates are given in section 4.2.9.

4.4 Discussion

Positive feedback of transcriptional regulation can control cellular differentiation (139). This was observed in the bimodal expression of PF cells. Because cell populations spontaneously relaxed to bimodal distributions in a constant environment, I assumed that switching was due to noisy gene expression. Using standard kinetic theory, I hypothesized that cells with a 1:1 ratio of low to high expressors would stochastically switch phenotypes at identical rates. However, cells sorted into high

and low sub-populations relaxed to their stationary distributions over different time scales (45 hours for low sorted, >122 hours for high sorted cells), implying that chemical kinetics models would be unable to reproduce PF cell dynamics (see section 6.12).

I then hypothesized that growth retardation from active rtTA was influencing the distribution of cells, causing low expressor cells to be more overly represented than would be expected by chemical kinetics alone. In order to quantify this effect, I estimated the instantaneous fitness of cells as a function of their internal cell states and their environment. This fitness "landscape" is similar to previously defined phenotypic fitness landscapes (140), where random, phenotypic plasticity allows cells to explore multiple phenotypes from a single genotype (141). By averaging instantaneous fitness over the *entire* cell population, I was able to measure the corresponding overall cell population fitness independently of noisy transitions.

I used the concept of cellular currents, inspired by probability currents (75,76), to estimate the memories of sub-populations. With the addition of a correcting cell fitness term, cell populations with 1:1 ratios of low to high expressors were predicted to have starkly unequal switching rates similar to those seen experimentally (see Fig. 4.4C, D, and sections 6.19, 6.20 for models incorporating chemical kinetics and cell growth).

The concept of cellular currents may allow switching rates to be inferred indirectly. Sorting can be stressful on cells, potentially precluding memory estimates. Furthermore, for fast switching cells, measuring division rates of sub-populations is extremely difficult, and subject to high noise. By using the concept of cellular currents, it may be possible to estimate cellular memory without performing difficult experimental measurements.

Previous papers have hypothesized that cells unable to sense their environments may evolve to maximize the mutual information between phenotypic switching rates and environmental switching rates (20,21). These papers assumed that cells only experienced costs from switching to unfit phenotypes. However, when cells experience a fitness decrease due to increased switching rates, cells are no longer under pressure to match their phenotypic switching rates to their environmental switching rates.

Do natural living systems endure a cost for increasing memory of stress tolerant states? Energetically, the answer is likely to be yes since energy must be expended to minimize noise induced switching. Bistable stress-tolerance and stress resistance states have been shown to emerge from implicit growth mediated feedback loops due to the expression cost or toxicity of various defense mechanisms (142-144). Examples of this include the toxin component of toxin-antitoxin systems (145,146), *hipQ* expression for type II persisters (147), and the TetA protein from the *tetRA* tetracycline resistance operon (148). Although these natural systems have defense and fitness costs

arising from a single protein, this is not unlike the cost (*rtTA*) and defense (yEGFP::ZeoR) of PF cells which were expressed from identical promoters, with identical regulation. Thus, when the genes involved in maintaining memory are toxic or metabolically expensive, the cost of maintaining high memory may become very significant.

This work may provide insights into the nature of cell population adaptation to drugs. The sweet spot of drug resistance provides intriguing insights into the nature of drug resistance and phenotypic plasticity (see section 4.3.8, Fig. 4.7). At low levels of induction, PF cells showed an overwhelming tendency to maintain the drug sensitive phenotype. However in high Zeocin environments, the drug resistant phenotype was overwhelmingly represented in the population. This effect was not due to increasing induction of the gene circuit, but was instead due to selection from Zeocin and a sufficiently high memory in the high state to maintain drug resistance. Similar drug resistance mechanisms may be responsible for the emergence of microbial or cancer drug resistance, where a minority of pre-existing drug-tolerant cells may survive drug treatment and propagate to form an increasingly drug-resistant population of cells (149-151).

Chapter 5 Conclusions

Synthetic gene expression systems based on *tetR* were used to study a wide range of phenotypic behaviors. Concordant with this wide range of phenotypic behaviors, even small changes to these gene expression systems resulted in significant, non-linear changes to cell behaviors.

Incorporating the effects of cell growth and division into kinetic modeling were essential to understand cell behaviors. Using constitutively expressing *tetR* gene expressions as a template (Chapter 2), I created a set of models to understand the effects of negative and positive autoregulation (Chapter 3 and Chapter 4, respectively) while incorporating the effects that living, dividing cells have on chemical kinetics.

For constitutively expressed *tetR* gene expression systems, previous models predicted that downstream noise would be proportional to upstream gene expression noise (52,91,92,94,96,97). My model predicted that if *tetR* gene expression were increased, while TetR noise remains constant, then downstream noise will increase. This mechanism of noise propagation is dependent on the living nature of cells, and suggests that cell growth creates a disequilibrium between environmental and cellular ATc concentration. This disequilibrium is due TetR sequestering incoming ATc. This model accounts for experimental behaviors (such as extreme sensitivity of cells to growth phase), and has a provided a basis for understanding more complicated gene expression systems.

Controlling constitutively expressed TetR gene expression systems at intermediate levels of induction is difficult. Because these gene expression systems are so sensitive to environmental ATc concentrations, small changes in ATc can result in large changes in gene expression. Furthermore, as cell growth rates change, TetR concentrations will fluctuate due to changes in production and dilution rates (143), and downstream gene expression will change dramatically.

The addition of an autoregulatory feedback loop eliminated these behaviors, and resulted in a linear dose-response. Because this linear behavior can be desirable for experimentalists trying to control genes of interest, it would be useful to understand when and how linearizers arise. Using modeling techniques, I found a set of conditions necessary to maintain a linear dose-response which was subsequently tested. These conditions are 1) high inducer retention (e.g., low inducer diffusion across cell membranes, high binding affinities to the negative regulator), 2) mutual and low degradation rates (e.g., from dilution), and 3) identical or linearly transformed up- and down-stream repressible promoters.

The linearizer will provide an excellent tool for studying the link between genotype and phenotype when gene expression is constant. However, because linearizers have relatively fast dynamics, and muted noise levels, it is not well suited to studying the effects of stochastic gene expression and memory on cell phenotypes.

In order to study the relationship between gene expression noise, cellular memory, and fitness, I studied cells with gene architectures forming positive feedback loops (PF). PF cells were created by placing the *tetR* variant *rtTA* (a transcriptional activator) under its own control. Consistent with observations that positive feedback loops are often involved in cellular differentiation (102-104), PF cells displayed two distinct phenotypes. Although it was initially thought that the ratio of switching rates between the two phenotypes were proportional to the ratio of cells with low and high expressor phenotypes, this hypothesis was rejected. It was subsequently found that rtTA was slowing cell growth, causing high expressor phenotypes to grow more slowly than low expressors.

Although this rtTA growth retardation effect is a significant caveat when studying phenotypic switching, it has provided insights in how the genotype encodes for phenotype. In order to understand the factors controlling cell population fitness, I developed two novel concepts to quantitatively predict how genotype, environment, and cellular memory interacted to determine cell population fitness. These concepts are "instantaneous fitness" and "cellular currents" which allowed me to infer phenotypic fitness landscapes independently of noisy gene expression, and estimate the rate at which cells transitioned between phenotypes from distributions of cells. These concepts make it possible to estimate growth and memory rates that may be impossible using traditional experimental methods.

This work raises several questions. These include whether properties of linearizer systems can be improved how fitness and memory guide evolution. Initial work to answer these questions are described in the following sections.

5.1 Future directions

5.1.1 Improving linearizer circuits

Although the linearizer is interesting from a theoretical point of view, it will also be useful for experimental study. Ideally, this gene expression system could be modified to study investigators' genes of interest, or could be used to create cheap, reproducing biosensors.

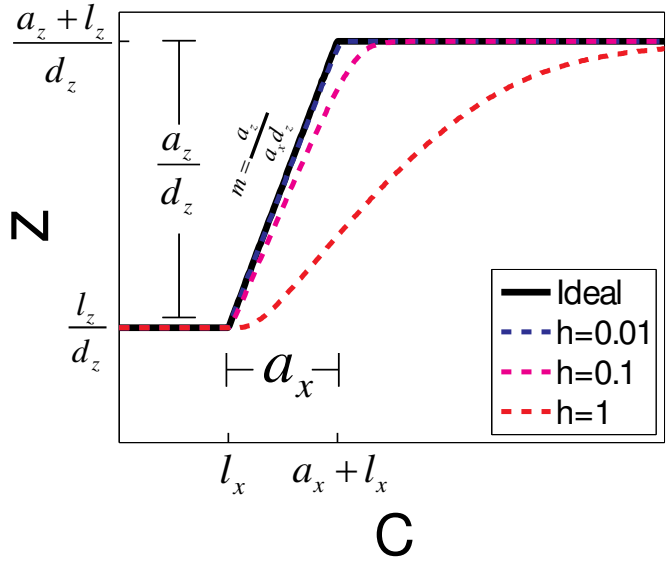


Figure 5.1. Modifying linearizer parameters

For an ideal linearizer circuit the following modifications to the parameters can shift, scale, and soften the dose-response.

Although it was shown that identical repressor and downstream promoters resulted in a linear dose-response, other parameters may be changed to enhance (or degrade) the behavior of linearizer circuits, or to create linearizers from different repressors and promoters.

For example, genes which act to decrease inducer concentrations within the cell may demonstrate reduced linearity. This may be an important caveat for investigators studying drug efflux genes.

Based on analysis of the simplified negative feedback model [39],

$$\begin{aligned} \dot{x} &= a_x F_x(x) + l_x - bxy - d_x x \\ \dot{y} &= C - bxy - hy \\ \dot{z} &= a_z F_z(x) + l_z - d_z z \end{aligned} \quad , \qquad [84]$$

it may be possible to control various aspects of the dose-response. For example, it may be possible to reduce the sensitivity of the downstream gene by increasing the upstream gene's maximum gene expression level. This would have the desirable property of making cells less sensitive to inducer concentrations (see Fig. 5.1.).

Alternatively, the linearizer can be modified to measure a chemical concentration over a narrow band of inducer concentrations. This is similar to previous work creating in-band detection systems which only respond to intermediate concentrations of inducers (152). Similarly, by increasing basal repressor (or *tetR*) expression, the linearizer may be made to measure specific, intermediate inducer concentrations.

5.1.2 Predicting evolutionary changes in populations

In order to understand behavior of cell distributions subject to stochastic transitions, I used the concept of cellular currents to estimate the memories of cell states (see section 1.3.4). For cells in a stationary distribution, the number of cells increasing their gene expression levels (or increasing current) should be equal to the number of cells with decreasing gene expression levels (or decreasing current), so that probabilities do not change. I noted that differences between sub-population fitness and overall cell population fitness could result in a disequilibrium between increasing and decreasing currents of cells. I described the disequilibrium as

$$\underbrace{N_T \int_0^{\varphi} [g(F) - g_T]\, p(0,\varphi)\, dF}_{\text{Selection}} = I_{L\to H}(\varphi) - I_{H\to L}(\varphi). \qquad [85]$$

where $g(F)$ is the instantaneous fitness of cells at F fluorescence, g_T is the overall cell population fitness, $p(F)$ is the probability density of cells with fluorescence F, N_T is the total number of cells, and $I_{L\to H}$ and $I_{H\to L}$ are the cellular currents of cells across the boundary φ. The subscript $L\to H$ denotes the current where low expressor cells switch to high expressor cells, while the subscript $H\to L$ denotes the current where high expressor cells transition to low expressor cells. This equality came from the equation describing changes in cell density as

$$\frac{\partial}{\partial t}[N_T\, p(F)] = \frac{\partial}{\partial F}[I_{L\to H}(F) - I_{H\to L}(F)] + g(F) N_T\, p(F) \qquad [86]$$

where $N_T p(F)$ is the density of cells at fluorescence F. Here I temporarily drop the dot notation to denote the time derivative, since the term inside of the derivative is the product of two terms. Noting the relationship, $g_T = \frac{\dot{N}_T}{N_T}$, the rate that the probability distribution changes is:

$$\frac{\partial}{\partial t}[p(F)] = \frac{\partial}{\partial F}\left[\frac{I_{L\to H}(F) - I_{H\to L}(F)}{N_T}\right] + g(F)p(F) - g_T\, p(F). \qquad [87]$$

I defined selection pressure as the rate at which cell growth reshapes the probability distribution function as:

$$s = g(F)p(F) - g_T p(F). \qquad [88]$$

Preliminary efforts to quantify this selective pressure have yielded intriguing results (see Fig. 5.2.A).

As a second measure of evolutionary pressure acting on cell populations, I performed a sensitivity analysis of population fitness to changes in the percentage of phenotypes within a population.

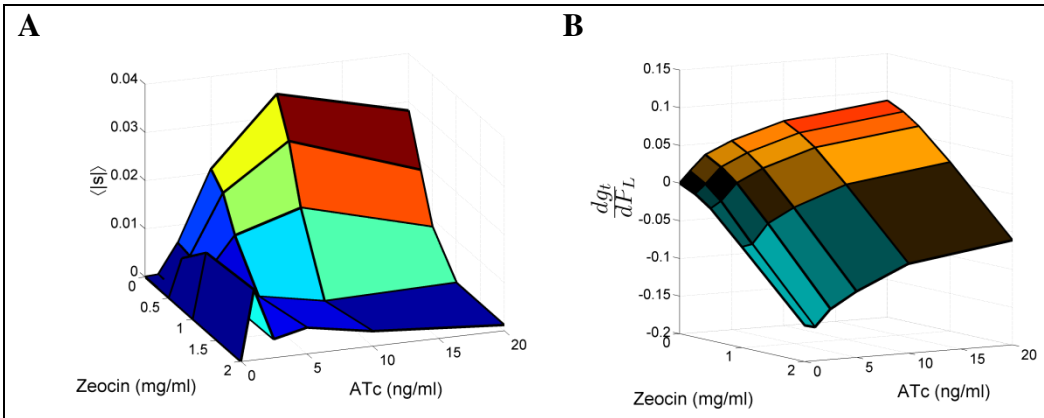


Figure 5.2. Selection pressure on PF cells

A) Shown are the average absolute selection pressures for PF cells. The average selection pressure suggests that cellular currents for cells at 1 ng/ml ATc and 2 mg/ml Zeocin are in disequilibrium due to cell growth. The selective pressure acting cells at 0 mg/ml ATc strongly increases as ATc is increased. B) Shown are the changes in overall growth rates corresponding to small changes in the population proportions. Blue sections predict that cells would benefit greatly from increased high expressors, while the red sections predict cells would benefit greatly by increasing the number of low expressors.

Using a two state model, I asked how much cell fitness would change as the number of low yEGFP::ZeoR expressing cells change. Based on the formula

$$g_t = g_L P_L + g_H (1 - P_L), \qquad [89]$$

I calculated the sensitivity of population fitness to the percentage of low expressor cells as

$$\frac{\partial g_t}{\partial P_L} = g_L - g_H . \qquad [90]$$

The results of the calculation suggest the strongest evolutionary pressures occur at low Zeocin/high ATc concentrations and high Zeocin/low ATc concentrations (see Fig. 5.2.B).

These metrics may predict how cells evolve over time, which may be used to ask how network architecture affects how quickly genes evolve (153-155).

Chapter 6 Appendix

6.1 Gal10-T123 Dizzy code

```
//Dizzy parameters for GAL10-T123
b=100; //ATc – TetR binding rate
l=10; //yEGFP basal transcription rate
lup=0.2; //TetR basal transcription rate

m=273; //maximal yEGFP transcription rate
mup=273; //maximal TetR expression

mu=3.5; //yEGFP mRNA degradation rate
muup=3.5; //TetR mRNA degradation rate

p=75; //yEGFP translation rate
d=75/2; //TetR translation rate

pi=0.3; //yEGFP degradation rate
delta=0.3; //TetR degradation rate

a=3.7; //Gal1 activation rate
aup=3.7; //Gal10 activation rate, changes depending on mutation
degraded=0; //number of degraded molecules

alpha=0.28; //Gal1 inactivation rate
alphaup=0.28; //Gal10 inactivation rate, changes depending on mutation
log2=0.6931;
rho=log2/(12/60); //TetR - tetO2 site dissociation rate
hindrance=5; //amount each TetR bound to tetO2 increases dissociation
r=0.1; //TetR- tetO2 binding affinity
diffusion=0.9242; //ATc diffusion rate across the cell membrane
c=279/2.25; //Constant relating ATc concentration to influx into the cell

M=l/mu; //yEGFP mRNA
P=M*p/pi; //yEGFP
Mup=mup*aup/(aup+alphaup)/muup+lup/muup; //TetR mRNA
D=Mup*d/delta; //Unbound TetR

H=0; //1 ATc bound to TetR
B=0; //2 ATc bound to TetR
R1=0; //1 TetR bound to T123 promoter
R2=0; //2 TetR bound to T123 promoter
R3=1; //3 TetR bound to T123 promoter
S=0; //0 TetR bound to T123 promoter
A=1; //Active T123 promoter
Aup=1;//Active Gal10 promoter
A0=0; //Inactive T123 promoter
Aup0=0; //Inactive Gal10 promoter

I=0; //ATc in the cell
Ie=0; //External ATc concentration, changes depending on ATc concentration in growth medium
```

Box 6.1 T123 parameters for Gillespie Simulations in Dizzy code

```
//Dizzy reactions for GAL10-T123
//ATc influx and outflux
Iinflux, Ie->Ie+I,          diffusion*c;
Ioutflux,          I->degraded,     diffusion;

//Promoter activation and inactivation
r1,       A->A0, alpha;
r2,       A0->A, a;
r3,       Aup->Aup0,      alphaup;
r4,       Aup0->Aup,      aup;

//TetR production and degradation
r5,       Aup->Aup+Mup,          mup;
r6,       Aup0->Aup0+Mup,        lup;
r7,       Mup->degraded,         muup;
r8,       Mup->Mup+D,  d;
r9,       D->degraded,    delta;

//TetR binding to ATc
r10,      D+I->H,          2*b;
r11,      H->degraded,    delta;
r12,      H+I->B,          b;
r13,      B->degraded,    delta;

//TetR binding to the T123 promoter
s1,       D+S->R1,         r;
s2,       R1->D+S,         rho;
s3,       D+R1->R2,        r;
s4,       R2->D+R1,        rho*hindrance;
s5,       D+R2->R3,        r;
s6,       R3->D+R2,        rho*hindrance*hindrance;

//yEGFP production rate
s7,       A0->A0+M,      l;
s8,       A+S->A+S+M, m;
s9,       M->degraded,    mu;
s10,      M->M+P,          p;
s11,      P->degraded,    pi;
```

Box 6.2 *Reactions for T123 Gillespie simulations in Dizzy code*

6.2 Gal10-T123 template model parameters

Parameter	Value	Source
a	3.7	(156)
a_{up}	wt:2.79 int1:6.82 int2:4.19 sev1:0.1964 sev2:0.1985	Fit from GAL10 mean and noise
b	100	(111)
c	124	Fit to wild-type dose-response
d	75/2	Half of yEGFP translation due to dimerization
hindrance	5	TetR hinders other TetR molecules from binding, (92)
i,o	0.924	(113)
l	10	Fit from uninduced yEGFP noise
l_{up}	0.2	Fit from GAL10 mean and noise
m,m_{up}	273	(157)
p	75	Fit to saturated noise levels
r	0.1	Fit from yEGFP peak noise
α	0.28	(156)
α_{up}	wt:1.05 int1:16.22 int2:41.56 sev1:24.84 sev2:22.68	Fit from GAL10 mean and noise
δ,π	0.3	Experimental cell division rate
μ,μ_{up}	3.5	Generic mRNA degradation rate, (157)
ρ	3.47	(158)

Table 6.1 *GAL10-T123 Template model parameters*

6.3 Linear Noise Approximation of the Gal10-T123 template model

The drift matrix for the template model (see Fig. 2.1) is

$$\mathbf{A}\left(\begin{bmatrix} A_{up} \\ A_{0up} \\ M_{up} \\ D \\ H \\ B \\ I \\ R_3 \\ R_2 \\ R_1 \\ S \\ A \\ A_0 \\ M \\ P \end{bmatrix}\right) = \begin{bmatrix} -\alpha_{up}A_{up} + aA_{0up} \\ \alpha_{up}A_{up} - aA_{0up} \\ l_{up}A_{0up} + m_{up}A_{up} - \mu_{up}M_{up} \\ dM_{up} - (2bI + \delta)D \\ 2bDI - (bI + \delta)H \\ bHI - \delta B \\ (-2bD - bH - o)I + ciI_E \\ rDR_2 - 3h^2\rho R_3 \\ -rDR_2 + 2rDR_1 + 3h^2\rho R_3 - 2h\rho R_2 \\ -2rDR_1 + 3rDS + 2h\rho R_2 - \rho R_1 \\ -3rDS + \rho R_1 \\ -\alpha A + aA_0 \\ \alpha A - aA_0 \\ mA + lA_0 - \mu M \\ pM - \pi P \end{bmatrix}, \quad [91]$$

This matrix was used to find the steady state solutions for the template model. To simplify the solution, I approximated the promoter dynamics by the steady state solution,

$$S = \frac{h^3\rho^3}{r^3D^3 + 3h^2r^2\rho D^2 + 3h^3r\rho^2D + h^3\rho^3}, \quad [92]$$

and used the equalities $A_{0up} = 1 - A_{up}$, and $A_0 = 1 - A$ to simplify the drift matrix as

$$\mathbf{A}\left(\begin{bmatrix} A_{up} \\ M_{up} \\ D \\ H \\ B \\ I \\ A \\ M \\ P \end{bmatrix}\right) = \begin{bmatrix} -\alpha_{up}A_{up} + a(1 - A_{up}) \\ l_{up}(1 - A_{up}) + m_{up}A_{up} - \mu_{up}M_{up} \\ dM_{up} - (2bI + \delta)D \\ 2bDI - (bI + \delta)H \\ bHI - \delta B \\ (-2bD - bH - o)I + ciI_E \\ -\alpha A + a(1 - A) \\ mA + l(1 - A) - \mu M \\ pM - \pi P \end{bmatrix}. \quad [93]$$

The diffusion matrix is

$$\mathbf{B} = \left(\begin{bmatrix} \mathbf{B_{up}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B_{bind}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B_{GFP}} \end{bmatrix} \right), \quad [94]$$

where

$$\mathbf{B_{up}} = \begin{bmatrix} \alpha_{up} A_{up} + a\left(1 - A_{up}\right) & 0 \\ 0 & l_{up}\left(1 - A_{up}\right) + m_{up} A_{up} + \mu_{up} M_{up} \end{bmatrix}, \quad [95]$$

$$\mathbf{B_{bind}} = \begin{bmatrix} dM_{up} + \left(2bI + \delta\right)D & -2bID & 0 & 2bID \\ -2bID & 2bDI + \left(bI + \delta\right)H & -bHI & -2bID + bHI \\ 0 & -bHI & bHI + \delta B & bHI \\ 2bID & -2bID + bHI & bHI & \left(2bD + bH + o\right)I + ciI_E \end{bmatrix}, \quad [96]$$

and

$$\mathbf{B_{GFP}} = \begin{bmatrix} \alpha A + a\left(1 - A\right) & 0 & 0 \\ 0 & mA + l\left(1 - A\right) + \mu M & 0 \\ 0 & 0 & pM + \pi P \end{bmatrix}. \quad [97]$$

The linear noise approximation of these matrices are shown in Fig. 2.4B. Alternatively, I approximated the active promoter, *S*, by the Hill function

$$S = \frac{1}{D + 1}, \quad [98]$$

in Fig. 2.4C. The parameters for the rate constants are given in section 6.2.

6.4 Gal10-T123 promoter model

The Linear Noise Approximation [14] was applied to the promoter model [31], with chemical species, drift [4], and diffusion matrices [5] defined as:

$$\mathbf{x} = \begin{bmatrix} x \\ z \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} a_x - gx \\ a_z F(x) + l_z - gz \end{bmatrix}, \text{and} \quad , \qquad [99]$$

$$\mathbf{B} = \begin{bmatrix} sa_x + gx & 0 \\ 0 & a_z F(x) + l_z + gz \end{bmatrix}$$

where $F(x) = \frac{C^n}{C^n + x}$. The phenomenological parameter *s* was used to scale the model's upstream noise to match experimental noise measurements. The Jacobian [13] of **A** is

$$\mathbf{J} = \begin{bmatrix} -g & 0 \\ -\frac{a_z}{x} F(x)[1 - F(x)] & -g \end{bmatrix}, \qquad [100]$$

and the steady state solution of **x**, **A**(**x***) = 0, is

$$\mathbf{x}^* = \begin{bmatrix} x^* \\ \frac{a_z}{g} F(x^*) + \frac{l_z}{g} \end{bmatrix}, \qquad [101]$$

where

$$x^* = \frac{a_x}{g}. \qquad [102]$$

The noise term for *x* (from [19]) is

$$\eta_x^2 = -\frac{B_{11}}{2J_{11}} = \frac{s+1}{2x^*}. \qquad [103]$$

The extrinsic noise contribution from free TetR becomes (from [19]):

$$\eta_z^2 = \frac{1}{2}\eta_x^2 \left(\frac{a_z F(x^*)[1 - F(x^*)]}{a_z F(x^*) + l_z} \right)^2 . \qquad [104]$$

The parameters used for the promoter model are given in table 6.2.

Parameter	Value	Source
a_x	wt:2500 int1:1153 int2:515 sev1:253 sev2:256	Obtained from measured GAL10 expression
a_z	2500	Similar to the template model, $\approx m\ p/\mu$
g	0.3	Experimental cell division rate
l	200	Similar to the template model, $\approx l\ p/\mu$
n	6.3	Fit from wt-T123 dose-response
s	wt: 383 int1:165 int2:87 sev1:46 sev2:50	Solved from the Linear Noise Approximation at C = 0 to match measured GAL10 (x) noise, $s = 2\left(\eta_{\mathrm{GAL10}}\frac{a_x}{g}\right)^2\frac{g}{a_x} - 1$

Table 6.2 Promoter model parameters

6.5 Gal10-T123 sequestration model

The Linear Noise Approximation [14] was applied to the sequestration model [35], with chemical species, drift [4], and diffusion matrices [5] defined as:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} a_x - bxy - gx \\ C - bxy - hy \\ a_z \dfrac{\theta^n}{\theta^n + x^n} + l_z - gz \end{bmatrix}, \text{ and} \qquad [105]$$

$$\mathbf{B} = \begin{bmatrix} sa_x + bxy + gx & -bxy & 0 \\ -bxy & C + bxy + hy & 0 \\ 0 & 0 & a_z \dfrac{\theta^n}{\theta^n + x^n} + l_z + gz \end{bmatrix}.$$

The Jacobian [13] of **A** is

$$\mathbf{J} = \begin{bmatrix} -by - g & -bx & 0 \\ -by & -bx - h & 0 \\ a_z \dfrac{n\theta^n x^{n-1}}{\left(\theta^n + x^n\right)^2} & 0 & -g \end{bmatrix}, \qquad [106]$$

and the steady state solution of **x**, $\mathbf{A}(\mathbf{x}^*) = 0$, is

$$\mathbf{x}^* = \begin{bmatrix} x^* \\ \dfrac{C}{bx^* - h} \\ \dfrac{a_z}{g}\dfrac{\theta^n}{\theta^n + (x^*)^n} + \dfrac{l_z}{g} \end{bmatrix},$$

where [107]

$$x^* = \frac{a_x b - bC - gh + \sqrt{\left(a_x b - bC - hg\right)^2 + 4a_x bgh}}{2bg}.$$

Although an analytical solution may be found, the result will be extremely complicated. The parameters used are shown in Table 6.3.

Parameter	Value	Source
a_x	wt:2500 int1:1153 int2:515 sev1:253 sev2:256	Obtained from measured GAL10 expression
a_z	2500	Similar to the template model, $\approx m\,p/\mu$
b	1	Reduced from the template model to account for distorted sharpness from the Linear Noise Approximation
g	0.3	Experimental cell division rate
h	0.924	(113)
l	200	Similar to the template model, $\approx l\,p/\mu$
n	1	Alternative hypothesis, low cooperativity
s	wt: 383 int1:165 int2:87 sev1:46 sev2:50	Solved from the Linear Noise Approximation at C = 0 to match measured GAL10 (x) noise, $s = 2\left(\eta_{\mathrm{GAL10}}\dfrac{a_x}{g}\right)^2 \dfrac{g}{a_x} - 1$
θ	30	Similar to the template model, $\approx \rho/r$

Table 6.3 Sequestration model parameters

6.6 Gal10-T123 sequestration model correlation matrix

Correlation is often used in addition to noise to understand the dynamics of a system. Included are the correlations between the molecules free TetR (x), internal ATc (y), and yEGFP (z) for the sequestration model [35].

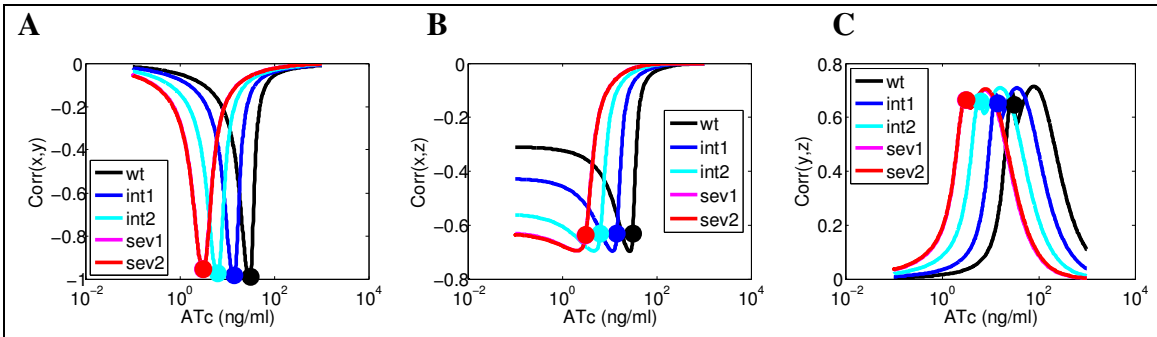


Figure 6.1. Correlation between molecules for sequestration model

The Linear Noise Approximation was used to predict the correlation between Free TetR and free ATc (A), free TetR and yEGFP (B), and free ATc and yEGFP (C) as ATc increases. The solid circles correspond to maximal yEGFP noise. Rates used are given in table 6.3.

6.7 Gal10-T123 sequestration model approximation

Assuming fast ATc dynamics with negligible contributions to noise, the sequestration model [35] chemical species, drift [4], and diffusion matrices [5] can be further approximated as:

$$\mathbf{x} = \begin{bmatrix} x \\ z \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} a_x - Cx - gx \\ a_z F(x) + l_z - gz \end{bmatrix}, \text{and} \quad , \qquad [108]$$

$$\mathbf{B} = \begin{bmatrix} sa_x + Cx + gx & 0 \\ 0 & a_z F(x) + l_z + gz \end{bmatrix}$$

where $F(x) = \dfrac{\theta^n}{\theta^n + x^n}$.

The Jacobian [13] of **A** is

$$\mathbf{J} = \begin{bmatrix} -C - g & 0 \\ -a_z \dfrac{n}{x} F(x)[1 - F(x)] & -g \end{bmatrix}, \qquad [109]$$

and the steady state solution of **x** is

$$\mathbf{x}^* = \begin{bmatrix} x^* \\ \frac{a_z}{g} F(x^*) + \frac{l_z}{g} \end{bmatrix}, \qquad [110]$$

where $x^* = \frac{a_x}{C+g}$

The noise term for x (from [19]) is

$$\eta_x^2 = -\frac{B_{11}}{2J_{11}} = \frac{s+1}{2x^*}. \qquad [111]$$

The extrinsic noise contribution from free TetR becomes (from [19]):

$$\eta_z^2 = \eta_x^2 \left(\frac{a_z n F(x^*)[1-F(x^*)]}{a_z F(x^*) + l_z} \right)^2 \frac{g}{C+2g} . \qquad [112]$$

Substituting TetR's noise gives

$$\eta_z^2 = \left(\frac{a_z n F(x)[1-F(x)]}{a_z F(x) + l_z} \right)^2 \frac{s+1}{4} \frac{g}{a_x} \left(1 + \frac{C/2}{C/2+g} \right). \qquad [113]$$

Using identical parameters to those for sequestration model, except n which was increased to 6 to account for the steep dose-response gives qualitatively similar results to the sequestration model as shown in Fig. 6.2.

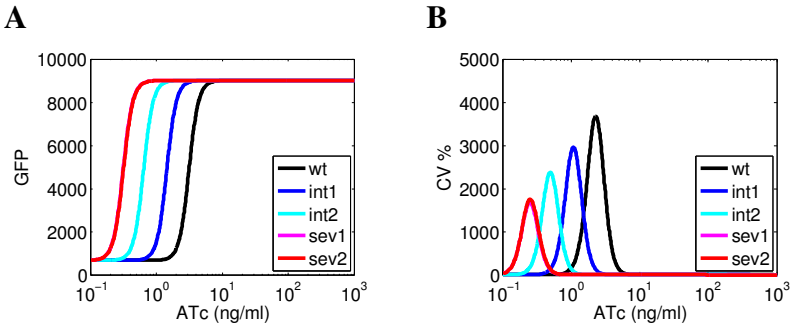


Figure 6.2. Dose-response and noise for approximated sequestration model

Approximating the sequestration by ignoring ATc influx into the cell gives qualitatively comparable behavior. Mean expression is reasonable (A), while the relative height between noise peaks is reasonable, it is greatly amplified (B). Rates used are given in table 6.3, except for the cooperativity of TetR binding to DNA, which is set to n = 6.

6.8 NR Dizzy code

```
b=100;//TetR ATc binding rate
l=2;// yEGFP mRNA leakage
lup=2;// TetR mRNA leakage

m=191;//Maximal yEGFP mRNA production
mup=273; //Maximal TetR mRNA production
mu=3.5; //yEGFP mRNA degradation rate
muup=3.5; //yEGFP mRNA degradation rate

p=75; //yEGFP translation rate
d=75/2; //TetR translation rate
pi=0.3; //yEGFP degradation/dilution rate
delta=0.3; //TetR degradation/dilution rate

a=3.7; //TBP binding rate to yEGFP promoter
aup=3.7; //TBP binding rate to TetR promoter
degraded=0; //Number of degraded molecules

alpha=0.28; //TBP unbinding rate from yEGFP promoter
alphaup=0.28; //TBP unbinding rate from TetR promoter
log2=0.6931; //ln(2)
rho=log2/(2/60); //TetR unbinding rate from a tetO2 site
r=0.2; //TetR binding rate to a tetO2 site
diffusion=0.9242; //ATc diffusion rate across the cell membrane
c=70; //Coefficient for ATc influx

basal=1; //Is there basal expression? 1 if the answer is yes.
M=l/mu; //Starting number of yEGFP mRNA molecules
P=M*p/pi; //Starting number of yEGFP molecules
Mup=mup*aup/(aup+alphaup)/muup+lup/muup; //Starting number of TetR mRNA
D=Mup*d/delta; //Starting number of TetR molecules

H=0; //TetR bound to 1 ATc
B=0; //TetR bound to 2 ATc
R1=0; //1 TetR bound to yEGFP promoter
R2=1; //2 TetR bound to yEGFP promoter
S=0; //0 TetR bound to yEGFP promoter

A=1; //TBP bound to yEGFP promoter
Aup=1; //TBP bound to TetR promoter
A0=0; //no TBP bound to yEGFP promoter
Aup0=0; //no TBP bound to TetR promoter

I=0; //ATc in the cell
Ie=10; //ATC in cell media, currently set to 10 ng/ml ATc
```

Box 6.3 NR parameters for Gillespie Simulations in Dizzy code

```
Iinflux,    Ie->Ie+I,   diffusion*c;//ATc influx into the cell
Ioutflux,   I->degraded,      diffusion;//ATc outflux out of the cell

r1,   A->A0,      alpha;//TBP unbinding to yEGFP promoter
r2,   A0->A,      a;//TBP binding yEGFP promoter
r3,   Aup->Aup0,  alphaup;//TBP unbinding to TetR promoter
r4,   Aup0->Aup,  aup;//TBP binding to TetR promoter


r5,   Aup->Aup+Mup,     mup; //TetR production
r6,   basal->basal+Mup, lup; //TetR basal production
r7,   Mup->degraded,    muup; //TetR mRNA degradation
r8,   Mup->Mup+D, d; //TetR production
r9,   D->degraded,      delta; //TetR degradation

r10,  D+I->H,     2*b; //ATc binds to TetR
r11,  H->degraded,      delta; //TetR with 1 ATc degrades
r12,  H+I->B,     b; //ATc binds to TetR
r13,  B->degraded,      delta; //TetR with 2 ATc degrades

s1,   D+S->R1,    r; //TetR binds to yEGFP promoter
s2,   R1->D+S,    rho; //TetR unbinds from yEGFP promoter
s3,   D+R1->R2,   r; //TetR binds to yEGFP promoter
s4,   R2->D+R1,   rho; //TetR unbinds from yEGFP promoter

s7,   basal->basal+M,   l; //yEGFP basal expression
s8,   A+S->A+S+M, m; //yEGFP expression
s9,   M->degraded,      mu; //yEGFP mRNA degraded
s10,  M->M+P,     p; //yEGFP production
s11,  P->degraded,      pi; //yEGFP degradation
```

Box 6.4 NR reactions for Gillespie Simulations in Dizzy code

6.9 NF Dizzy code

```
b=100;//TetR ATc binding rate
l=2;// yEGFP mRNA leakage
lup=2;// TetR mRNA leakage

m=191;//Maximal yEGFP mRNA production
mup=191; //Maximal TetR mRNA production
mu=3.5; //yEGFP mRNA degradation rate
muup=3.5; //yEGFP mRNA degradation rate

p=75; //yEGFP translation rate
d=75/2; //TetR translation rate
pi=0.3; //yEGFP degradation/dilution rate
delta=0.3; //TetR degradation/dilution rate

a=3.7; //TBP binding rate to yEGFP promoter
aup=3.7; //TBP binding rate to TetR promoter
degraded=0; //Number of degraded molecules

alpha=0.28; //TBP unbinding rate from yEGFP promoter
alphaup=0.28; //TBP unbinding rate from TetR promoter
log2=0.6931; //ln(2)
rho=log2/(2/60); //TetR unbinding rate from a tetO2 site
r=0.2; //TetR binding rate to a tetO2 site
diffusion=0.9242; //ATc diffusion rate across the cell membrane
c=70; //Coefficient for ATc influx

basal=1; //Is there basal expression? 1 if the answer is yes.
M=l/mu; //Starting number of yEGFP mRNA molecules
P=M*p/pi; //Starting number of yEGFP molecules
Mup=mup*aup/(aup+alphaup)/muup+lup/muup; //Starting number of TetR mRNA
D=Mup*d/delta; //Starting number of TetR molecules

H=0; //TetR bound to 1 ATc
B=0; //TetR bound to 2 ATc
R1=0; //1 TetR bound to yEGFP promoter
R2=1; //2 TetR bound to yEGFP promoter
S=0; //0 TetR bound to yEGFP promoter
R01=0; //1 TetR bound to TetR promoter
R02=1; //2 TetR bound to TetR promoter
S0=0; //0 TetR bound to TetR promoter

A=1; //TBP bound to yEGFP promoter
Aup=1; //TBP bound to TetR promoter
A0=0; //no TBP bound to yEGFP promoter
Aup0=0; //no TBP bound to TetR promoter

I=0; //ATc in the cell
Ie=10; //ATC in cell media, currently set to 10 ng/ml ATc
```

Box 6.5 NF parameters for Gillespie Simulations in Dizzy code

```
Iinflux,    Ie->Ie+I,   diffusion*c;//ATc influx into the cell
Ioutflux,   I->degraded,      diffusion;//ATc outflux out of the cell

r1,   A->A0,      alpha;//TBP unbinding to yEGFP promoter
r2,   A0->A,      a;//TBP binding yEGFP promoter
r3,   Aup->Aup0,  alphaup;//TBP unbinding to TetR promoter
r4,   Aup0->Aup,  aup;//TBP binding to TetR promoter


r5,   Aup+S0->Aup+S0+Mup,     mup; //TetR mRNA production
r6,   basal->basal+Mup, lup; //TetR basal production
r7,   Mup->degraded,    muup; //TetR mRNA degradation
r8,   Mup->Mup+D, d; //TetR production
r9,   D->degraded,      delta; //TetR degradation

r10,  D+I->H,     2*b; //ATc binds to TetR
r11,  H->degraded,      delta; //TetR with 1 ATc degrades
r12,  H+I->B,     b; //ATc binds to TetR
r13,  B->degraded,      delta; //TetR with 2 ATc degrades

s01,  D+S0->R01,  r; //TetR binds to TetR promoter
s02,  R01->D+S0,  rho; //TetR unbinds from TetR promoter
s03,  D+R01->R02, r; //TetR binds to TetR promoter
s04,  R02->D+R01, rho; //TetR unbinds from TetR promoter

s1,   D+S->R1,    r; //TetR binds to yEGFP promoter
s2,   R1->D+S,    rho; //TetR unbinds from yEGFP promoter
s3,   D+R1->R2,   r; //TetR binds to yEGFP promoter
s4,   R2->D+R1,   rho; //TetR unbinds from yEGFP promoter

s7,   basal->basal+M,   l; //yEGFP basal expression
s8,   A+S->A+S+M, m; //yEGFP expression
s9,   M->degraded,      mu; //yEGFP mRNA degraded
s10,  M->M+P,     p; //yEGFP production
s11,  P->degraded,      pi; //yEGFP degradation
```

Box 6.6 NF reactions for Gillespie Simulations in Dizzy code

6.10 NR and NF template model parameters

Parameter	Value	Source
a, a_{up}	3.7	(156)
b	100	(111)
c	70	Fit to dose-response
d	75/2	Half of yEGFP translation due to dimerization
i,o	0.924	(113)
l, l_{up}	2	Fit from uninduced yEGFP noise
m (WT)	273	(157)
m (D12)	273*0.7	(92,157)
p	75	Fit to saturated noise levels
r	0.2	Fit from yEGFP peak noise
α, α_{up}	0.28	(156)
δ,π	0.3	Experimental cell division rate
μ,μ_{up}	3.5	Generic mRNA degradation rate, (157)
ρ	21	(158)

Table 6.4 *NR and NF Template model parameters.*

6.11 Simplified NF model

The Linear Noise Approximation [14] was applied to the simplified model [35], with chemical species, drift [4], and diffusion matrices [5] defined as:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} a_x F_x(x) - bxy - gx \\ C - bxy - hy \\ a_z F_z(x) + l_z - gz \end{bmatrix}, \text{and} \qquad [114]$$

$$\mathbf{B} = \begin{bmatrix} sa_x F_x(x) + bxy + gx & -bxy & 0 \\ -bxy & C + bxy + hy & 0 \\ 0 & 0 & a_z F_z(x) + l_z + gz \end{bmatrix},$$

Where the promoter response is approximated by the Hill function

$$F(x) = \frac{\theta^n}{\theta^n + x^n}.$$

The Jacobian [13] of **A** is

$$\mathbf{J} = \begin{bmatrix} a_x \dfrac{\partial F_x(x)}{\partial x} - by - g & -bx & 0 \\ -by & -bx - h & 0 \\ a_z \dfrac{\partial F_z(x)}{\partial x} & 0 & -g \end{bmatrix}. \qquad [115]$$

Solving for the steady state values of the molecular species requires solving a polynomial of indeterminate rank. Thus, an analytical solution was not obtained. Solutions for the steady state were obtained numerically by minimizing $\mathbf{A}^T\mathbf{A}$ as a function of **x** by the fminsearch function in Matlab. The parameters used are shown in Table 6.5.

Parameter	Value	Source
a_x, a_z	2500	Obtained from GAL1 expression
b	100	(111)
g	0.3	Experimental cell division rate
h	1	(113)
l	2	Similar to the template model, $\approx l\,p/\mu$
n	1	No cooperativity assumed
θ	5	Low dissociation constant

Table 6.5 Sequestration model parameters

6.12 Fokker-Planck formulation of PF dynamics

I created a model (similar to previously defined models, [35,36] where rtTA was capable of activating both *rtTA* and *yEGFP::ZeoR* promoters. The reaction scheme is:

$$\begin{aligned}
\text{rtTA production:} &\xrightarrow{a_x F(x)+l_x} w \\
\text{inactive rtTA degradation/dilution:} &\; w \xrightarrow{\delta+g} \varnothing \\
\text{active rtTA degradation/dilution:} &\; x \xrightarrow{\delta+g} \varnothing \\
\text{rtTA ATc binding:} &\; w + C \xrightarrow{b} x \\
\text{yEGFP production:} &\xrightarrow{a_z F(x)+l_z} z \\
\text{yEGFP dilution:} &\; z \xrightarrow{g} \varnothing
\end{aligned} \quad , \qquad [116]$$

where the activation level of the P_{TETREG} promoter is given by the increasing Hill function

$$F(x) = \frac{x^n}{\theta^n + x^n}, \qquad [117]$$

while the variables w, x, and z correspond to free and inducer-bound intracellular activator, and reporter concentrations, respectively, and C is a control parameter proportional to the extracellular inducer concentration. Other parameters are a_x and a_z (rtTA and yEGFP::ZeoR protein synthesis rate), b (inducer-activator association rate), δ (rate of rtTA degradation), g (rate of dilution due to cell growth), l (basal expression in the absence of activation), θ (activator-promoter dissociation constant) and n (cooperativity of rtTA-MF for P_{TETREG} activation). In contrast to previous models, which explicitly account for ATc diffusion into the cell, this model makes a less stringent assumption that ATc saturates the cells and is unaffected by rtTA sequestration of ATc. This is justified by the fact rtTA has a lower affinity for ATc that is four orders of magnitude lower than that of TetR (86).

PF sub-population means were modeled by the set of equations:

$$\begin{aligned}
\dot{w} &= aF(x) - bwC - (\delta + g)w + l \\
\dot{x} &= bwC - (\delta + g)x \\
v &= x + w
\end{aligned} \quad . \qquad [118]$$

Parameters were obtained from previously used parameters, literature, and from fitting the model to sub-population means (see Table 6.6, Fig. 6.3.E). The resulting model displayed bistability for ATc concentrations greater than or equal to 1 ng/ml, possibly explaining how two distinct peaks could be seen in PF cells (see Fig. 4.1.B).

Assuming fast ATc binding dynamics, inactive rtTA can be described as

$$w \approx \frac{(\delta + g)x}{bC}. \qquad [119]$$

In terms of total rtTA,

$$v = x\left[1 + \frac{(\delta + g)}{bC}\right]. \quad [120]$$

Thus the dynamics for total rtTA can be written as

$$\dot{v} = \dot{x} + \dot{w} = aF\left(\frac{v}{1 + \frac{(\delta + g)}{bC}}\right) + l - (\delta + g)v. \quad [121]$$

Dividing these terms into production and degradation terms gives the Fokker-Planck variables

$$k_{\rightarrow} = aF\left(\frac{v}{1 + \frac{(\delta + g)}{bC}}\right) + l \quad [122]$$

and

$$k_{\leftarrow} = (\delta + g)v. \quad [123]$$

These terms were variables were used to estimate the distributions of total rtTA in the cell population as (71):

$$\pi(v) = \frac{A}{k_{\rightarrow} + k_{\leftarrow}} \exp\left(2\Omega \int \frac{k_{\leftarrow} - k_{\rightarrow}}{k_{\rightarrow} + k_{\leftarrow}} dv\right) \quad [124]$$

where *A* is a normalizing constant and Ω is a scaling constant for how strongly stochastic processes occur. However the resulting proportions of cells in the two modes diverged strongly from experimental results (see Fig. 6.3.). This divergence is likely due to the fact that low expressor cells grow faster than high expressor cells (see section 4.2.5).

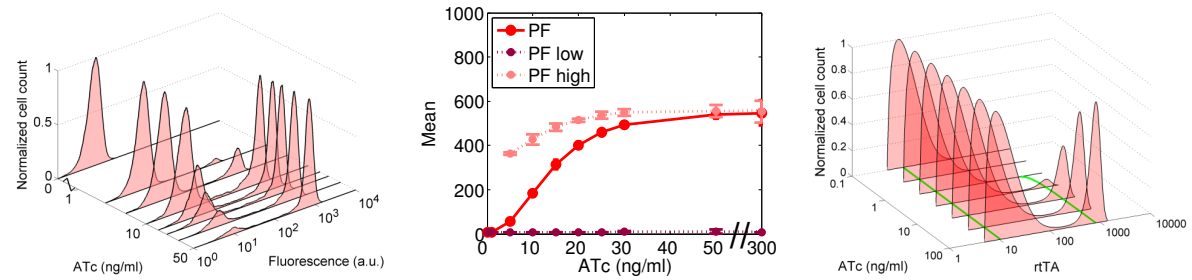


Figure 6.3. Modeling the simplified PF model

A) The fluorescence distributions of PF cells. B) Mean dose-responses of PF cells were used to fit parameters for the simplified the PF model. yEGFP::ZeoR fluorescence is shown by red lines. Low and high sub-population means are shown in purple and pink lines. C) These parameters were then used in the 1-Dimensional Fokker-Planck formulation to predict cell population fluorescence distributions. The solution to the steady state solutions are shown by green lines.

Parameter	Value	Source
a	1650	Obtained from relative expression levels of PF cells to NR cells.
b	0.1	(86)
δ	2.0794	(19)
g	0.24	Experimental cell division rate
l	1.1310	Fit
n	2	Constrained to be > 1 for bistability to occur. Non-integer values cannot be analytically solved
θ	267.5310	Fit
Ω	0.05	Fit

Table 6.6 Parameters for the simplified PF model

6.13 TetR and rtTA dependent growth rates

Gene expression systems based on TetR and yEGFP are only slightly affected by induction (see Fig. 6.4.A). Active rtTA appears to be more toxic than inactive rtTA (see Fig. 6.4.B).

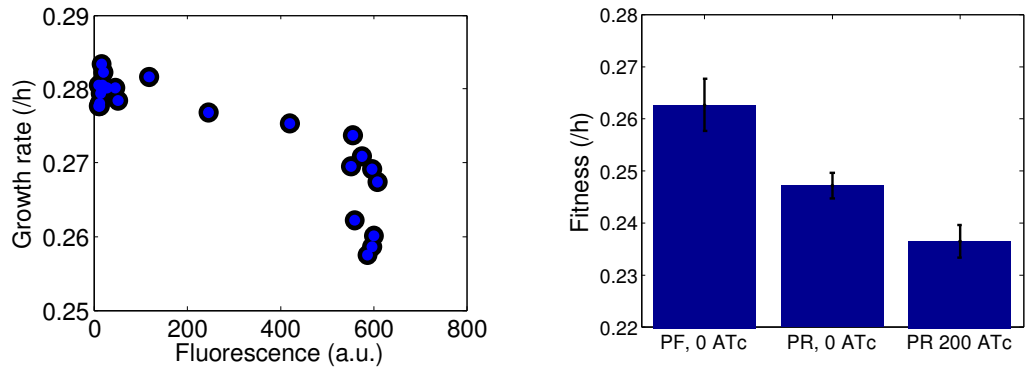


Figure 6.4. yEGFP::ZeoR and rtTA growth retardation

A) Growth rates of NR cells as a function of mean fluorescence. Growth rates are calculated from 2 OD600 measurements of cells grown overnight. B) Growth rates of uninduced PF cells, uninduced PR cells (constitutively expressing rtTA), and PR cells grown in high concentrations of ATc.

6.14 NR and NF yEGFP dynamics

yEGFP is often considered to be a stable protein. Thus, it dilutes out of the cell at the rate of cell division. This was tested for NR and NF cells which were pre-incubated at 0 ng/ml ATc and switched to 500 ng/ml and measured (see Fig. 6.5.A). Complementing this, a second experiment was also performed where NR and NF cells were pre-incubated at 500 ng/ml ATc to 0 ng/ml ATc and measured (see Fig. 6.5.B). The sequestration (section 6.5) and simplified (section 6.11) models for NR and NF suggests that after ATc diffuses in or out of the cells, the growth curves follow an exponential curve with exponential coefficients similar to cellular growth rates.

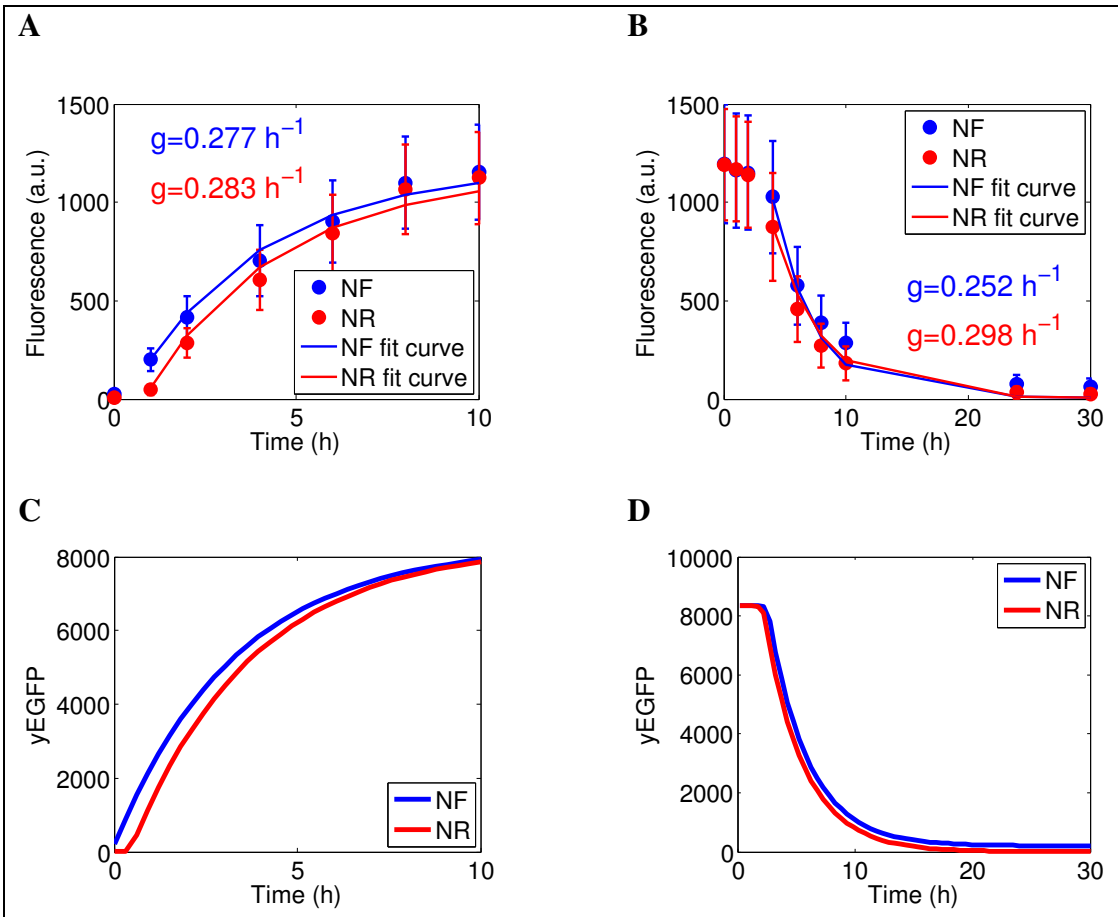
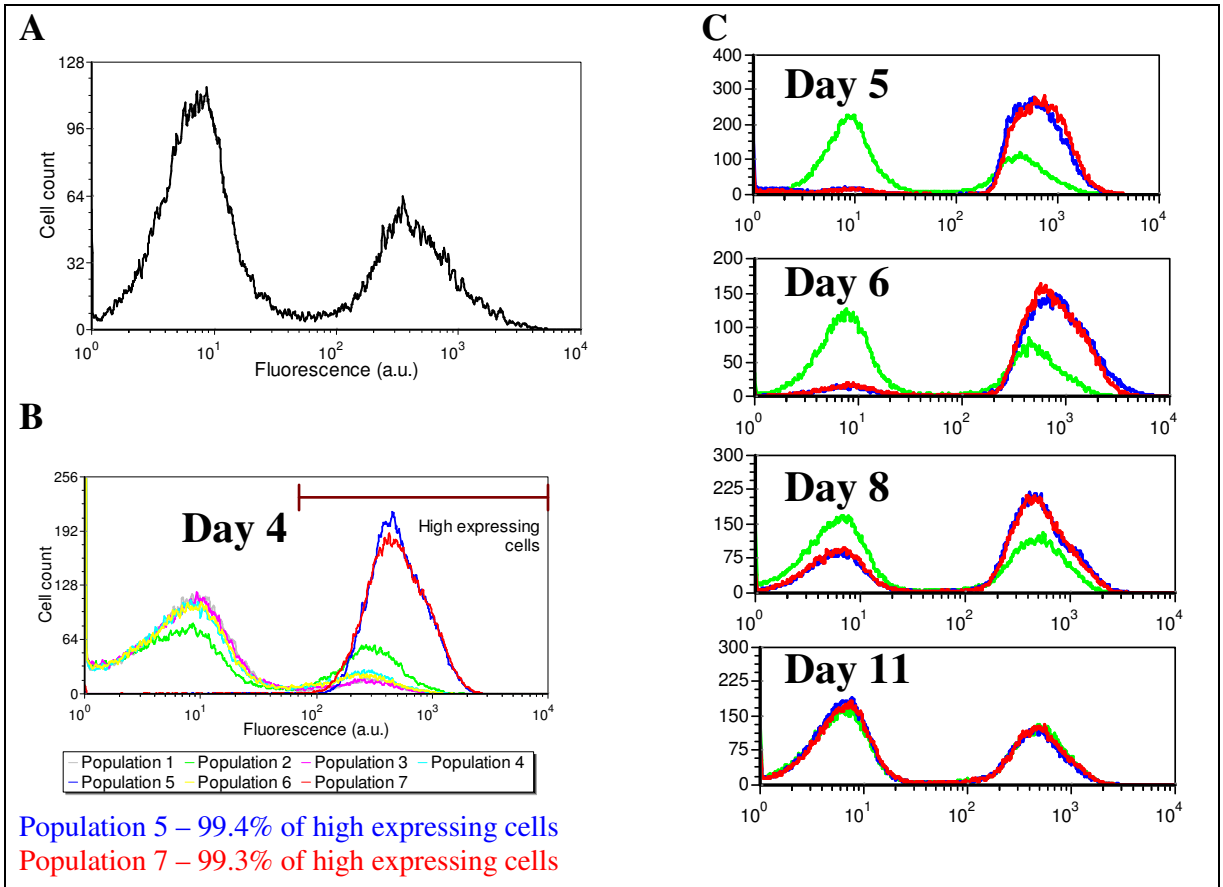


Figure 6.5. yEGFP degradation/dilution rate mirrors yeast growth rate

A)Time course measurements of NF and NR cells incubated in 0 ng/ml ATc switched to 500 ng/ml ATc at 0 hours, and B) time course measurements of NF and NR cells incubated in 500 ng/ml ATc and switched to 0 ng/ml ATc at 0 hours. After the cell's ATc concentrations have responded to the change in ATc concentrations, yEGFP becomes fully expressed or fully repressed and acts according to simple exponential curves. The slope of the exponential curves is comparable to measured cell growth rates supporting the hypothesis that yEGFP is diluting out of the cell. C and D) Corresponding simulations of NF and NR responses to changes in ATc were performed based on the simplified [36] and sequestration [35] models, respectively. The parameters used were $a_x = 2500$ *proteins* h^{-1}*(for NF) and* $a_x = 3600$ *proteins* h^{-1} *(for NR),* $a_z = 2500$*,* $b = 100$*,* $C = 60 * [ATc]$*,* $g = 0.3$*,* $h = 1$*,* $l_x = l_z = 0$*,* $n = 1$*, and* $\theta = 5$*. The WT GAL1 promoter for TetR in NR strains is approximately 1.4 times stronger than the D12 promoter used in NF strains, explaining the differences in two rate constants (91)*

6.15 Purification experiments confirm phenotypic switching

To confirm that cells were switching from high expressor states (instead of having infinitely high memories) Dmitry Nevozhay diluted media containing a bimodal PF cell population into fresh tubes such that the expected concentrations of cells per tube was 0.1 cell/tube. These media were incubated and resuspended every 12 hours. After the 4th day, cell populations were measured. Some tubes retained a predominantly high population which slowly relaxed to a bimodal population. Cells were consistently incubated in 10 ng/ml ATc.



Population 5 – 99.4% of high expressing cells
Population 7 – 99.3% of high expressing cells

Figure 6.6. Purified high expressor cells relax to bimodal populations

Serial dilution experiment for obtaining populations of PF cells grown from single cells. (A) Initial distribution of PF cells grown in 10 ng/ml ATc; (B) Populations of PF cells in different tubes after 4 days during the serial-dilution experiment. Populations 5 and 7 were highly enriched (>99%) in high expressing cells; (C) Monitoring Populations 5 and 7 (blue and red) and Population 2 (green, control bimodal population) demonstrated the strong memory of high expression, as long time was needed for the predominantly high expressing subpopulation switch to the bimodal state.

6.16 Deriving the cellular current

Chemical kinetics and cell growth have been modeled according to the cell population balance (CPB) model (57,159,160).

The density organisms within a small volume corresponding to physical (161,162), chemical (57,160), or phenotypic states (159) may be modeled according to the flux of organisms due to changes in their states as well as due to reproduction. The 1 dimensional formulation of the CPB for chemical reactions has the form:

$$\dot{u}(x,t)+\underbrace{\sum_{i=1}^{k}\left[\frac{\partial}{\partial x}L_i(x)u(x,t)\right]}_{\text{Flux}}+\underbrace{\Gamma(x)u(x,t)}_{\substack{\text{Decrease from}\\ \text{cell division}}}=\underbrace{2\int_{x}^{\infty}\Gamma(x')\psi(x,x')u(x',t)dx'}_{\substack{\text{Increase from}\\ \text{cell division}}}, \qquad [125]$$

where u is the density of cells with the chemical state x at time t, L_i is an operator describing how cells transition states change due to the i^{th} chemical reaction, Γ is the average rate at which cells divide, and ψ is the probability that during cell division the daughter cell will retain x molecules from the parent cell.

The concentration form of the CPB where the cell state is defined by the number of molecules divided by the cell volume ($F = x/V$) has been derived (57) as:

$$\begin{aligned}&\dot{u}(F,V,t)+\frac{\partial}{\partial F}\left\{\left[L_{\rightarrow}(F,t)+L_{\leftarrow}(F,t)-F\frac{\dot{V}(F,V)}{V}\right]u(F,V,t)\right\}\\&+\frac{\partial}{\partial V}\left[\dot{V}(F,V)u(F,V,t)\right]+\Gamma(F,V)u(F,V,t)=2\int_{0}^{\infty}\Gamma(F,V')\omega(V\mid F,V')u(F,V',t)dV'\end{aligned}, \qquad [126]$$

where $\omega(V|F,V')$ is the probability that a mother cell with volume V' will have a daughter cell with volume V depending on the concentration of molecules in the cell. I neglected noise from cell partitioning of volume, and assumed that relative partitioning is not affected by the cell state (i.e., the partition probability is given by the Dirac delta distribution, $\omega(V\mid F,V')=\delta(V'-kV)$, where k is a constant). Additionally, assuming that the change in cell growth is exponential ($\dot{V}(F,V)=\kappa(F)V$) and that there is practically no degradation reaction ($r_{\leftarrow}\approx 0$), the CPB [126] becomes

$$\begin{aligned}&\dot{u}(F,V,t)+\frac{\partial}{\partial F}\left\{\left[L_{\rightarrow}(F,t)-F\kappa(F)\right]u(F,V,t)\right\}\\&+\frac{\partial}{\partial V}\left[\dot{V}(F,V)u(F,V,t)\right]+\Gamma(F,V)u(F,V,t)=2\Gamma(F,kV)u(F,kV,t)\end{aligned}. \qquad [127]$$

Although there may be some dependence on cell state and cell volume, I assumed that the rate of cell division was completely due to the internal state of the cell $\left(\Gamma\left(F,V\right)\rightarrow\Gamma\left(F\right)\right)$. This is justified by the fact that cells with high levels of rtTA tend to be slightly larger, and subsequently are somewhat larger when they reproduce than cells with low rtTA concentrations. Eliminating the volume dependence of growth, [127] becomes:

$$\begin{aligned}&\dot{u}\left(F,V,t\right)+\frac{\partial}{\partial F}\left\{\left[L_{\rightarrow}\left(F,t\right)-F\kappa\left(F\right)\right]u\left(F,V,t\right)\right\}\\&+\frac{\partial}{\partial V}\left[\dot{V}\left(F,V\right)u\left(F,V,t\right)\right]+\Gamma\left(F\right)u\left(F,V,t\right)=2\Gamma\left(F\right)u\left(F,kV,t\right)\end{aligned}. \qquad [128]$$

I next integrated the CPB over all cell volumes. Assuming that

$$\dot{V}\left(F,V=0\right)=0, \qquad [129]$$

$$u\left(F,V=\infty,t\right)=0, \qquad [130]$$

and

$$\tilde{u}\left(F,t\right)=\int_{0}^{\infty}u\left(F,V,t\right)dV, \qquad [131]$$

the CPB [128] is further approximated as

$$\begin{aligned}&\dot{\tilde{u}}\left(F,t\right)+\frac{\partial}{\partial F}\left\{\left[L_{\rightarrow}\left(F,t\right)-F\kappa\left(F\right)\right]\right\}\tilde{u}\left(F,t\right)\\&+\Gamma\left(F\right)\tilde{u}\left(F,t\right)=2\Gamma\left(F\right)\tilde{u}\left(F,t\right)\end{aligned}. \qquad [132]$$

I changed the notation so that the flux terms are denoted by the cellular currents, so that

$$I_{\rightarrow}\left(F,t\right)=L_{\rightarrow}\left(F,t\right)\tilde{u}\left(F,t\right) \qquad [133]$$

and

$$I_{\leftarrow}\left(F,t\right)=F\kappa\left(F\right)\tilde{u}\left(F,t\right). \qquad [134]$$

Defining the density of cells as being equal to the probability distribution of cells ($p(F,t)$) times the total number of cells (N_T),

$$\tilde{u}\left(F,t\right)=N_{T}\,p\left(F,t\right) \qquad [135]$$

and integrating the CPB over the interval of fluorescence concentrations from φ_1 to φ_2 gives

$$\dot{N}\left(\varphi_1,\varphi_2,t\right)=\underbrace{I_{\rightarrow}\left(\varphi_1,t\right)-I_{\rightarrow}\left(\varphi_2,t\right)}_{\text{Increasing fluorescence}}-\underbrace{I_{\leftarrow}\left(\varphi_1,t\right)+I_{\leftarrow}\left(\varphi_2,t\right)}_{\text{Decreasing fluorescence}}+\underbrace{g\left(\varphi_1,\varphi_2,t\right)N\left(\varphi_1,\varphi_2,t\right)}_{\text{Fitness}}, \qquad [136]$$

where $g(\varphi_1,\varphi_2,t)=\dfrac{\int_{\varphi_1}^{\varphi_2}\Gamma(F)p(F,t)dF}{\int_{\varphi_1}^{\varphi_2}p(F,t)dF}$, and $N(\varphi_1,\varphi_2)=\int_{\varphi_1}^{\varphi_2}\tilde{u}(F,t)dF$. This equation is similar to the cellular current equation in section 4.2.7. Numerical simulations of the cellular current are given in section 6.17.

6.17 Numerical verification of cellular current

In order to numerically test how well cellular currents predict rise and fall rates, I performed a set of cell simulations (see Fig. 6.7A). Cells produced a protein whose concentration could consequently increase protein production, forming a positive feedback loop. Cells grew at a rate depending on cell volume and protein concentrations. Cell cycles progressed at a rate dependent on protein concentrations. When a cell cycle reached a threshold, two daughters were created in the cells place, with half the (integer) number of proteins and volume of the parent. If protein numbers were odd, one of the daughters received an odd number of proteins, while the second daughter received an even number of proteins. Cells were randomly killed to allow longer simulations to be performed. Pseudocode is shown in box 6.7.

These simulations were used to create a stationary distribution of cells (see Fig. 6.7B), which were used along with cellular currents to estimate the fraction of cells that left a boundary over time (i.e., *r* and *f*). These escape rates were directly measured from the simulations, as the fraction of cells that crossed a boundary at any particular time point. These simulated escape rates were compared to the cellular current escape rates (see Fig. 6.7C and D).

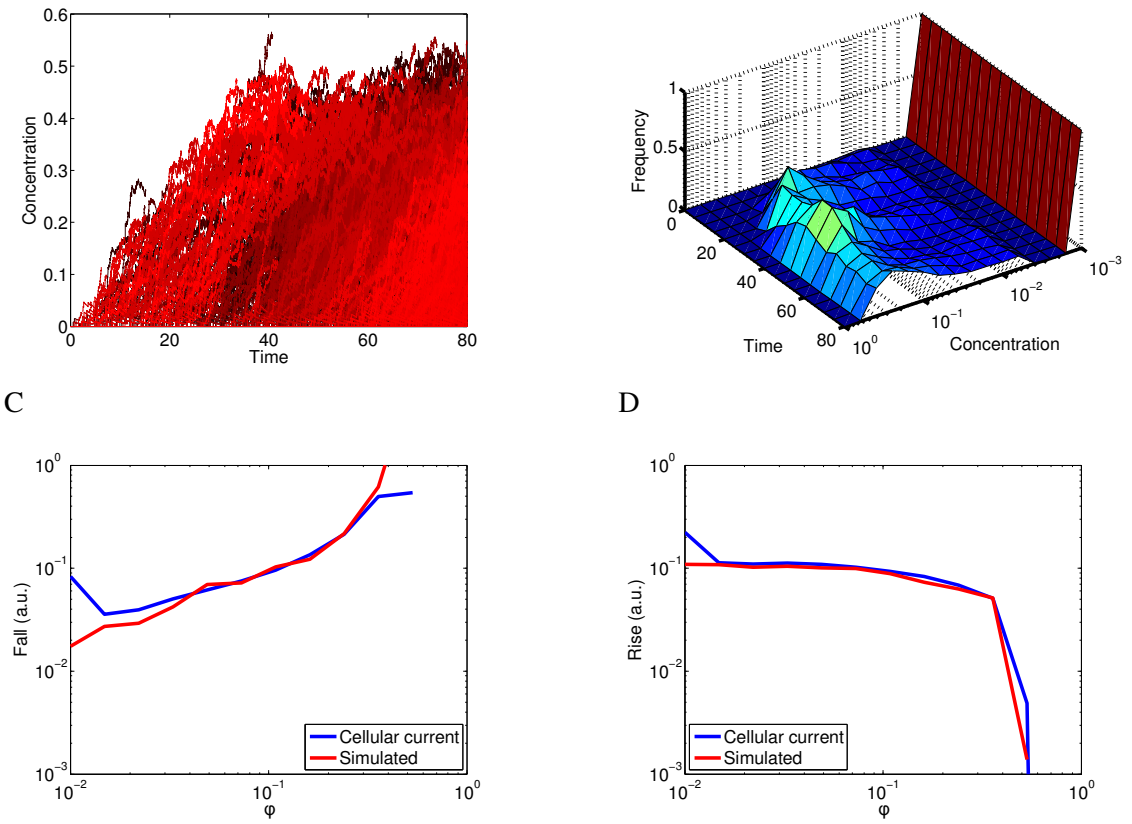


Figure 6.7. Comparison of cellular current predictions to numerical simulations

A) More than 10000 reproducing cells were simulated over a period of 80 time units. B) Frequency histograms of protein concentrations over time. These simulations were used to estimate the fall (C) and rise (D) rates across a set of arbitrary boundaries (φ), which were compared with cellular current estimates. The two methods diverge at low boundaries, due to the property that protein numbers are integers, and are not evenly split during cell division. The two methods also diverge at high boundaries, as the probability distribution becomes less precise.

I assumed that two parameters describe the chemical state of the cell, X and V, where X is the number of proteins, and V is the size of the cell. The concentration of proteins is $x = X/V$. The protein production rate is $p(x) = a\frac{x}{x+\theta} + l$, where a is maximum production, θ is the concentration where half maximal production occurs, and l is the basal protein production rate. The growth/division rate is $\Gamma(x) = g_0 \frac{\rho}{\rho + x} + g_{\min}$, where $g_{\min}$ is the minimum division/growth rate, g_0 is the increase in growth due to the cell being in an ideal chemical state, ρ is the concentration of proteins where the contribution of growth from g_0 is half its maximum value. The pseudo-code for this process is:

[**x**, **t**]=simulation(X,V,t,Δt,$t_{\max}$)
$C \leftarrow 0$
$D \leftarrow 0$
$i \leftarrow 0$
while true
$X(t_{i+1}) \leftarrow X(t_i) + Pois(p(x(t_i))\Delta t)$
$V(t_{i+1}) \leftarrow V(t_i) + \Gamma(x(t_i))\Delta t$
$C \leftarrow C + \Gamma(x(t_i))\Delta t$
$D \leftarrow Pois(\delta \Delta t)$
$x(t_{i+1}) \leftarrow X(t_{i+1})/V(t_{i+1})$
$t_{i+1} \leftarrow t_i + \Delta t$
$i \leftarrow i + 1$
if $t > t_{\max}$
return $(\{\mathbf{x}\}, \{\mathbf{t}\})$
end
if $D \geq 1$
return $(\{\mathbf{x}\}, \{\mathbf{t}\})$
end
if $C \geq \ln(2)$
$\mathbf{d_1}, \mathbf{t_1} \leftarrow$simulation($\lceil P(t)/2 \rceil$, $V(t)/2$, t, Δt, $t_{\max}$)
$\mathbf{d_2}, \mathbf{t_2} \leftarrow$simulation($\lfloor P(t)/2 \rfloor$, $V(t)/2$, t, Δt, $t_{\max}$)
return $(\{\mathbf{x}, \mathbf{d_1}, \mathbf{d_2}\}, \{\mathbf{t}, \mathbf{t_1}, \mathbf{t_2}\})$
end
end

Pois denotes a Poissonian random variable. I used the parameters $a = 10$, $\theta = 0.1$, $l = 0.1$, $g_0 = 0.2$, $\rho = 0.1$, $g_{\min} = 0.1$, $\delta = 0.2$, $\Delta t = 0.001$, and $t_{\max} = 80$. Cells were initialized at $P_0 = 0$, $V_0 = 100$. Unitless rates were used.

Box 6.7 Pseudocode for cell simulations

6.18 Comparison of memory estimates

I compared two methods of estimating memory in cell sub-populations; the cellular current method (see section 4.3.5), and fitting the two state model to cells relaxing to equilibrium (see section 4.2.8). Cellular current based memory estimates were obtained from 3 cell strains using flow cytometry to obtain population distributions and from measuring the exponential growth curves of cells over 4 days (see Fig. 6.8.). These 3 strains are 1) NR, which constitutively expresses the *tetR* repressor, which subsequently blocks a downstream yEGFP reporter strains (see section 2.2.1, Fig. 3.2. for details), 2) PF cells which have *rtTA* (transcriptional activator) controlling its own expression, as well as the downstream *yEGFP::zeoR* fusion gene (see section 4.2.1, Fig. 4.1 for details), and 3) PF-PG cells which have *rtTA* controlling its own expression as well as the *yEGFP:: PDR5* fusion gene. The PDR5 gene encodes for a multi-drug resistance protein that pumps drugs out of the cell (163). It is unknown if it acts to pump out doxycycline or ATc.

Dmitry Nevozhay and Junchen Diao performed four sorting experiments on the 3 strains of cells at different levels of induction. The ratio of sorted cells over time was used to fit the two state model, and estimate the switching rates of these cells at different levels of induction (see Fig. 6.9.).

Strain	Induction	2 state fit		cellular current	
		τ_L (hours)	τ_H (hours)	τ_L (hours)	τ_H (hours)
PF	5 ng/ml ATc	58	10^{13}	53	169
PF	10 ng/ml ATc	24	990	20	216
NR	80 ng/ml ATc	11	12	14	9
PF-PG	0.05 mg/ml Dox	8	5	11	7

Table 6.7 Comparison of memory estimates for 3 cell strains

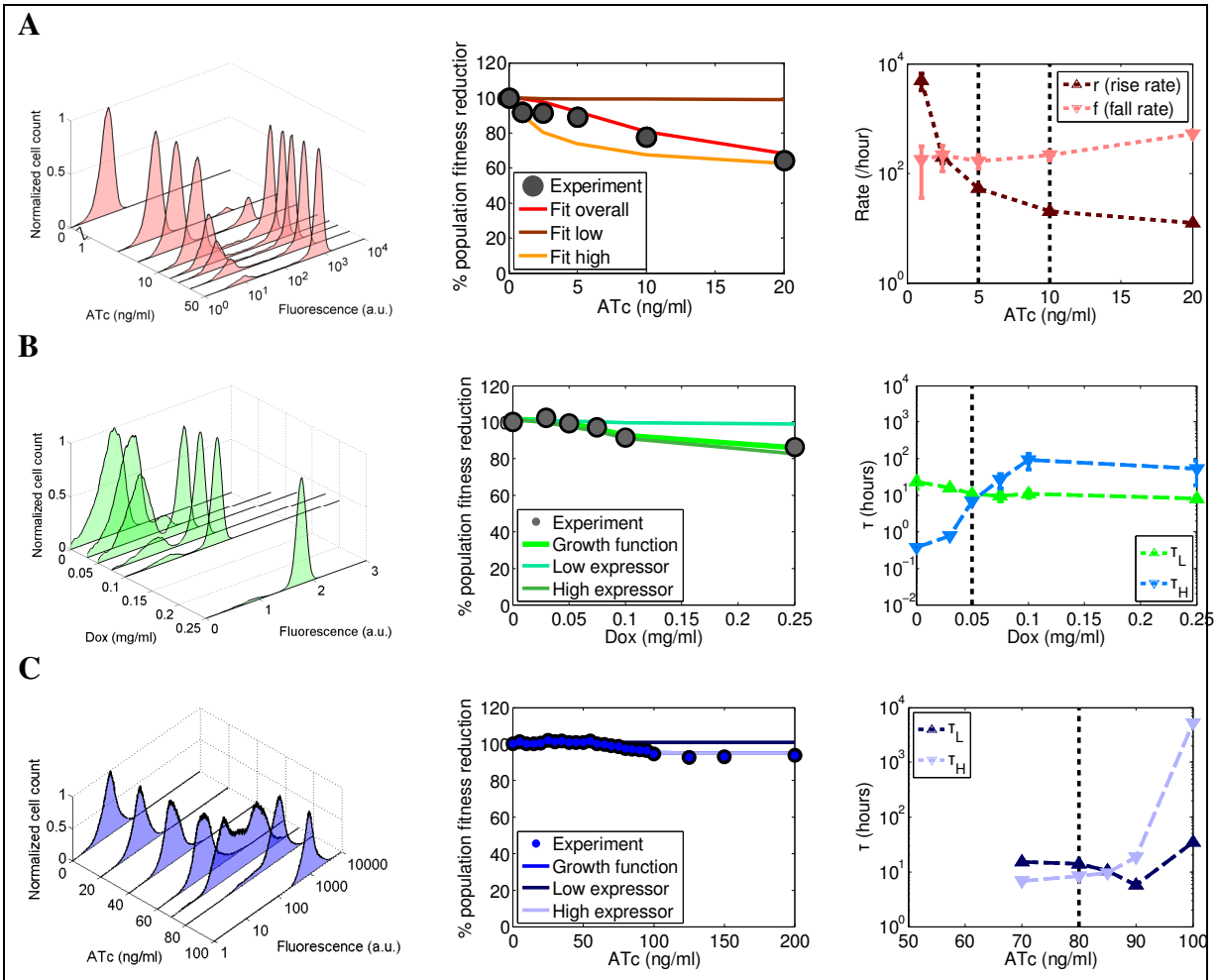


Figure 6.8. Cellular current memory estimates

Cellular current estimates of memory were performed for the 3 cell strains PF (A), PF-PG (B), and NR (C). The cell density functions were obtained by flow cytometry (left column). Cell growth curves were then established over a variety of inducer concentrations (middle column). These density functions and growth curve functions were used to estimate cellular memory over wide ranges of inducers (right column). The vertical dashed lines denote concentrations where cellular memories were estimated by fitting the ratio of sorted cells over time to a two state cell dynamics model (see Fig. 6.9. for fits, section 4.2.8 for details of model).

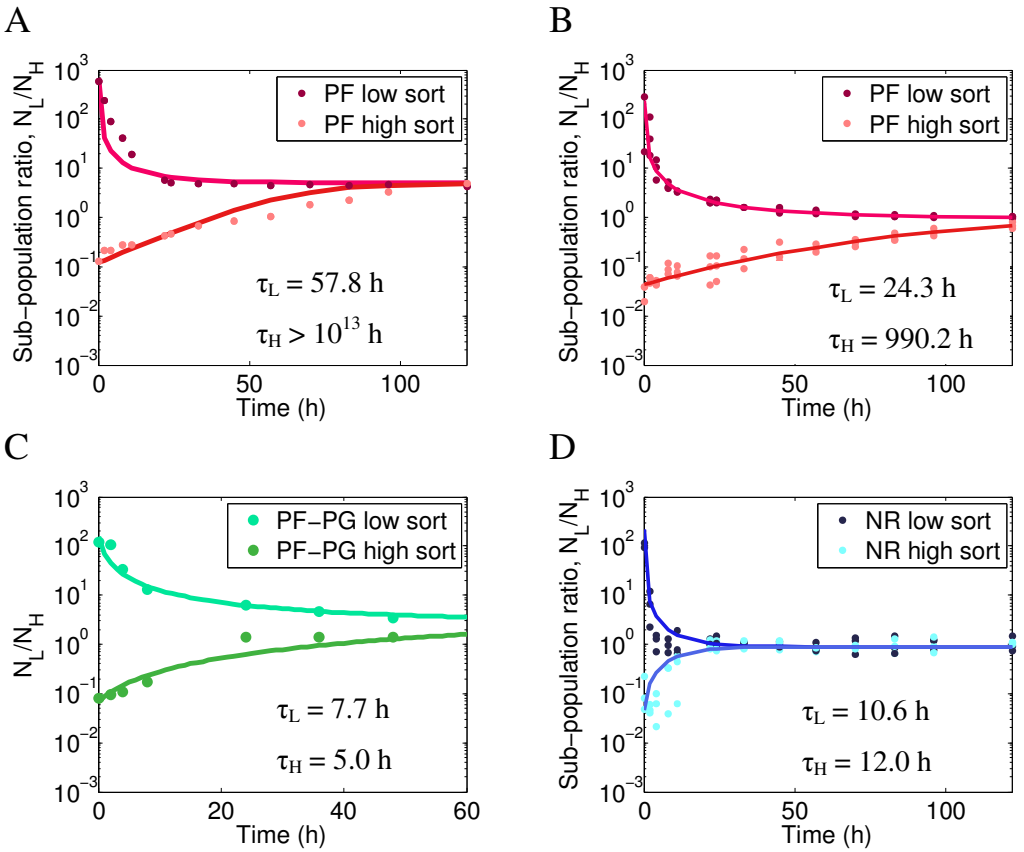


Figure 6.9. 2 state fit memory estimates

Fit memory estimates of PF cells incubated at 5 ng/ml (A) and 10 ng/ml (B) ATc concentrations were performed. Memory estimates were also performed for sorted PF-PG cells at 0.05 mg/ml Doxycycline (C), and NR cells at 80 ng/ml ATc (D).

The 2 state model fits and the cellular current models gave comparable estimates of cellular memory (see table 6.7). Some of the discrepancies may be attributed to inducer degradation over time (since these measurements were made on different days), distributional variation (since distributions change depending on cell growth rates and other factors), low cell counts (if cell counts in a flow cytometry bin average less than 1, the distributions are smoothed to make the probability distribution non-zero), and the possibility that sorted cells may not have constant switching rates over time. The last factor is likely most responsible for the major discrepancies between the PF measurements, since cells at intermediate fluorescence values were discarded. During cell sorting high expressor cells had to travel further (biochemically) to reach their boundaries than cells not subjected to cell sorting.

6.19 Stochastic simulation of PF cells

In order to understand how PF gene distributions arose due to induction of the rtTA positive feedback circuit, I created a biochemical reaction model for cells similar to those used for NR and NF cell types (see Fig. 6.10A). The internal state of the cell determined how quickly cell division occurred (see Fig. 6.10B). These simulations were able to reproduce experimental behavior seen in PF cells (see Fig. 6.10C). The Dizzy code used to simulate the internal cell state is given in box *6.8* and *6.9*.

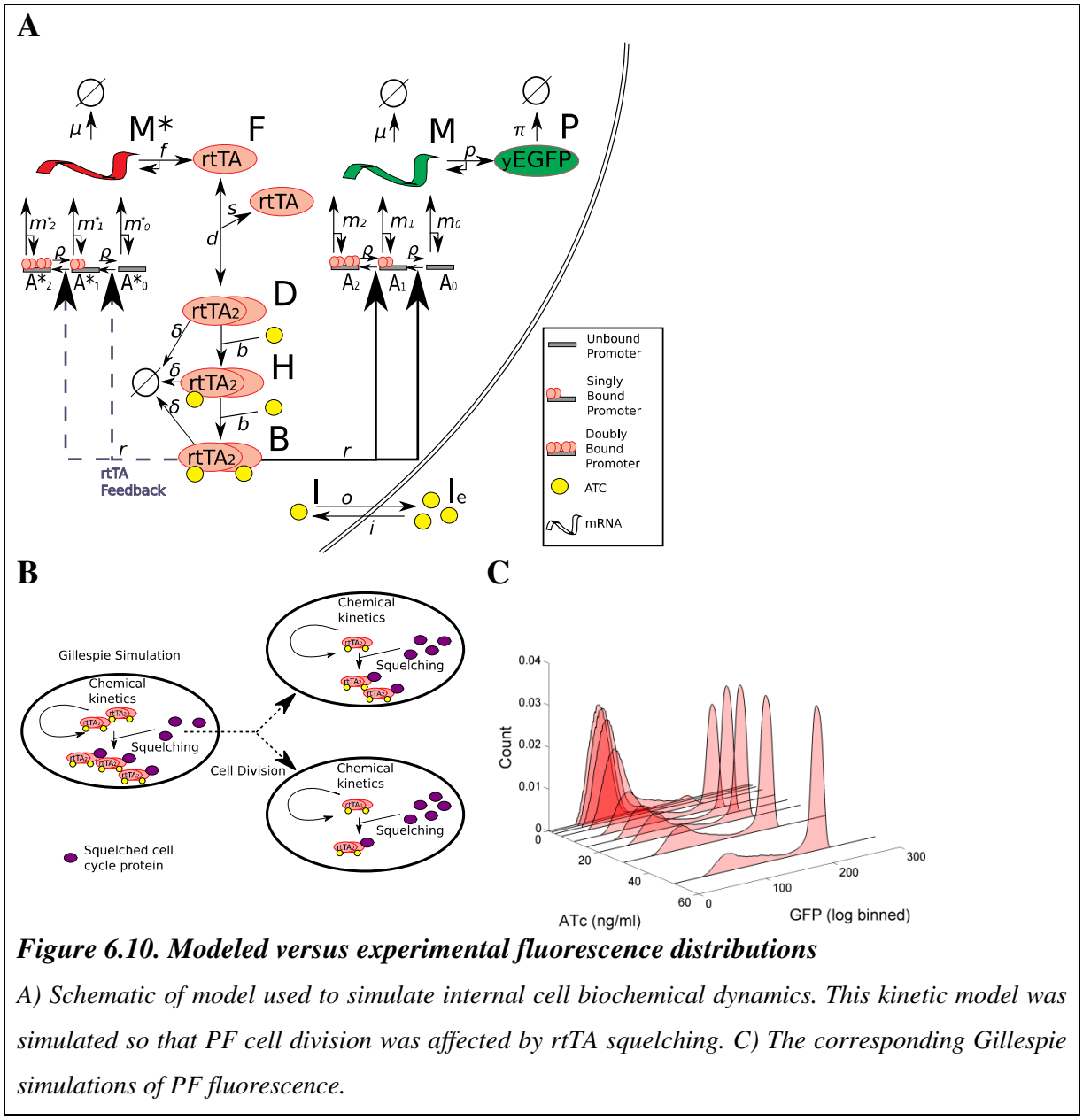


Figure 6.10. Modeled versus experimental fluorescence distributions

A) Schematic of model used to simulate internal cell biochemical dynamics. This kinetic model was simulated so that PF cell division was affected by rtTA squelching. C) The corresponding Gillespie simulations of PF fluorescence.

```
log2=0.6931;
b=0.005;//ATc rtTA binding rate
l=0.3;//yEGFP basal transcription rate
lup=l;//rtTA basal transcription rate

m=180;//yEGFP transcription rate
mup=180;//rtTA transcription rate
mu=3.5;//yEGFP mRNA degradation rate
muup=3.5;//rtTA mRNA degradation rate
p=75;//yEGFP translation rate
d=75/2;//rtTA translation + dimerization rate
pi=log2/26;//yEGFP degradation rate
delta=log2/(20/60);//rtTA degradation rate
degraded=0;//molecular species for degraded molecules

rho=log2/(6.67/60);//rtTA-tetO2 unbind rate
r=0.06;//rtTA-tetO2 binding rate
diffusion=log2/(45/60);//ATc diffusion across membrane
c=60;//ATc influx parameter

Wtotal=2500;//Proteins involved in cell cycle
squelch=0.5;//rtTA affinity for cell cycle affecting proteins
death=0.17;//cell death rate
growth=0.24/(0.15*0.15*Wtotal);//Rate at which cell cycle proceeds

W=Wtotal;//cell cycle protein
BW=0;//W bound to rtTA

M=0;//yEGFP mRNA
P=0;//yEGFP
Mup=0;//rtTA mRNA
D=0;//rtTA dimers
H=0;//rtTA bound to ATc
B=0;//rtTA bound to 2 ATc

R1=0;//rtTA bound yEGFP promoter
R2=0;//2 rtTAs bound to yEGFP promoter
S=1;//no rtTA bound to yEGFP promoter
RW1=0;//rtTA bound yEGFP promoter
RW2=0;//2 rtTAs bound to yEGFP promoter
R1up=0;//rtTA bound rtTA promoter
R2up=0;//2 rtTAs bound to rtTA promoter
Sup=1;//no rtTA bound to rtTA promoter
RW1up=0;//rtTA bound rtTA promoter
RW2up=0;//2 rtTAs bound to rtTA promoter

cell=1;//Living cell
dead=0;//Dead cell
cycle=1/(0.15*0.15);// cell division occurs when cycle=90
constitutive=1;//1 if ATc is in media
I=0;//Free ATc in cell
Ie=10;//ATc outside of cell
```

Box 6.8 PF constants for Gillespie Simulations in Dizzy code

```
//ATc influx and outflux
Iin,   constitutive->constitutive+I,    diffusion*c*Ie;
Iout,  I->degraded,              diffusion;

//rtTA binding to rtTA promoter dynamics
t1,    B+Sup->R1up,        r;
t2,    R1up->B+Sup,        rho;
t3,    B+R1up->R2up,       r;
t4,    R2up->B+R1up,       rho;
t12,   BW+Sup->RW1up,             r;
t22,   RW1up->BW+Sup,             rho;
t32,   BW+RW1up->RW2up,    r;
t42,   RW2up->BW+RW1up,    rho;

//rtTA central dogma dynamics
TetR1, Sup->Sup+Mup,       lup;
TetR2, R1up->R1up+Mup,            mup;
TetR3, R2up->R2up+Mup,            mup;
TetR4, RW1up->RW1up+Mup,   mup;
TetR5, RW2up->RW2up+Mup,   mup;
TetR6, Mup->degraded,             muup;
TetR7, Mup->Mup+D,         d;
TetR8, D->degraded,        delta;

//rtTA-ATc binding
r9,    D+I->H,                    2*b;
r10,   H->I,               delta;
r11,   H+I->B,                    b;
r12,   B->I+I,                    delta;

//rtTA binding to yEGFP promoter dynamics
s1,    B+S->R1,            r;
s2,    R1->B+S,            rho;
s3,    B+R1->R2,           r;
s4,    R2->B+R1,           rho;
s12,   BW+S->RW1,          r;
s22,   RW1->BW+S,          rho;
s32,   BW+RW1->RW2,        r;
s42,   RW2->BW+RW1,        rho;

//yEGFP central dogma dynamics
GFP1,  S->S+M,                    l;
GFP2,  R1->R1+M,           m;
GFP3,  R2->R2+M,           m;
GFP4,  RW1->RW1+M,         m;
GFP5,  RW2->RW2+M,         m;
GFP6,  M->degraded,        mu;
GFP7,  M->M+P,                    p;
GFP8,  P->degraded,        pi;

//Squelching dynamics
c1,    B+W->BW,            squelch;
c2,    BW->W+I+I,          delta;
c3,    cell->dead,         death;
c4,    cell+W->cell+W+cycle,      growth;
```

Box 6.9 PF reactions for Gillespie Simulations in Dizzy code

6.20 PF template model parameters

Parameter	Value	Source
b	0.005	(86)
c	60	Fit to dose-response
d	75/2	Half of yEGFP translation due to dimerization
i,o	0.924	(113)
l	0.3	Fit from uninduced yEGFP noise
l_{up}	1	Fit from uninduced rtTA noise
m,m_{up}	180	(157), rtTA expression was observed to be 2/3 NR expression
p	75	Fit to saturated noise levels
r	0.06	(86)
α, α_{up}	0.28	(156)
δ	ln(2)/(20/60)	(19,114)
π	ln(2)/26	(108)
μ,μ_{up}	3.5	Generic mRNA degradation rate, (157)
ρ	6.2	(158)

Table 6.8 *P Template model parameters.*

6.21 Loewe additivity does not affect PF fitness predictions

Instead of assuming Bliss independence to predict cell fitness, where the reduction in fitness was:

$$\gamma = \gamma_Z \times \gamma_x,$$

I asked whether Loewe additivity changed the predictions. Loewe additivity was defined as

$$\gamma = \frac{\alpha\chi}{\alpha\chi + \alpha Z_i + \beta x + \iota Z_i x}$$

where ι is the cooperativity of the two toxic molecules. A positive value of ι implies synergism, while a negative value of ι implies antagonism between the two molecules. In this instance, Bliss independence is equivalent to $\iota = 1$. The resulting predictions are given in Fig. 6.11.

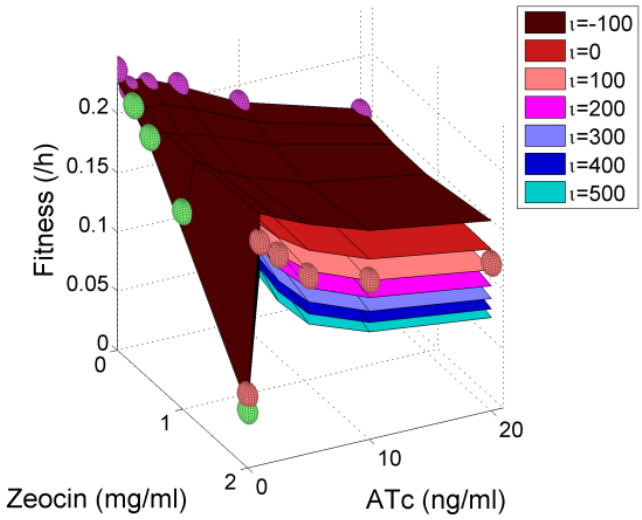


Figure 6.11. Cell population fitness predictions do not change with Loewe additivity

Fitness prediction based on memory and Loewe additivity give similar results to those based on Bliss independence. Experimental results may suggest slight synergism between rtTA toxicity and Zeocin toxicity at ι = 100. Experimental data points are shown by spheres, with purple spheres corresponding to 0 Zeocin, green spheres corresponding to 0 ATc, and red spheres corresponding to 2 mg/ml Zeocin.

6.22 The "sweet spot" occurs when switching rates are less than growth rates

I asked whether the observed sweet spot in PF cells was sensitive to the low and high expressor memories. As the switching rates become faster, memory decreases. I scaled the switching rates by a constant k ($r' = kr, f' = kf$), decreasing memory of the two states. As the switching rates became comparable to growth rates (at k = 10), the sweet spot was dampened, and disappeared at high switching rates (see Fig. 6.12.).

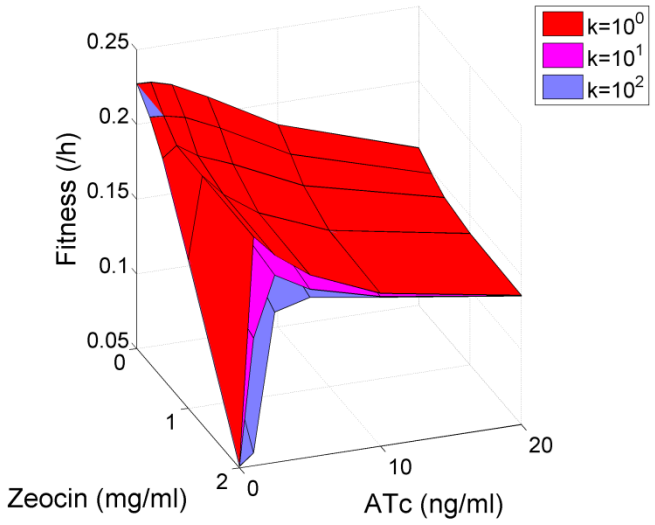


Figure 6.12. Cell fitness as memory is decreased

The estimated switching rates were scaled by the factor k. As switching rates reach parity with growth rates (k = 10), the sweet spot disappears.

6.23 Microscopy verification of cellular memory

Cells were measured over 40 hours using brightfield and fluorescence microscopy measurements. Using custom tracking algorithms cells were tracked. Low fluorescence expressors switched to high expression at a much greater rate than high expressors transitioned to the low state (see Fig. 6.13.). Using a threshold of 2000 arbitrary fluorescence units, I measured the memories of the two cell states. I counted the percentage of cells that crossed fluorescence boundary per unit time, to observe the rise and fall rates of r = 28.7×10-3 h^{-1} and f = 3.9×10-3 h^{-1}, consistent with cellular current predictions (see section 6.18).

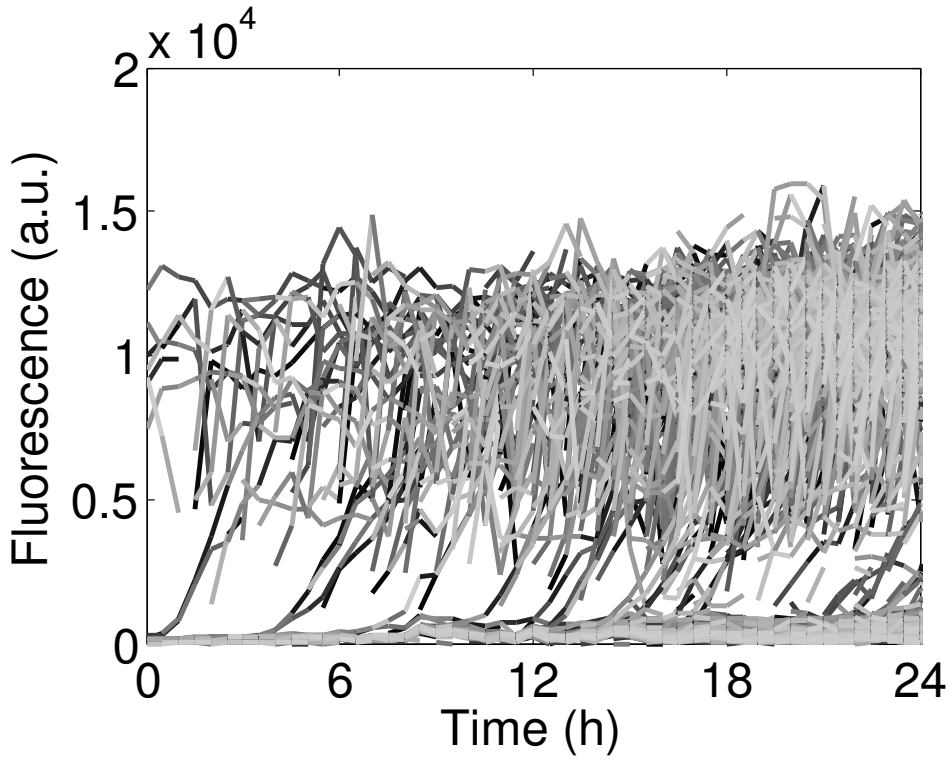


Figure 6.13. Time course of bimodal PF cell population

PF cells were grown in doxycycline concentrations so that the approximate ratio of low to high yEGFP::ZeoR expressors was 1:1. Cells transition to high expressor states more frequently than they transition to low expressor states.

Bibliography

1. Collins, F. S. 1999. Medical and societal consequences of the Human Genome Project. New England Journal of Medicine 341:28-37.

2. Babcock, E. B. and R. E. Clausen. 1918. Genetics in relation to agriculture: McGraw-Hill.

3. Altman, G. H., F. Diaz, C. Jakuba, T. Calabro, R. L. Horan, J. Chen, H. Lu, J. Richmond, and D. L. Kaplan. 2003. Silk-based biomaterials. Biomaterials 24:401-416.

4. McLaren, J. S. 2000. Future renewable resource needs: will genomics help? Journal of Chemical Technology and Biotechnology 75:927-932.

5. Crick, F. 1970. Central dogma of molecular biology. Nature 227:561-563.

6. Galton, F. 1876. The history of twins, as a criterion of the relative powers of nature and nurture. The Journal of the Anthropological Institute of Great Britain and Ireland 5:391-406.

7. Novick, A. and M. Weiner. 1957. Enzyme induction as an all-or-none phenomenon. Proceedings of the National Academy of Sciences of the United States of America 43:553-566.

8. Newman, J. R., S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman. 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:840-846.

9. Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329:533-538.

10. Veening, J. W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen. 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proceedings of the National Academy of Sciences of the United States of America 105:4393-4398.

11. Cohen, A. A., N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, T. Danon, N. Perzov, and U. Alon. 2008. Dynamic proteomics of individual cancer cells in response to a drug. Science 322:1511-1516.

12. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences of the United States of America 99:12795-12800.

13. Hasty, J., J. Pradines, M. Dolnik, and J. J. Collins. 2000. Noise-based switches and amplifiers for gene expression. Proceedings of the National Academy of Sciences of the United States of America 97:2075-2080.

14. Arkin, A., J. Ross, and H. H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage {lambda}-infected Escherichia coli cells. Genetics 149:1633–1648.

15. Balázsi, G., A. van Oudenaarden, and J. J. Collins. 2011. Cellular decision making and biological noise: from microbes to mammals. Cell 144:910-925.

16. Raj, A. and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216-226.

17. Thattai, M. and A. Van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences of the United States of America 98:8614-8619.

18. Ochab-Marcinek, A. and M. Tabaka. 2010. Bimodal gene expression in noncooperative regulatory systems. Proceedings of the National Academy of Sciences of the United States of America 107:22096-22101.

19. To, T. L. and N. Maheshri. 2010. Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science 327:1142-1145.

20. Kussell, E. and S. Leibler. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. Science 309:2075-2078.

21. Thattai, M. and A. Van Oudenaarden. 2004. Stochastic gene expression in fluctuating environments. Genetics 167:523–530.

22. Kalmar, T., C. Lim, P. Hayward, S. Muñoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, and A. M. Arias. 2009. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biology 7:e1000149.

23. Galhardo, R. S., P. Hastings, and S. M. Rosenberg. 2007. Mutation as a stress response and the regulation of evolvability. Critical Reviews in Biochemistry and Molecular Biology 42:399-435.

24. Weinberger, L. S., R. D. Dar, and M. L. Simpson. 2008. Transient-mediated fate determination in a transcriptional circuit of HIV. Nature Genetics 40:466-470.

25. Acar, M., A. Becskei, and A. van Oudenaarden. 2005. Enhancement of cellular memory by reducing stochastic transitions. Nature 435:228-232.

26. Acar, M., J. T. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. Nature Genetics 40:471-475.

27. Giepmans, B. N. G., S. R. Adams, M. H. Ellisman, and R. Y. Tsien. 2006. The fluorescent toolbox for assessing protein location and function. Science's STKE 312:217.

28. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. Science 297:1183-1186.

29. Andrianantoandro, E., S. Basu, D. K. Karig, and R. Weiss. 2006. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology 2.

30. Sauro, H. M. 2008. Modularity defined. Molecular Systems Biology 4.

31. Ellis, T., X. Wang, and J. J. Collins. 2009. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nature Biotechnology 27:465-471.

32. Zhang, J., R. E. Campbell, A. Y. Ting, and R. Y. Tsien. 2002. Creating new fluorescent probes for cell biology. Nature Reviews Molecular Cell Biology 3:906-918.

33. May, T., H. Hauser, and D. Wirth. 2006. Current status of transcriptional regulation systems. Cytotechnology 50:109-119.

34. Brown, R. 1828. XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philosophical Magazine Series 2 4:161-173.

35. Einstein, A. 1905. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Annalen der Physik 17:549-560.

36. Raser, J. M. and E. K. O'Shea. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010-2013.

37. Elf, J., G. W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316:1191-1194.

38. McAdams, H. H. and A. Arkin. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends in Genetics 15:65-69.

39. Kierzek, A. M., J. Zaim, and P. Zielenkiewicz. 2001. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. Journal of Biological Chemistry 276:8165-8172.

40. Blake, W. J., M. Kærn, C. R. Cantor, and J. J. Collins. 2003. Noise in eukaryotic gene expression. Nature 422:633-637.

41. Raser, J. M. and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. Science 304:1811.

42. Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. Nature Genetics 31:69-73.

43. van Kampen, N. G. 1992. Stochastic processes in physics and chemistry: North Holland.

44. Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22:403-434.

45. Gillespie, D. T. 2000. The chemical Langevin equation. The Journal of Chemical Physics 113:297-306.

46. Elf, J. and M. Ehrenberg. 2003. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Research 13:2475–2484.

47. Iacus, S. M. 2008. Simulation and inference for stochastic differential equations: with R examples: Springer Verlag.

48. Uhlenbeck, G. E. and L. S. Ornstein. 1930. On the theory of the Brownian motion. Physical Review 36:823–841.

49. Scott, M., B. Ingalls, and M. Kærn. 2006. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. Chaos 16:026107-026107-026115.

50. Ito, Y. and K. Uchida. 2010. Formulas for intrinsic noise evaluation in oscillatory genetic networks. Journal of Theoretical Biology 267:223-234.

51. Paulsson, J. 2004. Summing up the noise in gene networks. Nature 427:415-418.

52. Pedraza, J. M. and A. van Oudenaarden. 2005. Noise propagation in gene networks. Science 307:1965-1969.

53. Khanin, R. and D. J. Higham. 2008. Chemical Master Equation and Langevin regimes for a gene transcription model. Theoretical Computer Science 408:31-40.

54. Bar-Even, A., J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai. 2006. Noise in protein expression scales with natural protein abundance. Nature Genetics 38:636-643.

55. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81:2340-2361.

56. Eden, E., N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo, and U. Alon. 2011. Proteome half-life dynamics in living human cells. Science's STKE 331:764-768.

57. Stamatakis, M. 2010. Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches. Chemical Engineering Science 65:1008-1015.

58. Lu, T., D. Volfson, L. Tsimring, and J. Hasty. 2004. Cellular growth and division in the Gillespie algorithm. Systems Biology 1:121-128.

59. Charlebois, D. A., J. Intosalmi, D. Fraser, and M. Kærn. 2011. An Algorithm for the Stochastic Simulation of Gene Expression and Heterogeneous Population Dynamics. Arxiv preprint arXiv:1110.6469.

60. Fantes, P. 1977. Control of cell size and cycle time in Schizosaccharomyces pombe. Journal of Cell Science 24:51-67.

61. Kar, S., W. T. Baumann, M. R. Paul, and J. J. Tyson. 2009. Exploring the roles of noise in the eukaryotic cell cycle. Proceedings of the National Academy of Sciences of the United States of America 106:6471–6476.

62. Bhalla, U. S., P. T. Ram, and R. Iyengar. 2002. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science 297:1018-1023.

63. Liu, R. Y., D. Fioravante, S. Shah, and J. H. Byrne. 2008. cAMP response element-binding protein 1 feedback loop is necessary for consolidation of long-term synaptic facilitation in Aplysia. The Journal of Neuroscience 28:1970-1976.

64. Wolf, D. M., L. Fontaine-Bodin, I. Bischofs, G. Price, J. Keasling, and A. P. Arkin. 2008. Memory in microbes: quantifying history-dependent behavior in a bacterium. PLoS One 3:e1700.

65. Turner, B. M. 2002. Cellular memory and the histone code. Cell 111:285-291.

66. Xiong, W. and J. E. Ferrell. 2003. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature 426:460-465.

67. Igoshin, O. A., R. Alves, and M. A. Savageau. 2008. Hysteretic and graded responses in bacterial two-component signal transduction. Molecular Microbiology 68:1196-1215.

68. Demongeot, J., M. Kaufman, and R. Thomas. 2000. Positive feedback circuits and memory. Comptes Rendus de l'Acadₑmie des Sciences-Series III-Sciences de la Vie 323:69-79.

69. Bagowski, C. P. and J. E. Ferrell. 2001. Bistability in the JNK cascade. Current Biology 11:1176-1182.

70. Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. 2005. Gene regulation at the single-cell level. Science 307:1962-1965.

71. Bialek, W. Stability and noise in biochemical switches; 2001. The MIT Press. p 103-109.

72. Klein, C. T. 1998. Hysteresis-driven structure formation in biochemical networks. Journal of Theoretical Biology 194:263-274.

73. Gillespie, D. T. 1981. On the calculation of mean first passage times for simple random walks. The Journal of Chemical Physics 74:5295-5299.

74. Seshadri, V., B. J. West, and K. Lindenberg. 1980. Analytic theory of extrema. III. Results for master equations with application to unimolecular decomposition. The Journal of Chemical Physics 72:1145-1151.

75. Ebeling, W. and I. M. Sokolov. 2005. Statistical thermodynamics and stochastic theory of nonequilibrium systems: World Scientific Publishing Co. Inc.

76. Risken, H. 1996. The Fokker-Planck equation: Methods of solution and applications: Springer Verlag.

77. Süel, G. M., R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. 2007. Tunability and noise dependence in differentiation dynamics. Science 315:1716-1719.

78. Ellermeier, C. D., E. C. Hobbs, J. E. Gonzalez-Pastor, and R. Losick. 2006. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. Cell 124:549-559.

79. González-Pastor, J. E., E. C. Hobbs, and R. Losick. 2003. Cannibalism by sporulating bacteria. Science 301:510-513.

80. Nguyen, T., Q. G. Phan, L. P. Duong, K. P. Bertrand, and R. E. Lenski. 1989. Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of Escherichia coli K12. Molecular Biology and Evolution 6:213-225.

81. Wolf, D. M., V. V. Vazirani, and A. P. Arkin. 2005. Diversity in times of adversity: probabilistic strategies in microbial survival games. Journal of Theoretical Biology 234:227-253.

82. Wolf, D. M., V. V. Vazirani, and A. P. Arkin. 2005. A microbial modified prisoner's dilemma game: how frequency-dependent selection can lead to random phase variation. Journal of Theoretical Biology 234:255-262.

83. Degenkolb, J., M. Takahashi, G. Ellestad, and W. Hillen. 1991. Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for

tetracycline analogs with the Tet repressor. Antimicrobial Agents and Chemotherapy 35:1591–1595.

84. Meier, I., L. V. Wray, and W. Hillen. 1988. Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. The EMBO Journal 7:567-572.

85. Takahashi, M., L. Altschmied, and W. Hillen. 1986. Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements. Evidence for divalent metal ion requirement and energy transfer. Journal of Molecular Biology 187:341-348.

86. Kamionka, A., J. Bogdanska Urbaniak, O. Scholz, and W. Hillen. 2004. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. Nucleic Acids Research 32:842–847.

87. Gossen, M., S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard. 1995. Transcriptional activation by tetracyclines in mammalian cells. Science 268:1766-1769.

88. Furth, P. A., L. St Onge, H. Böger, P. Gruss, M. Gossen, A. Kistner, H. Bujard, and L. Hennighausen. 1994. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proceedings of the National Academy of Sciences of the United States of America 91:9302-9306.

89. Akagi, K., M. Kanai, H. Saya, T. Kozu, and A. Berns. 2001. A novel tetracycline-dependent transactivator with E2F4 transcriptional activation domain. Nucleic Acids Research 29:e23.

90. Morimoto, M. and R. Kopan. 2009. rtTA toxicity limits the usefulness of the SP-C-rtTA transgenic mouse. Developmental Biology 325:171-178.

91. Blake, W. J., G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. Molecular Cell 24:853-865.

92. Murphy, K. F., G. Balázsi, and J. J. Collins. 2007. Combinatorial promoter design for engineering noisy gene expression. Proceedings of the National Academy of Sciences of the United States of America 104:12726-12731.

93. Yao, F., C. Theopold, D. Hoeller, O. Bleiziffer, and Z. Lu. 2006. Highly efficient regulation of gene expression by tetracycline in a replication-defective herpes simplex viral vector. Molecular Therapy 13:1133-1141.

94. Alon, U. 2007. An introduction to systems biology: design principles of biological circuits: CRC Press.

95. Dublanche, Y., K. Michalodimitrakis, N. Kümmerer, M. Foglierini, and L. Serrano. 2006. Noise in transcription negative feedback loops: simulation and experimental analysis. Molecular Systems Biology 2.

96. Hooshangi, S., S. Thiberge, and R. Weiss. 2005. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. Proceedings of the National Academy of Sciences of the United States of America 102:3581-3586.

97. Rosenfeld, N., M. B. Elowitz, and U. Alon. 2002. Negative autoregulation speeds the response times of transcription networks. Journal of Molecular Biology 323:785-793.

98. Huang, C. Y. and J. E. Ferrell. 1996. Ultrasensitivity in the mitogen-activated protein kinase cascade. Proceedings of the National Academy of Sciences of the United States of America 93:10078-10083.

99. Becskei, A. and L. Serrano. 2000. Engineering stability in gene networks by autoregulation. Nature 405:590-593.

100. Becskei, A., B. Séraphin, and L. Serrano. 2001. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. The EMBO Journal 20:2528-2535.

101. May, T., L. Eccleston, S. Herrmann, H. Hauser, J. Goncalves, and D. Wirth. 2008. Bimodal and hysteretic expression in mammalian cells from a synthetic gene circuit. PLoS One 3:590-593.

102. Ferrell, J. E. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Current Opinion in Cell Biology 14:140-148.

103. Smolen, P., D. A. Baxter, and J. H. Byrne. 2000. Mathematical modeling of gene networks. Neuron 26:567-580.

104. Thomas, R. 1994. The role of feedback circuits: positive feedback circuits are a necessary condition for positive real eigenvalues of the Jacobian matrix. Berichte der Bunsengesellschaft für Physikalische Chemie 98:1148-1151.

105. Buchler, N. E. and F. R. Cross. 2009. Protein sequestration generates a flexible ultrasensitive response in a genetic network. Molecular Systems Biology 5.

106. Buchler, N. E. and M. Louis. 2008. Molecular titration and ultrasensitivity in regulatory networks. Journal of Molecular Biology 384:1106-1119.

107. Adams, R. M., K. F. Murphy, X. Wang, G. Balázsi, and J. J. Collins. 2010. Tuning and controlling gene expression noise in synthetic gene networks. Nucleic Acids Research 38:2712-2726.

108. Corish, P. and C. Tyler-Smith. 1999. Attenuation of green fluorescent protein half-life in mammalian cells. Protein Engineering 12:1035-1040.

109. Ramsey, S., D. Orrell, and H. Bolouri. 2005. Dizzy: stochastic simulation of large-scale genetic regulatory networks. Journal of Bioinformatics and Computational Biology 3:415-436.

110. Adams, R. M., D. Nevozhay, K. F. Murphy, K. Josi , and G. Balázsi. 2009. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene

expression. Proceedings of the National Academy of Sciences of the United States of America 106:5123–5128.

111. Schubert, P., K. Pfleiderer, and W. Hillen. 2004. Tet repressor residues indirectly recognizing anhydrotetracycline. European Journal of Biochemistry 271:2144-2152.

112. Hillen, W., C. Gatz, L. Altschmied, K. Schollmeier, and I. Meier. 1983. Control of expression of the Tn10-encoded tetracycline resistance genes*:: Equilibrium and kinetic investigation of the regulatory reactions. Journal of Molecular Biology 169:707-721.

113. Sigler, A., P. Schubert, W. Hillen, and M. Niederweis. 2000. Permeation of tetracyclines through membranes of liposomes and Escherichia coli. European Journal of Biochemistry 267:527-534.

114. Resch, M., H. Striegl, E. M. Henssler, M. Sevvana, C. Egerer-Sieber, E. Schiltz, W. Hillen, and Y. A. Muller. 2008. A protein functional leap: how a single mutation reverses the function of the transcription regulator TetR. Nucleic Acids Research 36:4390-4401.

115. Von Bertalanffy, L. 1950. The theory of open systems in physics and biology. Science 111:23-29.

116. Meadows, D. H. 2008. Thinking in systems: A primer: Chelsea Green Publishing.

117. Bratsun, D., D. Volfson, L. S. Tsimring, and J. Hasty. 2005. Delay-induced stochastic oscillations in gene regulation. Proceedings of the National Academy of Sciences of the United States of America 102:14593-14598.

118. Marquez-Lago, T. T. and J. Stelling. 2010. Counter-intuitive stochastic behavior of simple gene circuits with negative feedback. Biophysical Journal 98:1742-1750.

119. Becskei, A. 2009. Linearization through distortion: a new facet of negative feedback in signalling. Molecular Systems Biology 5.

120. Sauro, H. M. and B. Ingalls. 2007. MAPK cascades as feedback amplifiers. Arxiv preprint arXiv:0710.5195.

121. Ermentrout, B., M. Pascal, and B. Gutkin. 2001. The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators. Neural Computation 13:1285-1310.

122. Adams, R. M., D. Nevozhay, E. Van Itallie, M. R. Bennett, and G. Balázsi. 2012. Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit. PLoS Computational Biology 8:e1002480.

123. Urlinger, S., U. Baron, M. Thellmann, M. T. Hasan, H. Bujard, and W. Hillen. 2000. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. Proceedings of the National Academy of Sciences of the United States of America 97:7963-7968.

124. Kosugi, Y., J. Ikebe, N. Shitara, and K. Takakura. 1986. Graphical presentation of multidimensional flow histogram using hexagonal segmentation. Cytometry 7:291-294.

125. Shapiro, H. 2003. Flow cytometry: problems, parameters, probes, and principles. In Practical Flow Cytometry. 4th ed. 18.

126. Gilbert, D., D. M. Heery, R. Losson, P. Chambon, and Y. Lemoine. 1993. Estradiol-inducible squelching and cell growth arrest by a chimeric VP16-estrogen receptor expressed in Saccharomyces cerevisiae: suppression by an allele of PDR1. Molecular and Cellular Biology 13:462-472.

127. Baron, U., M. Gossen, and H. Bujard. 1997. Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. Nucleic Acids Research 25:2723–2729.

128. Dumas, P., M. Bergdoll, C. Cagnon, and J. Masson. 1994. Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. The EMBO Journal 13:2483–2492.

129. Sugiyama, M., T. Kumagai, H. Matsuo, Z. Alam Bhuiyan, K. Ueda, H. Mochizuki, N. Nakamura, and J. E. Davies. 1995. Overproduction of the bleomycin-binding proteins from bleomycin-producing Streptomyces verticillus and a methicillin-resistant Staphylococcus aureus in Escherichia coli and their immunological characterisation. FEBS Letters 362:80-84.

130. Nowak, M. A. 2006. Evolutionary Dynamics: Exploring the Equations of Life: Belknap Press.

131. Lenski, R. E. and M. Travisano. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proceedings of the National Academy of Sciences of the United States of America 91:6808-6814.

132. Fong, S. S., A. R. Joyce, and B. O. Palsson. 2005. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Research 15:1365-1372.

133. Nur, N. 1987. Population Growth Rate and the Measurement of Fitness: A Critical Reflection. Oikos 48:338-341.

134. Orr, H. A. 2009. Fitness and its role in evolutionary genetics. Nature Reviews Genetics 10:531-539.

135. Charlebois, D. A., N. Abdennur, and M. Kærn. 2011. Gene expression noise facilitates adaptation and drug resistance independently of mutation. Arxiv preprint arXiv:1110.6467.

136. Zhuravel, D., D. Fraser, S. St-Pierre, L. Tepliakova, W. L. Pang, J. Hasty, and M. Kærn. 2010. Phenotypic impact of regulatory noise in cellular stress-response pathways. Systems and Synthetic Biology 4:105-116.

137. Goldoni, M. and C. Johansson. 2007. A mathematical approach to study combined effects of toxicants in vitro: evaluation of the Bliss independence criterion and the Loewe additivity model. Toxicology in vitro 21:759-769.

138. Greco, W. R., G. Bravo, and J. C. Parsons. 1995. The search for synergy: a critical review from a response surface perspective. Pharmacological Reviews 47:331-385.

139. Chickarmane, V. and C. Peterson. 2008. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. PLoS One 3:e3478.

140. Orr, H. 2005. Theories of adaptation: what they do and don't say. Genetics of Adaptation 123:3-13.

141. Borenstein, E., I. Meilijson, and E. Ruppin. 2006. The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes. Journal of Evolutionary Biology 19:1555-1570.

142. Veening, J. W., W. K. Smits, and O. P. Kuipers. 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annual Review of Microbiology 62:193-210.

143. Klumpp, S., Z. Zhang, and T. Hwa. 2009. Growth rate-dependent global effects on gene expression in bacteria. Cell 139:1366-1375.

144. Tan, C., P. Marguet, and L. You. 2009. Emergent bistability by a growth-modulating positive feedback circuit. Nature Chemical Biology 5:842-848.

145. Van Melderen, L. and M. Saavedra De Bast. 2009. Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genetics 5:e1000437.

146. Rotem, E., A. Loinger, I. Ronin, I. Levin-Reisman, C. Gabay, N. Shoresh, O. Biham, and N. Q. Balaban. 2010. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. Proceedings of the National Academy of Sciences of the United States of America 107:12541-12546.

147. Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. Science 305:1622-1625.

148. Nguyen, T. N., Q. G. Phan, L. P. Duong, K. P. Bertrand, and R. E. Lenski. 1989. Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of Escherichia coli K12. Molecular Biology and Evolution 6:213-225.

149. Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. Science's STKE 305:1622-1625

150. Allison, K. R., M. P. Brynildsen, and J. J. Collins. 2011. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. Nature 473:216-220.

151. Sharma, S. V., D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, and M. A. Fischbach. 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141:69-80.

152. Basu, S., D. Karig, and R. Weiss. 2003. Engineering signal processing in cells: Towards molecular concentration band detection. Natural Computing 2:463-478.

153. Komurov, K. and M. White. 2007. Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Molecular Systems Biology 3.

154. Levy, E. D. and J. B. Pereira-Leal. 2008. Evolution and dynamics of protein interactions and networks. Current Opinion in Structural Biology 18:349-357.

155. Cagatay, T., M. Turcotte, M. B. Elowitz, J. Garcia-Ojalvo, and G. M. Suel. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell 139:512-522.

156. Perez-Howard, G. M., P. A. Weil, and J. M. Beechem. 1995. Yeast TATA binding protein interaction with DNA: fluorescence determination of oligomeric state, equilibrium binding, on-rate, and dissociation kinetics. Biochemistry 34:8005-8017.

157. García-Martínez, J., A. n. Aranda, and J. E. Pérez-Ortín. 2004. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Molecular Cell 15:303-313.

158. Hillen, W., C. Gatz, L. Altschmied, K. Schollmeier, and I. Meier. 1983. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. Journal of Molecular Biology 169:707-721.

159. Henson, M. A. 2003. Dynamic modeling of microbial cell populations. Current Opinion in Biotechnology 14:460-467.

160. Stamatakis, M. and K. Zygourakis. 2010. A mathematical and computational approach for integrating the major sources of cell population heterogeneity. Journal of Theoretical Biology 266:41-61.

161. Levin, S. A. 1976. Population dynamic models in heterogeneous environments. Annual Review of Ecology and Systematics 7:287-310.

162. Newman, W. I. and C. Sagan. 1981. Galactic civilizations: Population dynamics and interstellar diffusion. Icarus 46:293-327.

163. Egner, R., Y. Mahe, R. Pandjaitan, and K. Kuchler. 1995. Endocytosis and vacuolar degradation of the plasma membrane-localized Pdr5 ATP-binding cassette multidrug transporter in Saccharomyces cerevisiae. Molecular and Cellular Biology 15:5879–5887.

Vita

Rhys Michael Adams was born in Spring, Texas on December 16, 1982. Rhys received his Bachelor of Arts in computer science in 2003, with a minor in physics. In 2006 Rhys obtained a Master of Science in Bioinformatics. He joined the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences in the fall of 2006.