Signal Processing and Decision Making in Single Cells

by

Jerome Thomas Mettetal II

B.S., Physics
Virginia Polytechnic Institute and State University, 2003

B.S., Mathematics
Virginia Polytechnic Institute and State University, 2003

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

Massachusetts Institute of Technology

February 2008

Signature of author: _____
Department of Physics
January 10, 2008

Certified by: _____
Alexander van Oudenaarden
Associate Professor of Physics
Thesis Supervisor

Accepted by: _____
Thomas Greytak
Professor of Physics
Associate Department Head of Education

Signal Processing and Decision Making in Single Cells

by

Jerome Thomas Mettetal II

Submitted to the Department of Physics
on January 11, 2008, in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Physics

Abstract

Cells are not simple passive observers oblivious to their environment, but sense and adapt to environmental changes in order to thrive. In addition to sensing the presence of signals in the environment, cells can extract information relating to the dynamics and spatial location of these signals and implement a response to these extracellular perturbations. This work examines a variety of signal-processing and decision-making processes across several different organisms.

To explore the connection between biological network topology and temporal signal processing, we study how periodic signals are propagated in the Hog1 osmotic-response pathway of the budding yeast *Saccharomyces cerevisiae*. Utilizing systems identification tools from control engineering, we study how the cells rapidly and robustly maintain osmotic homeostasis. By measuring the expression level of key proteins we begin to understand how fluctuating environments regulate gene expression.

The *lac* operon in *Escherichia coli* has the ability to display a bistable, "all-or-nothing" response to sugar. To understand how noise drives transitions between these two stable states, we measure switching dynamics in a population of cells. A simple model is constructed that can make predictions about system behavior unavailable from a deterministic model. Further, by measuring individual switching events in a similar bistable system implemented in the Galactose utilization pathway of *Saccharomyces cerevisiae*, we find that correlations in switching times of related individuals can be explained in terms of correlations in levels of key regulatory proteins.

Many single celled organisms, such as the slime mold *Dictyostelium discoideum*, can sense and respond to concentration gradients of extracellular signaling molecules. We find that the cells' ability to detect an extracellular signal is influenced by an asymmetric intracellular signal, which varies in direction and magnitude from cell-to-cell. Further, a model that accounts for both signals predicts the observed population response to directed stimuli.

Finally, we explore a "bet-hedging" strategy for fluctuating environments with an engineered population of *Saccharomyces cerevisiae* cells that randomly switch between two phenotypes. Each phenotype is fit to one of two alternating environments. We find that to optimize fitness, cells must tune the phenotypic transition rates in accordance with the rate of environmental transitions.

Thesis Supervisor: Alexander van Oudenaarden
Title: Associate Professor of Physics

Table of Contents

This work is dedicated to my parents, Ann and Jerome, who have instilled in me a love of learning without which this work would not have been possible.

1 Introduction

Cells do not simply observe their environment passively; instead they constantly react to and interact with it. In order to thrive they must continually adapt to environmental changes and utilize available information about potential harm or benefit conferred by extracellular conditions. They have an extensive array of mechanisms to sense and process signals from the external environment, make decisions based on these signals, and finally implement a response. Specialized network modules of interacting proteins and genes execute these various tasks and are capable of processing temporal signals (e.g. changes in concentration of environmental factors) as well as spatial signals such as gradients of metabolite or cell-to-cell signaling molecules. Not only must these networks be highly specific, they must also be able to function robustly in the presence of noise arising from the inherent stochastic effects that result from the discrete nature of biochemical processes.

This work will explore several aspects of signal processing and decision-making in single cells. Chapter 2 provides examples of a few key concepts used extensively in later chapters, such as the link between network dynamics and structure, decision making in a simple gene-regulatory network, and cell-to-cell variability (noise) introduced through stochastic gene expression. Chapter 3 explores the link between network structure and dynamic signal processing, while Chapter 4 focuses on the effects of cell-to-cell variability on decision making in both temporal and spatial signaling pathways. Finally, Chapter 5 examines decision making at the level of a population by asking how a population of cells can improve fitness and "hedge its bets" by utilizing biochemical noise to increase diversity in times of uncertainty. The remainder of this introduction describes many of the individual problems addressed in later chapters and briefly summarizes the key results.

Probing Network Dynamics with Periodic Stimuli

Cells use networks of interacting proteins and genes to perform desired functions, and the topology of these networks has a large impact on how that cellular system will process and respond to time-dependent signals. Recent work has focused on better understanding the network's structure and its dynamic properties. The dynamic properties of many simple network structural motifs have been well studied both experimentally and theoretically (Alon 2007). While endogenous systems are often

composed of simple recurrent motifs, these natural systems are almost always much more complex, suggesting that a full understanding of network function cannot be reached from a simple reductionist perspective. Further complicating the analysis of endogenous networks, the reactions in signaling cascades operate over a wide variety of timescales ranging from rapid ligand-protein interactions to slower downstream gene expression. Thus, it is often not clear which pathway components and reactions dominate a system's output dynamics and should be included in modeling efforts. In Chapter 3.1 we explore directly the link between network structure and signal processing in a well-studied regulatory network. We find a noteworthy relation between the network structure and the dynamic response of the system. In addition, we are able to utilize the signal processing properties of the network under specific perturbations to better understand the general properties of the network structure and function.

Mitogen-activated protein kinase (MAPK) pathways are highly conserved from yeast to humans and are responsible for mediating a cellular response to various extracellular signals, including stress and developmental cues. To learn how signals are processed in a MAPK cascade, we begin by using periodic stimuli to experimentally measure the frequency response of the osmo-adaptation pathway in *Saccharomyces cerevisiae* (Hohmann 2002). This network contains multiple feedback loops, which act to maintain intracellular osmolyte concentrations in response to changes in the extracellular osmolarity. These feedbacks utilize diverse mechanisms such as protein-protein interactions, gene expression, metabolic activity and membrane channel regulation.

We constructed a flow chamber apparatus capable of applying square wave pulses of osmolyte to cells with temporal resolution on the order of seconds while simultaneously imaging cells to quantify the system activity. Using the chamber, the network's output (assayed by nuclear translocation of the MAPK Hog1) is measured in response to a broad frequency range of square wave signals of the osmolyte NaCl. System-identification methods from control engineering are used to infer a concise predictive model of the dynamics of the MAPK Hog1. Possible model architectures that can give rise to the observed signaling dynamics are explored and then compared with known genetic and biochemical data. We find that the dynamics of osmo-adaptation is dominated by a fast-acting negative feedback through Hog1 that does not require synthesis of new proteins.

Given that *de novo* protein synthesis is not important for cells to adapt to osmotic shocks, it is natural to ask what the role of new protein synthesis is in the osmotic shock

response. This was accomplished by measuring the response of cells to pulses of osmolyte when cells were unable to produce new proteins. We found that at elevated osmo-shocks the additional, much slower, negative feedback acting through gene expression allows cells to adapt faster to future stimuli once cells have already adapted to previous stimuli. This analysis sheds light on how the multiple feedback structure of the osmo-adaptation network allows cells to respond rapidly and robustly to fluctuating extracellular environments.

Gene Expression in Fluctuating Environments

Gene expression is often the cornerstone of cellular responses to a variety of environmental stimuli. For example, as we explore in Chapter 3.1, when presented with an increase in extracellular osmolyte, cells synthesize proteins responsible for producing intracellular osmolytes, thereby counteracting the increase in extracellular osmotic pressure. Cells must decide how much protein to produce given a specific extracellular signal; if they produce too little they will not effectively balance out the extracellular signal, whereas if they produce too much, they will overshoot the desired response. Several factors complicate the decision of how much protein to produce. First, protein production is not an instantaneous process. From the initiation of transcription, it can take up to 30 to 60 minutes for the proteins to become fully functional because of the delays in transcription, translation, protein folding and post-translational modification. Further, once proteins are produced, they are often degraded slowly. This means that any newly synthesized proteins likely remain in the cell for long periods of time. This can lead to significant overcompensation if the external stimulus is short lived, which can be potentially toxic to the cell. Finally, proteins production is not "free"; cells must be sure that the benefit of the new proteins justifies the resources used to producing them.

We chose to measure how cells decide what quantity of proteins to produce when presented with a temporally changing environment. As a reporter of gene expression, green fluorescent protein (GFP) was incorporated under the control of a promoter that is activated in response to increases in extracellular osmolarity. Cells were then exposed to a slow but steadily increasing concentration of extracellular osmolyte, while the amount of GFP was measured in a population of cells. We find that a larger concentration of protein is produced when cells are exposed to more rapid increases in osmotic pressure than when the increase is gradual. This finding is consistent with a model of gene expression where cells are attempting to predict the

concentration of extracellular osmolyte at some point in the future by measuring the current rate of change. We combine this analysis with the analysis of the activation of the upstream signaling cascade presented in Chapter 3.1 and discuss the relative contribution to signal processing for upstream and downstream components.

Discrete Decisions and Stochastic Transitions

In some cases, rather than choosing and fine-tuning a response to an environmental perturbation out of a continuum of possibilities, cells will make a decision between discrete courses of action. For example, in the development of multi-cellular organisms, each cell has identical genetic information but implements a different genetic program based on its cell-type. As the organism develops, each cell must make a set of discrete decisions about which type of cell to become, and then it must remember these decisions indefinitely. Another example of a discrete choice is the regulation of genes responsible for utilizing an energy source present in the environment. Cells make this decision depending on whether there is enough of the metabolite present to make its metabolism beneficial. Implementing these decisions typically involves stably modifying the expression levels of key genes, leading to a distinct gene-expression pattern and phenotypic state.

Recent work with simple endogenous (Ozbudak et al. 2004; Acar et al. 2005; Suel et al. 2006) and synthetic (Gardner et al. 2000; Becskei et al. 2001) gene networks has elucidated how these stable gene expression states are established. At their core, these networks typically consist of a positive feedback loop that allows for two or more gene expression states to be stable in a given extracellular environment. Simple deterministic models have been developed which can explain how these multiple steady states are formed (Chung et al. 1996; Ozbudak et al. 2004). At the same time more complex models have attempted to describe how the discrete and stochastic nature of molecular events inside cells can affect the proportion of cells in each stable phenotypic state when the system has reached a steady state (Gardner et al. 2000). Although much work has been done recently to understand how the stochastic nature of gene expression can affect variability of protein levels in single cells (noise) (Arkin et al. 1998; Thattai et al. 2001; Ozbudak et al. 2002; Paulsson 2004; Hooshangi et al. 2005; Pedraza et al. 2005), it is still an open question as to how noise affects the dynamics of the decision-making processes in single cells.

Chapter 4.1 and (Mettetal et al. 2006) experimentally examine a system whose dynamics are heavily influenced by stochastic effects. Distributions of protein numbers in a population are measured as a function of time in the *Escherichia coli* lactose uptake network (*lac* operon). This network is well known for presenting an all-or-nothing (ON or OFF) gene expression response depending on the environmental conditions. Populations are exposed to time varying environmental conditions, which cause them to transition between the stable phenotypes, while gene expression levels are measured in single cells. These data show that two distinct types of transitions between steady states occur: ballistic and stochastic. Ballistic transitions are characterized by a population that transitions together in a synchronous manner from one state to the other. On the other hand, stochastic transitions occur when individual cells randomly switch from one state to another independently of the other cells in the population. A dynamic stochastic model is then introduced to show that predicting the dynamic distributions of protein levels characterizing these transitions requires only a few noise parameters in addition to the rates that characterize a deterministic model. Whereas the deterministic model cannot fully capture the observed behavior, the stochastic model correctly predicts0 the experimental dynamics without any fit parameters. These results provide a proof of principle for the possibility of faithfully predicting dynamic population distributions from deterministic models supplemented by a stochastic component that captures the major noise sources.

Correlations in Switching Dynamics

When cells divide, an entire pattern of gene expression can be passed from mother to daughter via inherited proteins and epigenetic chromatin modifications. Once cell division is complete, stochastic processes within cells cause this similarity to diminish, with closely related cells growing less similar over time as a result. Recent investigations of gene expression in single cells revealed the decorrelation time between related individuals (Rosenfeld et al., 2005) resulting from the inherently stochastic nature of mRNA and protein production. Nevertheless, it is still unclear how the correlation time affects decision making. In Chapter 4.2 and in (Kaufmann et al. 2007) we measure inheritance of a gene-expression state, and its subsequent decorrelation, in single yeast cells.

To accomplish this, a genetic network is engineered such that individual cells stochastically transition between two semi-stable states (ON and OFF) even in a

constant environment. We demonstrate that the switch times from the OFF to ON states are well described by a constant-rate stochastic process. At the same time, we find that even several generations after cells have physically separated many pairs of closely related cells switch in near synchrony. We quantify this effect by measuring how likely a mother cell is to have switched given that the daughter cell has already switched. This yields a conditional probability distribution very different from the exponential one found in the entire population of switching cells. We measure the extent to which this correlation between switching cells persists by comparing our results with a model Poisson process. Together, these findings demonstrate the inheritance of a dynamic gene expression state whose post-division changes include both random factors arising from noise as well as correlated factors that originate in two related cells' shared history. To explain the behavior observed, we construct a stochastic Monte Carlo model that demonstrates that our major findings can be explained by burst-like fluctuations in a single regulatory protein. The model retains Poisson switching statistics while also explaining the simultaneous switching time observed between distantly related cells. A counter-intuitive aspect of the model is that the long-time correlations are actually induced by randomness in the production rate of the regulatory protein, and the correlations increase as more noise is introduced. Finally, we discuss potential benefits and hindrances imposed by having cellular decisions occurring with synchrony in a population.

Transduction and Fidelity of Spatial Signals

It is generally assumed that single cells in an isogenic population exhibit the same behavior when exposed to an identical environment. However, it is becoming increasingly clear that even in a genetically identical population, gene expression levels can vary significantly from cell-to-cell giving rise to non-genetic individuality. It is an open question whether a conceptually similar individuality can be observed in other cellular activities, such as signal transduction. For example, it is unknown how the fidelity of sensing a directed extracellular cue varies from cell-to-cell. Many studies have been performed on sensing directional signals in single cells, however, most work on this topic ignores cell-to-cell variability and has instead focused on understanding the deterministic mechanisms by which extracellular gradients were transduced and processed (Insall et al. 1994; Lilly et al. 1995; Xiao et al. 1997; Parent et al. 1998; Levchenko et al. 2002; Janetopoulos et al. 2004; Manahan et al. 2004; Van Haastert et al. 2004). In Chapter

4.3 and in (Samadani et al. 2006) we explore the concept of population variability and signal fidelity in the gradient sensing response of single *Dictyostelium* cells when exposed to repeated spatio-temporal pulses of the chemoattractant cyclic adenosine monophosphate (cAMP). We find the response of a single cell to be highly reproducible from pulse-to-pulse. In contrast, a large variability in the response direction and magnitude is observed between cells, even when different cells are exposed to the same pulse. We find that this variability comes from two sources: 1) random orientation of an intracellular asymmetry between cells and 2) cellular individuality in the magnitude of this intracellular asymmetry. To explain these results, we propose that the effective signal a cell detects is the product of the extracellular cAMP signal and the randomly oriented intracellular signal. Using this model we successfully predict the observed population response to extracellular signals. This stochastic aspect of directional sensing might, on one hand, fundamentally limit the fidelity of signal detection and, on the other hand, might be beneficial by diversifying phenotypes in an isogenic population.

Bet Hedging as a Strategy for Surviving Fluctuating Environments

A classic problem in population and evolutionary genetics is to understand how a population optimizes its fitness in fluctuating environments (Cohen 1966; Levins 1968; Schaffer 1974; Stearns 1976). Rather than maintaining a phenotypically homogenous population, a population might enhance its fitness by allowing individual cells to stochastically transition between multiple phenotypes, ensuring that some cells are always prepared for an unforeseen environmental fluctuation. Recent theoretical models (Lachmann et al. 1996; Thattai et al. 2004; Kussell et al. 2005; Kussell et al. 2005; Wolf et al. 2005) have proposed that such a strategy is optimized when the transition rate between phenotypic states is tuned to match the rate of fluctuations in the environment. In Chapter 5.1 we experimentally explore how switching rates impact the population fitness using the galactose utilization network of *Saccharomyces cerevisiae*.

A yeast strain is genetically engineered that randomly transitions between two phenotypes due to stochastic gene expression (Paulsson 2004; Kaern et al. 2005; Samoilov et al. 2006; Kaufmann et al. 2007; Maheshri et al. 2007). Each phenotype is designed to confer a growth advantage over the other phenotype in a certain environment. A custom turbidostat setup was built to accurately (2% error) measure the population growth rates over long periods of time (up to 2 weeks). We compared the growth of two populations with different switching rates between the phenotypic states

after environmental transitions. We find that fast-switching populations outgrow slow-switchers when the environment fluctuates rapidly whereas slow-switching phenotypes out compete fast switchers when the environment changes more rarely. These results confirm the theoretical findings that, in order to enhance population fitness, cells may be able to tune their inter-phenotype switching rate to the frequency of environmental changes.

2 Background

2.1 Quantitative Methods

Fully understanding the cellular systems that are responsible for signal processing and decision-making in single cells requires that we sidestep the reductionist approach historically taken in biology research. It is not enough to study individual components or interactions alone; rather, the entire network must be studied collectively. By combining traditional biochemical techniques with mathematical models, the systems level approach has been extremely useful in understanding a variety of biological phenomena at the level of emergent network behavior (Ideker et al. 2001; Kitano 2002). This chapter explores several simple systems that illustrate key concepts relating to the processing of dynamic signals, decision-making, and stochastic effects (noise) in single cells. First, the properties of bistability are explored in a simple gene regulatory network with positive feedback. The second section uses Fourier analysis to explore how dynamic signals are propagated through gene regulatory networks. Finally, the last section details how stochastic effects in simple biochemical reactions and reviews how this noise affects cell-to-cell variability in gene expression levels.

2.2 Decision Making: Positive Feedback and Hysteresis

Positive feedback is a common gene regulatory motif that is important for cellular decision making. For example, cells can use this type of network in order to implement a "memory device" where a gene-expression program remains in action after a transient stimulus has been removed (Gardner et al. 2000; Xiong et al. 2003; Ozbudak et al. 2004; Acar et al. 2005). At their core, these networks function by using the proteins produced from a gene to positively regulate (directly or indirectly) the promoter of that gene. When the transcription factor is present at very low levels, transcription of its gene locus is infrequent and the level of transcription factor remains low. On the other hand, when the protein is present in large quantities, transcription occurs at a high rate and the protein levels remain high. The concept of bistability induced memory is used extensively in several systems studied in Chapters 4.1, 4.2, and 5.1. This section demonstrates how memory and hysteresis arises in the context of a simple positive feedback network, and while the systems explored in later chapters are more complex, they share many similarities with the example given here.

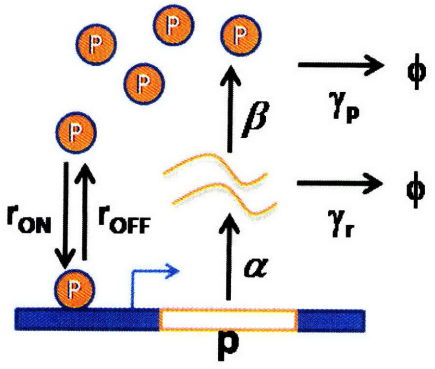


Figure 2.2.1 Simple feedback in a gene regulatory network. The gene *p* is transcribed to mRNA and translated into protein P. Protein P can bind to the promoter increasing the rate of expression of the protein P.

To understand how bistability can arise when positive feedback is present in a network, we begin by examining a system that has a positive feedback but does not exhibit bistability. The system is shown in Figure 2.2.1 where the transcription rate of a gene is regulated by its own gene product. In this scenario, the gene *p* encodes for protein P. The protein P binds to a promoter upstream of the gene with a single binding site. The rate at which the protein binds to its site on the promoter is first order in protein

levels and written as, $r_{ON}P$, where P is the concentration of the protein. Bound proteins can then unbind with a rate r_{OFF}, which does not depend on the concentration of free proteins, but rather on the intrinsic affinity between the protein and the DNA. mRNA (quantified by R) is produced from the DNA with rate α only when a protein is bound to the binding site, while protein is produced from mRNA with rate β. mRNA is degraded linearly with a rate constant γ_R while protein is degraded linearly with a rate constant γ_p. The fraction of time a protein is bound to the binding site is written f_{bound}, and these processes are summarized in the following equations:

$$\frac{d}{dt}R = \alpha f_{bound} - \gamma_R R$$
$$\frac{d}{dt}P = \beta R - \gamma_P P$$
$$r_{ON}P(1 - f_{bound}) = r_{OFF} f_{bound}$$

We set the time derivatives of R and P to zero in order to find steady state solutions to the system of equations, resulting in the following relations:

$$f_{bound} = \frac{r_{ON}P}{r_{ON}P + r_{OFF}}$$
$$R = \frac{\alpha}{\gamma_R} f_{bound}$$
$$P = \frac{\beta}{\gamma_P} R = \frac{\alpha\beta}{\gamma_P \gamma_R} f_{bound}$$

Depending on the exact values of the parameters solving for P can give either two or one steady state protein concentrations. For $\alpha\beta r_{ON} > \gamma_P \gamma_R r_{OFF}$, the steady state solutions are:

$$P = \{0, \frac{\alpha\beta}{\gamma_P \gamma_R} - \frac{r_{OFF}}{r_{ON}}\},$$

where the first steady state is unstable and the second is stable. When $\alpha\beta r_{ON} < \gamma_P \gamma_R r_{OFF}$, there is a single stable steady state at P=0. In either case, the system always has a single stable steady state. The steady state is shown graphically in Figure 2.2.2 as

intersections between the rate of production and rate of degradation of proteins. From Figure 2.2.2 it is clear that since the slopes of both curves are monotonic, they will intersect at a maximum of two points leading to a single steady state solution, and to get more than two steady states one of the two curves will need a non-monotonic slope.

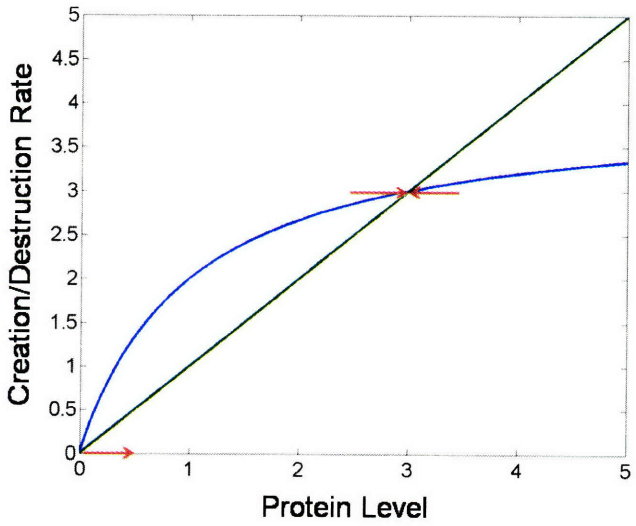


Figure 2.2.2 Rate of protein creation (blue) and destruction (green) as a function of protein levels, *P*. When the production is greater than degradation, protein levels increase (red arrows to right), but when destruction is greater than production the protein levels decrease (red arrows to left). This means that there will always be at most a single stable steady state (red arrows point towards steady state) at either one or zero non-steady states (red arrows point away from steady state).

Bistability can be achieved very simply by modifying the example above and introducing multiple interacting binding sites at the promoter of the gene. Interactions between the proteins can cause the second protein to bind more or less rapidly when a first protein has already bound to one of the two binding sites (Figure 2.2.3). Assuming identical binding sites, a_1 and a_2 are the rates for going from zero bound to one bound and from one bound to two bound, respectively. b_2 and b_1 are the rates for going from two bound to one bound and from one bound to zero bound, respectively. In this case the fraction of time DNA spends with one protein bound is f_1 and with two bound is f_2, and assuming that protein is available in excess the dynamics of the bound fractions are given by the following equations:

$$\dot{f}_0 = b_1 f_1 - a_1 f_0$$
$$\dot{f}_1 = a_1 f_0 + b_2 f_2 - b_1 f_1 - a_2 f_1$$
$$\dot{f}_2 = a_2 f_1 - b_2 f_2$$

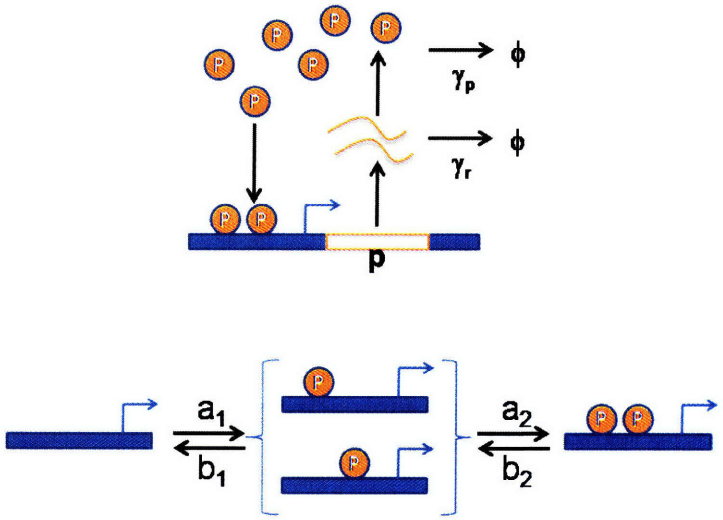


Figure 2.2.3 Autoregulatory network with dual identical binding sites. Transcription rates are low when there are zero or one proteins attached to the DNA. When two proteins are attached, transcription occurs at a higher rate. Proteins bind and unbind to the DNA preferentially when there is already a single protein occupying the other binding site.

Assuming that the reactions are in steady state and that there is a small basal rate of mRNA production (*basal*) in the absence of two bound proteins the equations governing production of the protein will be:

$$f_2 = \frac{1}{1 + \frac{b_2}{a_2 P}(1 + \frac{b_1}{a_1 P})}$$
$$R = \frac{\alpha}{\gamma_R} f_2 + basal$$
$$P = \frac{\beta}{\gamma_P} R = \frac{\beta}{\gamma_P}(\frac{\alpha}{\gamma_R} f_2 + basal)$$

This set of equations can have one, two, or three steady state solutions (Figure 2.2.4). For small values of α, there is only one steady state solution. At intermediate values,

there are three steady states. The lowest and highest steady states are stable, while the middle steady state unstable. In this range of parameters, identical cells can have different levels of protein depending on which stable steady state they are in. For large rates of protein production, there is only one large steady state with a large amount of protein.

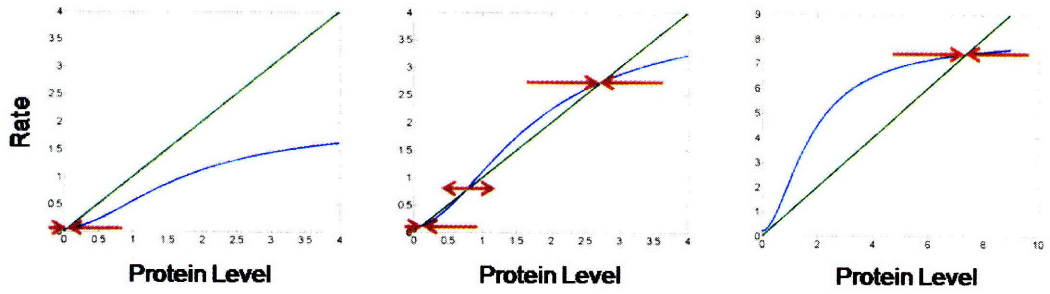


Figure 2.2.4 Steady states for the multiple identical interacting binding site system. Blue line is protein production rate, while green line is protein degradation rate. When α is small (left panel) there is a single stable steady state at low protein levels. For intermediate values of α (center panel) three steady states are possible, a low stable steady state and a high stable steady state are separated by an unstable steady state. For high values of α there is only a single steady state at high protein levels. Parameters used are $\beta = 10$, $\gamma_p = 1$, $\gamma_r = 10$, $basal = 0.025$, $a_1 = 1$, $a_2 = 1$, $b_1 = 5$, $b_2 = 0.5$.

These steady states are graphed in Figure 2.2.5, where the stable and unstable steady are shown in green as the parameter alpha is varied. Because of the arrangement of steady states, this system can exhibit hysteresis (Figure 2.2.5). Cells initially prepared in the ON (OFF) state with large (small) α, will remain in the ON (OFF) state when α is slowly decreased (increased) into the bistable region all three steady states exist for a single value of α. Since these states are stable, cells will only leave these states if α is modified so that the states are no longer stable (Figure 2.2.5, red arrows), or if stochastic effects drive a rare transition between the states (Kramers 1940; Risken 1996; Acar et al. 2005). This is similar to thermal noise driving transitions between local minima in an energy landscape. Here the stable states correspond to the local minima, the local maxima correspond to the unstable steady state, and thermal noise is similar to the noise introduced by stochastic effects due to gene expression and other biochemical processes.

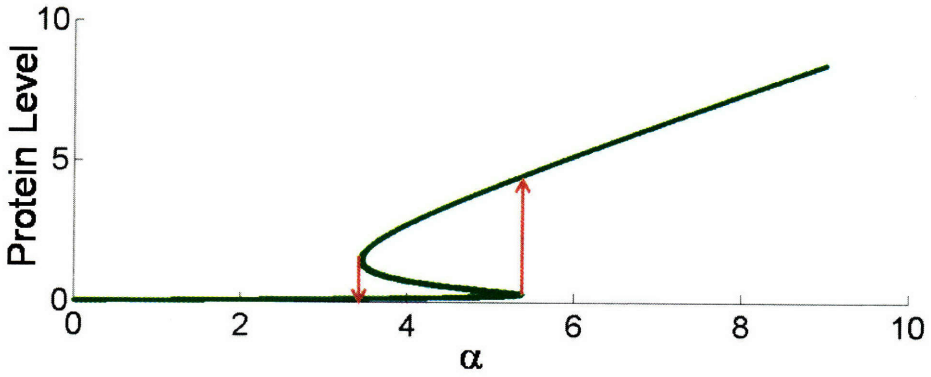


Figure 2.2.5 Steady states plotted as a function of the parameter α. Low values of alpha have a low stable steady state protein level, while high values have a stable high protein level. Intermediate values of α can stably support either high or low levels of protein.

2.3 Signal Processing: Network Structure and System Dynamics

To illustrate the connection between network topology and system dynamics, we consider the dynamic signal-processing performed by a single gene that encodes for a protein P. Here it is assumed that the rate of transcription of this gene is instantaneously affected by the concentration of a transcription factor, S, whose concentration is varying over time. Further, the protein is assumed to degrade linearly with a rate γ_p. By also assuming that the production and degradation of mRNA is very fast compared to the dynamics of protein levels, we can write the following relation for the concentration of protein, p:

$$\dot{p} = f(s) - \gamma_p p$$

Here $f(s)$ is the rate of protein production (accounting for both transcription and translation) for any given concentration of transcription factor, s. A schematic representation of this process is shown in Figure 2.3.1.

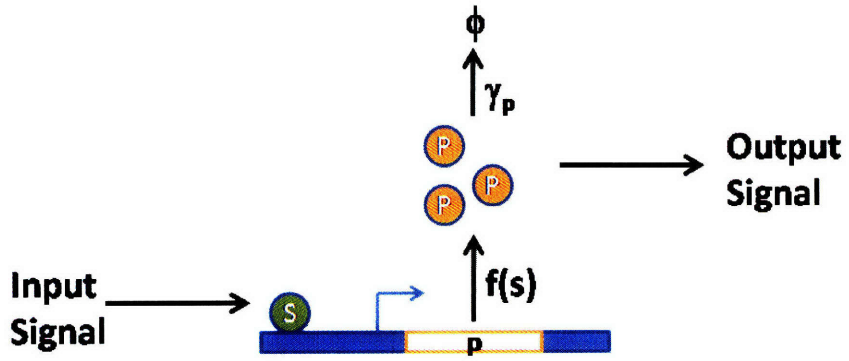


Figure 2.3.1 A single gene, p, under control of a transcription factor S. Molecules of S bind to the promoter region of the gene and cause the rate of gene expression to change (quantified by $f(s)$). Proteins, P, are linearly degraded with rate γ_p. The time dependent concentration of protein is the module output. The filtering properties of this module are obtained by analyzing this system in the frequency domain.

If we assume that the fluctuations in the concentration of s are sufficiently small, we can use a linear approximation of this relationship:

$$\delta\dot{p} = f(s_0) + \left(\left. \frac{df(s)}{ds} \right|_{s=s_0} \right) \delta s - \gamma_p (p_0 + \delta p)$$

Here we defined δs and δp as deviations from a basal activity s_0 and p_0:

$$\delta s = s - s_0$$
$$\delta p = p - p_0 = p - \frac{1}{\gamma_p} \left(\left. \frac{df(s)}{ds} \right|_{s=s_0} \right)$$

By defining the Fourier transform of δs and δp as:

$$\delta\tilde{s}(\omega) = \int_{-\infty}^{\infty} s(t) e^{i\omega t} dt$$
$$\delta\tilde{p}(\omega) = \int_{-\infty}^{\infty} p(t) e^{i\omega t} dt$$

the dynamic relation between δs and δp can be rewritten in Fourier space as:

$$\delta\tilde{p} = \frac{1}{i\omega + \gamma_p} \left(\left. \frac{df(s)}{ds} \right|_{s=s_0} \right) \delta\tilde{s}$$

This relation allows one to determine the "filtering" properties of this gene as a function of the signal's frequency. Figure 2.3.2 shows the amplitude and phase of the output signal in response to an input of constant amplitude and phase, $s(t) = s_0 + \cos(\omega t)$. For low frequencies, the signal is propagated without damping and without a significant delay. At higher frequencies, the signal is quickly filtered out by the slow accumulation and degradation of proteins. The output signal for a given frequency input signal is determined by two parameters alone. First the derivative $\left(\left. \frac{df(s)}{ds} \right|_{s=s_0} \right)$ determines the magnitude of the output because it quantifies how sensitively p depends on changes in s. Similarly γ_p determines the magnitude because it controls how quickly the protein is

degraded. As the only parameter containing units of time, γ_p sets the 'cutoff frequency' (Figure 2.3.2) at which point the amplitude of the output signal begins to decrease.

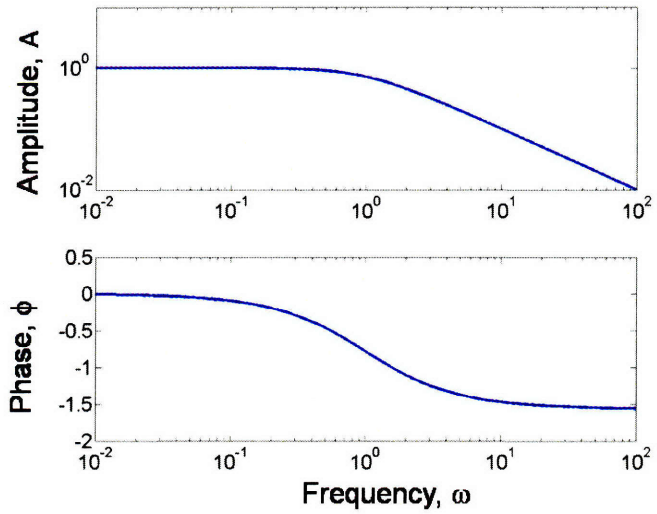


Figure 2.3.2 Filtering properties of a single gene with unit gain. The top panel is the amplitude of the output when the upstream transcription factor oscillates at frequency ω with unit amplitude. The bottom panel is the delay, represented as a phase shift ϕ. For frequencies lower than $\omega = 1$, the gene output faithfully follows the input with little delay. At high frequencies the system filters out the fluctuations in transcription factor. The parameter γ is set to one in this calculation.

Next, we examine the filtering properties of a more complex gene regulatory network similar to that of the 'repressilator' constructed by (Elowitz et al. 2000). By utilizing nonlinear effects and large gains, this synthetic gene regulatory network is known to produce spontaneous oscillations, suggesting that the simplified model analyzed here should have a non-trivial frequency-dependent behavior. This network consists of three genes: *x*, *y*, and *z* each encoding for a protein X, Y and Z respectively. Gene *x* is regulated by protein Z as well as by an upstream transcription factor S. Protein X binds to the promoter of *y* and regulates its expression, while Y binds to the promoter region of *z* and regulates its expression. The concentration of transcription factor S is therefore the input of the module and the concentration of protein X is the output. This process is shown in Figure 2.3.3.

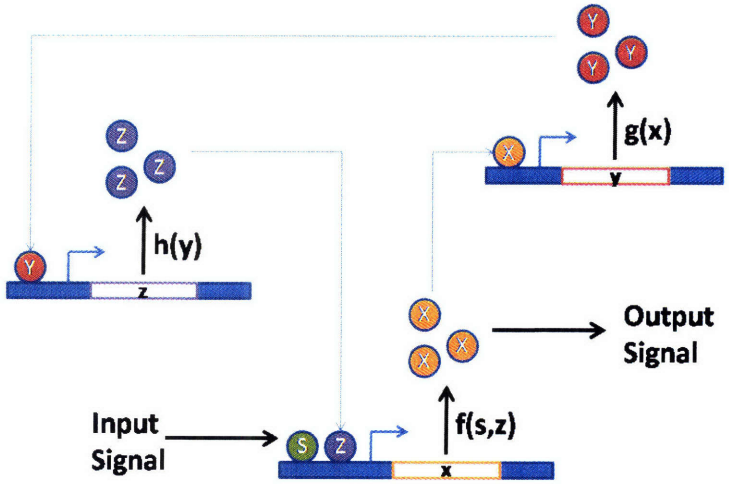


Figure 2.3.3 Signal processing module comprised of three genes, regulating each other in a 'ring-like' structure. When these proteins repress expression of the downstream genes, this network is similar to that of the repressilator (Elowitz et al. 2000). Each protein effectively regulates its own expression through an intermediate gene, and this leads to delays which can make this network especially sensitive to specific driving frequencies.

The genes y and z form a feedback loop on the expression of x, and the delay associated with this cascade will cause the network to be more sensitive to some input frequencies than to others. Again, assuming first-order protein degradation of rate γ for all proteins and fast mRNA production degradation kinetics, the dynamic equations governing protein levels (Elowitz et al. 2000) can be written as:

$$\dot{x} = f(s,z) - \gamma x$$
$$\dot{y} = g(x) - \gamma y$$
$$\dot{z} = h(y) - \gamma z$$

After defining deviations from a baseline network activity,

$$\delta s = s - s_0$$
$$\delta x = x - x_0 = x - \frac{f(s_0, z_0)}{\gamma}$$
$$\delta y = y - y_0 = y - \frac{g(x_0)}{\gamma}$$
$$\delta z = z - z_0 = z - \frac{h(y_0)}{\gamma}$$

and linearizing about this state, the dynamic equations become:

$$\delta\dot{x} = f_s'(s_0, z_0)\delta s + f_z'(s_0, z_0)\delta z - \gamma\delta x$$
$$\delta\dot{y} = g'(x_0)\delta x - \gamma\delta y$$
$$\delta\dot{z} = h'(y_0)\delta y - \gamma\delta z$$

Here f'_s and f'_z are the partial derivatives of f with respect to s and z respectively at the point s_0, z_0. By defining Fourier transforms of the dynamic variables as before, the dynamic equations can be written in frequency space as:

$$i\omega\delta\tilde{x} = f_s'(s_0, z_0)\delta\tilde{s} + f_z'(s_0, z_0)\delta\tilde{z} - \gamma\delta\tilde{x}$$
$$i\omega\delta\tilde{y} = g'(x_0)\delta\tilde{x} - \gamma\delta\tilde{y}$$
$$i\omega\delta\tilde{z} = h'(y_0)\delta\tilde{y} - \gamma\delta\tilde{z}$$

Finally, by solving for the Fourier transform of δx alone, the dynamic relationship between input and output can be written as a single equation:

$$\delta\tilde{x} = \delta\tilde{s}\frac{(\gamma + i\omega)^2 f_s'}{(\gamma + i\omega)^3 - f_z' g' h'}$$

The frequency response is presented as a Bode plot in Figure 2.3.4. Because of the delay associated with the feedback through genes y and z, we see that the amplitude exhibits a peak in the response. The behavior is similar to that of a band-pass filter where a narrow band of frequencies are allowed to pass through while frequencies outside of the band are filtered out of the output signal.

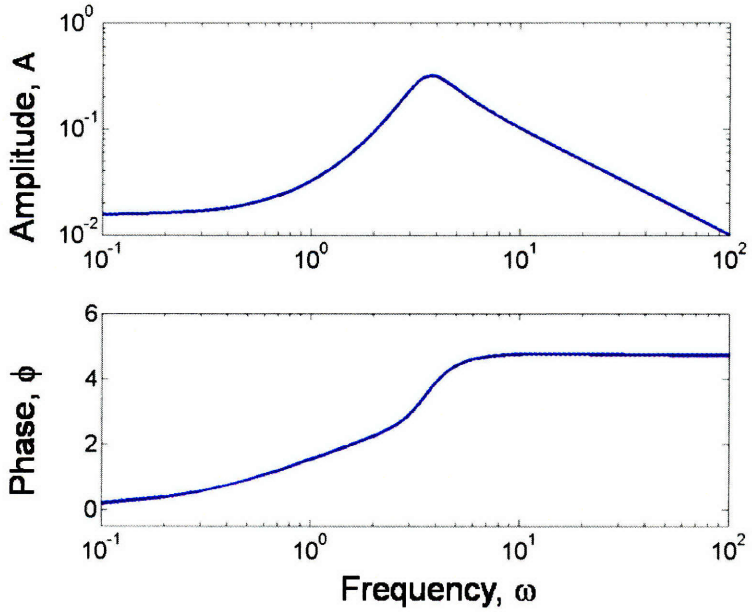


Figure 2.3.4 Frequency response of the three gene network. Top panel is the response amplitude as a function of frequency, while bottom panel is the phase shift. This plot was generated using $\gamma = 1$, and a gain of -4 for each gene.

These two examples illustrate the signal processing properties of two simple gene regulatory networks. As seen in the first example a single gene controlled by an upstream transcription factor will have certain low-pass filtering properties that act to constrain the range of signals to which the downstream protein levels will be sensitive. The second example demonstrates how these modules can be connected to perform more complex filtering properties. By exploiting the inherent delay present in a signaling cascade (here gene *y* affecting gene *z*) a gene regulatory network may be highly sensitive to a narrow band of frequencies while filtering out higher and lower frequencies. It is worth noting that the example presented here does not autonomously oscillate; given a non-fluctuating input signal, oscillations in the network will decay over time. Nevertheless, large gains combined with non-linearities can cause a gene regulatory network with an identical structure to oscillate in the absence of a driving signal (Elowitz et al. 2000). While the calculations presented here use linearized equations and therefore do not capture this behavior, they still predict that the network will resonate strongly at a specific input frequency.

2.4 Noise in Gene Expression

Biological networks frequently contain chemical reactions where reactants are present in low numbers of discrete molecules, and this introduces inherent stochasticity to the processes performed by these systems. The numbers of particular molecules in cells can vary widely (from a single copy for DNA to tens or hundreds of thousands of copies for a single species of protein), and the timing of individual reactions occurs in a random manner. From simple statistical arguments, when the number of molecules is relatively abundant, the expected cell-to-cell variability from the intrinsic stochasticity of that process should be low. On the other hand, processes involving molecules with low numbers or rare events, such as mRNA transcription from DNA, will yield larger amounts of variability. Cell-to-cell variability due to stochastic gene expression has historically been observed in a variety of contexts (Delbruck 1945; Benzer 1953; Lederberg et al. 1956; Novick et al. 1957; Powell 1958; Maloney et al. 1973; Spudich et al. 1976), but only recently have the origins and details of this noise been experimentally and theoretically understood. This section will introduce some of these concepts by discussing the intrinsic noise of a constitutively expressed gene as well as the level of noise in a gene controlled by a noisy upstream transcription factor.

The process of mRNA production from the DNA template on a constitutively expressed gene is sufficient to illustrate how stochastic effects are introduced during biological processes. Here the DNA is transcribed by RNA polymerase resulting in mRNA. mRNA is assumed to be degraded in a Poisson manner. The rate of production of mRNA (m) is written as α, while the degradation rate for m molecules of mRNA is γm, linear with the number of mRNA molecules. This process is summarized in Figure 2.4.1.

$$p(m-1) \underset{\gamma m}{\overset{\alpha}{\rightleftharpoons}} p(m) \underset{\gamma(m+1)}{\overset{\alpha}{\rightleftharpoons}} p(m+1)$$

Figure 2.4.1 The rate of change of the probability of having m molecules of mRNA depends on the probability of having m-1 and m+1 molecules of mRNA. Right arrows indicate rate of production of mRNA molecules, while left arrows represent their rate of degradation.

The probability of having a cell with m mRNA molecules is written as $p(m)$. The process of creation and destruction can then be written in master equation form:

$$\dot{p}(m) = -p(m)(m\gamma + \alpha) + (m+1)\gamma p(m+1) + \alpha p(m-1).$$

When this process is stationary, $\dot{p}(m) = 0$ and the equation has solutions of the form:

$$p(m) = \left(\frac{\alpha}{\gamma}\right)^m \frac{e^{-\frac{\alpha}{\gamma}}}{m!}.$$

This is a simple Poisson distribution, and therefore has the following properties:

$$\langle p(m) \rangle = \frac{\alpha}{\gamma}$$

$$\langle p(m)^2 \rangle - \langle p(m) \rangle^2 = \frac{\alpha}{\gamma}$$

Therefore, the expectation is that from a single gene the variance in the number of mRNA will be of equal to the mean number of mRNA's present in the cell.

Fluctuations in protein levels are also important for affecting a cell's phenotype and therefore noise in protein levels must be quantified in a similar manner as fluctuations in mRNA levels. Before degradation, a single mRNA molecule may be translated multiple times; each round of translation by a ribosome results in the production of a new protein. This has the effect of amplifying the noise present in mRNA levels leading to variability in protein numbers beyond those expected from simple Poisson statistics (i.e. variance in protein levels will be much larger than the average number of proteins). A similar calculation has been performed which takes into account fluctuations in protein levels due to the discrete nature of the proteins in addition to the discrete level of mRNA inside cells. The resulting variability in protein levels arising from the intrinsic noise of mRNA and protein production and destruction can be approximated as $\langle \delta n^2 \rangle = \langle n \rangle (b+1)$ (Thattai et al. 2001; Ozbudak et al. 2002), where n is the number of proteins present in a cell, and b is the average number of proteins produced from a single mRNA. Because the proteins are produced from a single mRNA rapidly in

relation to the protein lifetime, the process of transcription from a single mRNA is commonly referred to as a 'burst' of translation, and *b* is referred to as the 'burst size'. Again, it is worth noting that this bursting process is amplifying the noise present in the mRNA levels, causing proteins to have a variance, which is much larger than the mean number of proteins.

While this example demonstrates how noise can be intrinsically generated by a genetic network, variability in protein levels can also be introduced from extrinsic sources. This is because other factors (e.g. numbers of ribosomes or upstream transcription factors) often directly influence the rate of expression of a gene, and any variability in the levels of these molecules will be transmitted directly into fluctuations in the downstream gene of interest. This effect has been calculated theoretically (Thattai et al. 2002; Paulsson 2004) and quantified experimentally (Hooshangi et al. 2005; Pedraza et al. 2005; Rosenfeld et al. 2005) in simple gene-regulatory cascades where noise from upstream genes is propagated into downstream gene expression levels.

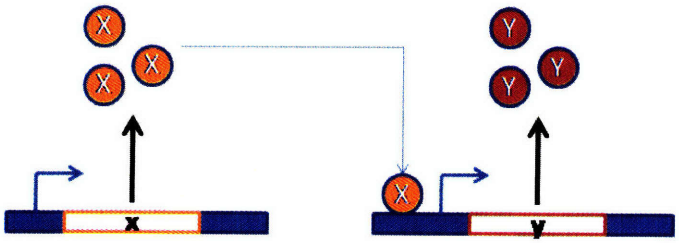


Figure 2.4.2 Regulatory cascade used to demonstrate propagation of noise from an upstream to downstream gene. The first gene is constitutively (and noisily expressed) while X proteins can bind to the promoter region of *y*, regulating the production of Y proteins.

The propagation of noise from one gene to the next can be understood by examining the following model system shown in Figure 2.4.2. Here protein X is constitutively produced with rate α_X, while fluctuations in X are propagated downstream into the transcription rate of gene *y*. Both proteins are assumed to be lost linearly with rate γ. The rate of protein production for gene *y* is $f(X)$ which is linearized about the steady state and written as $\alpha_Y + \beta(X-\alpha_X/\gamma)$. These linearized equations can be written for the dynamics of the two proteins, using the Langevin formalism:

$$\dot{X} = \alpha_X - \gamma X + \eta_X$$
$$\dot{Y} = \alpha_Y + \beta(X - \frac{\alpha_X}{\gamma}) - \gamma Y + \eta_Y$$

Here η_X and η_Y are random noise terms representing the stochastic variability introduced by intrinsic fluctuations arising from the stochastic nature of *X* and *Y*'s gene expression. Here we assume that the fluctuations propagated from *X* into *Y* are sufficiently small that η_Y is not significantly changed. Statistically, η_X and η_Y are derivatives of a Wiener process and are uncorrelated random processes with strengths characterized by the definitions:

$$<\eta>=0$$
$$<\eta_{X,Y}(t)\eta_{X,Y}(t+\Delta t)>=\eta_{0-X,Y}^2\delta(\Delta t)\,.$$

Here η_{0-X}^2 and η_{0-Y}^2 capture the magnitude of intrinsic fluctuations in *X* and *Y*. Defining deviations from steady state values for *X* and *Y* as $\delta X = X - \frac{\alpha_X}{\gamma}$ and $\delta Y = Y - \frac{\alpha_Y}{\gamma}$, and Fourier transforming results in the following relations:

$$(i\omega+\gamma)\delta\tilde{X} = \tilde{\eta}_X$$
$$(i\omega+\gamma)\delta\tilde{Y} = \beta\delta\tilde{X} + \tilde{\eta}_Y$$

Solving for $\left\langle \delta\tilde{X}\delta\tilde{X}^* \right\rangle$ and applying the Wiener-Khintchine Theorem results in a variance for *X* given by:

$$\langle \delta X^2 \rangle = \frac{\eta_0^2}{2\gamma}$$

Similarly, solving for $\delta\tilde{Y}$ yields:

$$(i\omega + \gamma)\delta\tilde{Y} = \beta\frac{\tilde{\eta}_X}{(i\omega + \gamma)} + \tilde{\eta}_Y$$

Applying the Wiener-Khintchine Theorem and noting that the intrinsic fluctuations between X and Y are uncorrelated, $\langle \tilde{\eta}_X \tilde{\eta}_Y \rangle = 0$, results in:

$$\langle \delta Y^2 \rangle = \frac{\eta_{0-Y}^2}{2\gamma} + \beta^2 \frac{\eta_{0-X}^2}{4\gamma^3}$$

Similar to previously derived equations (Paulsson 2004), the resulting variance in *Y* is the sum of noise from two sources. The first term on the RHS characterizes fluctuations due to noise intrinsic to the expression of *Y* proteins, while the second term on the RHS characterizes the noise propagated into the protein levels by noise in the upstream protein, *X*. Intuitively, the magnitude of this noise is directly dependent on how strongly *X* influences *Y*, characterized by β.

3 Biological Systems and Dynamic Signals

3.1 Probing Network Dynamics with Periodic Stimuli

3.1.1 Introduction

Cells have evolved intricate and elaborate mechanisms to sense and respond to environmental changes, and these mechanisms involve complex systems of biochemical reactions that occurring over a wide range of timescales. For instance, the processes involved in a cellular response can be very fast, such as association and dissociation between a ligand and its receptor, or very slow, such as protein synthesis. Though a system may be comprised of hundreds of reactions, often only a few of these reactions will dictate the system dynamics. Thus, building models that accurately reflect system dynamics requires identifying the dominant processes and measuring them quantitatively. Unfortunately, identification of such processes is often difficult, and many models instead incorporate present knowledge of all known biochemical reactions in the system. Although occasionally successful (Tyson et al. 2001; Slepchenko et al. 2003; Kitano et al. 2005; Klipp et al. 2005), this exhaustive approach can still suffer from missing information such as unknown interactions or parameters.

In this chapter we utilize systems engineering tools to study how oscillatory signals propagate through a signal transduction cascade, thereby enabling us to identify the properties of key reactions and to make a simple model of system dynamics. The cornerstone of this approach is to measure cascade output in response to input signals oscillating at different frequencies (Oppenheim 1983; Ljung 1999; Westwick 2003). For instance, fast reactions can be studied by stimulating the cells with a rapidly oscillating input; under such conditions, slow reactions cannot respond strongly to the rapidly changing signal. By contrast, the contribution of slow reactions can be measured with slowly oscillating inputs, where fast reactions equilibrate and cease to affect system dynamics. By comparing the frequency response of the wild-type network to that of mutant strains, the molecular underpinnings of the dominant time-scales can be determined. Studies of neural and other physiological systems have employed these principles from systems theory (Westwick 2003), while pioneering work has begun to apply control theory concepts to cellular networks (Savageau 1976; Block et al. 1982; Yi et al. 2000; Samoilov et al. 2002; Ingalls 2004; Lipan et al. 2005; Sontag 2005).

We focus on the high osmolarity glycerol (HOG) Mitogen-activated protein kinase (MAPK) cascade in the budding yeast *Saccharomyces cerevisiae*. This cascade forms a

core module of the hyperosmotic-shock-response system and is particularly well suited to frequency-response analysis for several reasons. The system contains multiple negative feedback loops that differentially regulate osmoadaptation on different time-scales (Hohmann 2002; Klipp et al. 2005; Gennemark et al. 2006), but it is still unclear which negative feedback loop(s) dominates the system dynamics. Additionally, it is unknown if the different feedback loops have distinct biological functions. Here, we determine properties of the main negative feedback loops in the HOG network and arrive at a concise predictive model of the signaling dynamics. Furthermore, by analyzing the system's dynamics over a range of osmotic shock strengths, we begin to understand how the multiple feedback architecture might be beneficial for osmoadaptation in fluctuating environments.

3.1.2 Network Architecture

The system we study with periodic stimuli is the osmoregulatory network in *Saccharomyces cerevisiae*. This system has been extensively studied (Hohmann 2002), and many of the salient interactions are shown in Figure 3.1.1. The network is able to keep intracellular osmotic pressure in check with the extracellular osmotic pressure through a set of negative feedback loops. Changes in osmotic pressure are sensed through several mechanisms. First, the transmembrane proteins Sho1 and Sln1 are activated in response to osmotic shock, and both proteins lead to downstream increase in Pbs2 activity. Doubly phosphorylated Pbs2 acts as a kinase to phosphorylate two sites on the MapK Hog1 causing it to be activated. Active Hog1 preferentially translocates to the nucleus where it can activate gene expression for a large number of genes, notably Gpd1 and Gpp2, which increase the rate of glycerol synthesis. Glycerol is used as an intracellular osmolyte by the cell to increase the internal osmotic pressure and counteract any externally applied osmotic perturbations. In addition to this feedback loop through Hog1, a second feedback loop exists through the membrane glycerol transporter Fps1. Under hyper-osmotic stress, Fps1 channels close increasing the intracellular glycerol concentration, while under hypo-osmotic stress, Fps1 channels open decreasing the intracellular glycerol concentration.

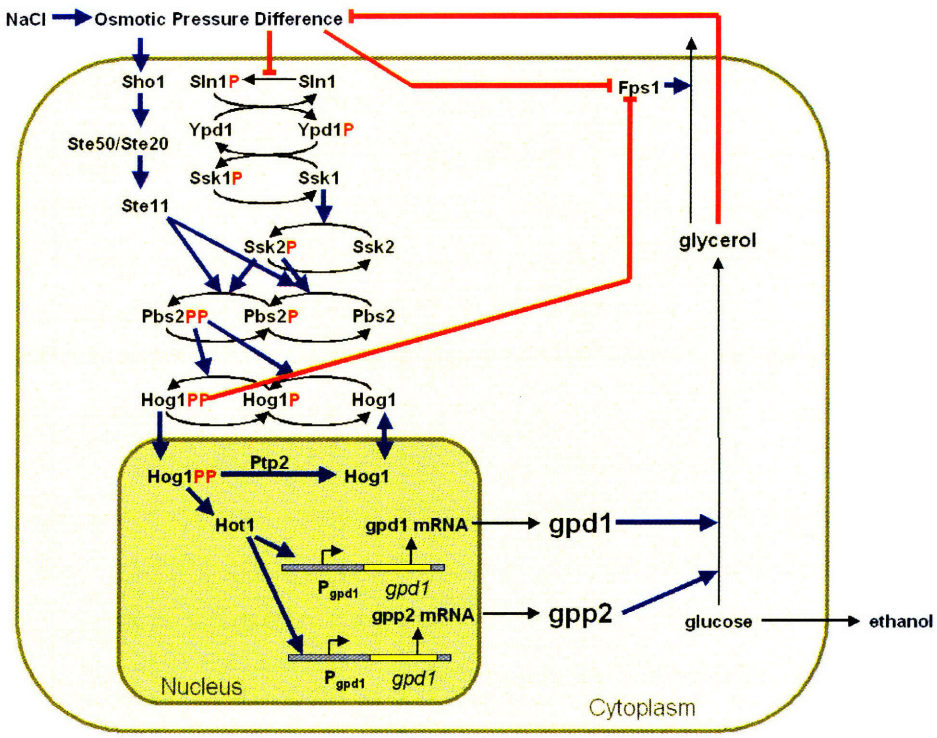


Figure 3.1.1 Structure of the osmoregulatory network studied in this chapter. The network is characterized by a multiple negative feedback structure. One feedback occurs through osmotic pressure interacting with Fps1 channels, while the second occurs with Hog1's proposed interaction with Fps1. The final feedback loop occurs when Hog1 upregulates gene expression of Gpd1 and Gpp2 which increase the production of glycerol.

3.1.3 Experimental Setup

Cells are periodically stressed with an osmotic shock, using a flow chamber and a computer-actuated valve to apply square-wave pulses of media with and without 0.2 M NaCl while concurrently imaging cells and measuring their Hog1 localization (Figure 3.1.2). The chamber is constructed by cutting a channel into an adhesive gasket with a thickness of 0.17 um (Grace Biolabs), which is then sandwiched between a microscope slide and glass coverslip. Cells in log-phase growth from an overnight low-osmolarity culture are concentrated (1mL of cells is spun at 13,000 rpm and resuspended in 0.1mL) and immobilized on the coverslip with Concanavalin A. Cells are then visualized with an inverted microscope (Nikon TE2000) and 100x phase contrast objective. Media is constantly removed from the flow cell with a peristaltic pump (Pharmacia) coupled to a 125mL reservoir, which dampens sudden changes in pressure. A computer-controlled valve selects the source of replacement media.

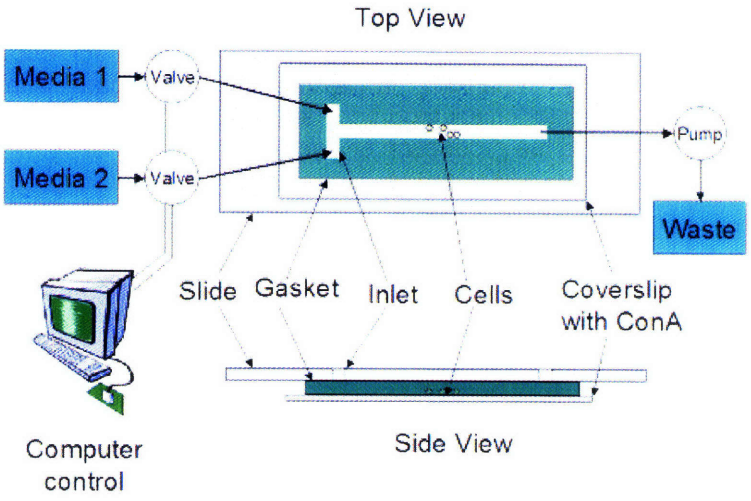


Figure 3.1.2 Schematic diagram of the flow-cell setup. A pump removes media from the flow chamber while a valve constantly selects from one of two fresh media sources.

To measure the time the flow chamber takes to equilibrate, a small concentration of Rhodamine (Sigma-Aldrich) is added to the first media source. The average fluorescence of a region in the middle of the chamber visualized with a 4x objective is shown in Figure 3.1.4. The rapid equilibration of the fluorescent signal indicates that the concentration of NaCl should also equilibrate in the chamber within 3-5 seconds.

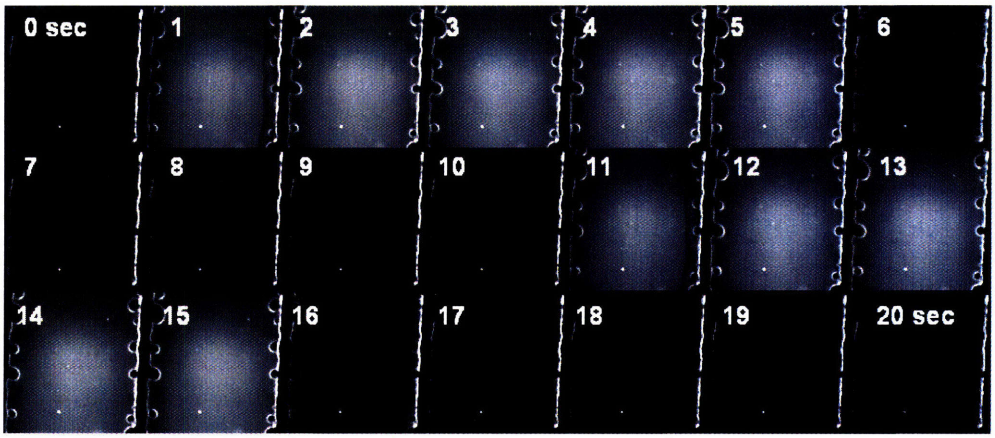


Figure 3.1.3 Fluorescent images of Rhodamine in a flow chamber using 4x objective.

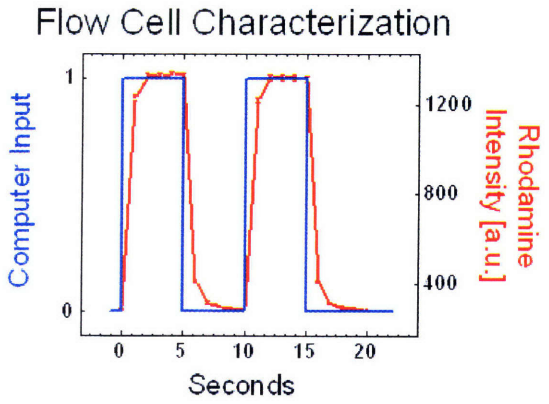


Figure 3.1.4 Flow cell characterization indicates that the time-constant for complete media replacement in the flow chamber is significantly less than 10 seconds.

Upon hyperosmotic shock, membrane proteins trigger a signal-transduction cascade that culminates in the double phosphorylation and activation of the MAPK Hog1, which is primarily cytoplasmic prior to the osmoshock (Ferrigno et al. 1998). When phosphorylated, Hog1 is rapidly accumulated into the nucleus. (Ferrigno et al. 1998; Reiser et al. 1999). Once the osmotic balance has been restored, either through changes to the extracellular environment or the intracellular osmolyte concentration, Hog1 is no longer activated by the cascade, and the amount of nuclear Hog1 rapidly decreases. We simultaneously monitored the cellular localization of a Hog1-YFP fusion protein and Nrd1-RFP, a red fluorescent protein fused to a strictly nuclear protein. To quantify the degree of Hog1 nuclear localization, we define the dynamic response

function, $R(t) = \left\langle \frac{< YFP >_{nucleus}}{< YFP >_{cell}} \right\rangle_{population}$, as the ratio (averaged over many cells) of mean YFP pixel intensity in the nucleus and mean YFP pixel intensity in the whole cell.

An time series of this process is presented in (Figure 3.1.5A). Here cells are initially stressed with 0.2M NaCl for 4 min at which point the NaCl stimulus is removed. At t = 0 the YFP signal is mostly homogeneously distributed within the cell, with a slight excess in signal coinciding with the nucleus. This is reflected in the response measurement that starts around R = 1.2 in the absence of stimulus. It has been reported that unphosphorylated Hog1 can be imported into the nucleus (Westfall et al. 2006), and this measurement indicates that the basal import rate is slightly greater than the basal export rate. The nuclear YFP signal increases rapidly within the first 1-2 minutes of stimulation and saturates at a response signal of roughly R = 1.6. This suggests that the signal propagation though the cascade is fast (order of 1 minute). Similarly, upon removal of stimulus, the nuclear YFP signal decreases rapidly over 1-2 minutes, eventually returning to the prestimulus level. This indicates that the turnover of Hog1-YFP in the nucleus is reversible, and that the turnover rate (import/export) of the proteins occurs on a timescale less than a minute or two.

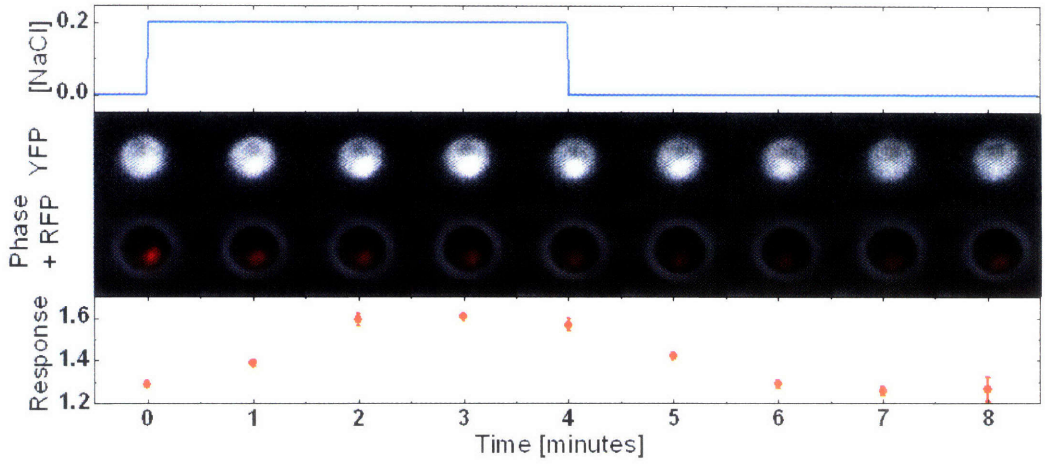


Figure 3.1.5 Localization of the fusion protein Hog1-YFP acts as a reporter of the osmotic stress pathway activity. Upon addition of extracellular NaCl, Hog1-YFP translocates to the nucleus, which is visualized by the localization of Nrd1-RFP. When the stimulus is removed Hog1-YFP rapidly translocates back to the cytoplasm. The population average translocation response (red circles) is defined by the ratio of average YFP fluorescence in the nucleus to the average YFP fluorescence. Bars represent standard error from pulse-to-pulse.

3.1.4 Frequency Response Data

Since we have demonstrated that we can quantify the response of the cells to alternating environments, we next apply simple periodic square-wave stimuli using our automatic flow cell apparatus and observe the average response in a population of cells. An example of this is shown in Figure 3.1.6, where the driving signal, NaCl concentration, oscillates between a concentration of 0 M and 0.2 M with a period of 16 minutes.

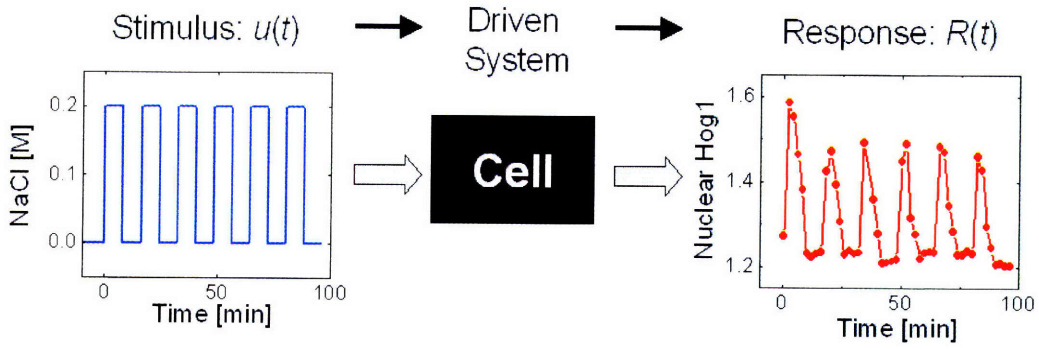


Figure 3.1.6 Square wave oscillations in the input of extracellular NaCl (blue line) with 16 min period drive oscillations of Hog1-YFP translocation with the same frequency in a population of cells (red circles).

The response of the system was measured when stimulated with square-wave pulses with periods ranging from T_0 = 2 minutes to T_0 = 128 minutes and the raw R(t) responses are presented in Figure 3.1.7. Two different behaviors are reached at the upper and lower end of this range. For rapidly oscillating signals, T_0 = 2 minutes for example, the addition and removal of NaCl is too fast for the network to respond fully to each pulse of osmolyte. Here, cells oscillate weakly at the driving frequency, and the main feature of the R(t) curve shows that Hog1 initially accumulates in the nucleus within about 2 minutes. After about 5 minutes it begins to leave, and after 10 minutes the signal has returned to prestimulus levels. This is most likely occurring because the system is averaging away the high frequency fluctuations and responding as if there were a simple 0.1 M NaCl step function input. On the other end of the spectrum, T_0 = 64min for example, the R(t) curves initially increase rapidly in response to each individual pulse, and return to prestimulus levels after about 10 minutes of the initiation of each pulse. This is occurring because the cells are in each environment long enough to fully respond and finally adapt to each pulse of osmolyte. It is also worth noting that

the cells respond in a very similar manner to each subsequent pulse, suggesting their memory of the previous pulses is short lived (less than about 16 minutes).

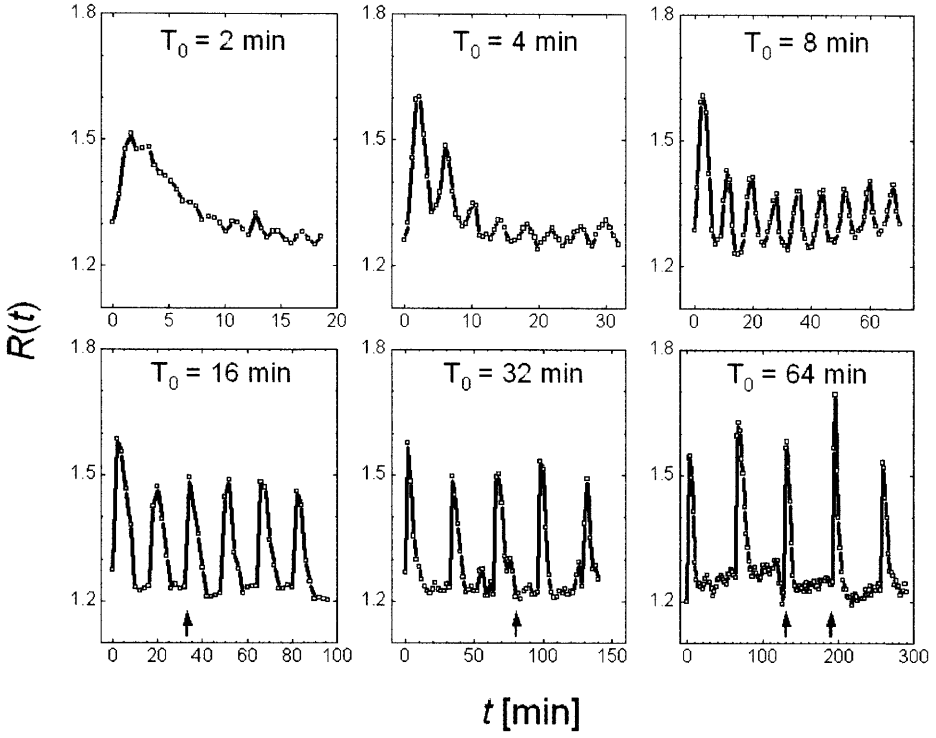


Figure 3.1.7 Example $R(t)$ traces for the wild-type strain. Cells are stimulated with 0.2M square-wave pulses of NaCl with periods, T_0, ranging from 2 min to 64 min. Black arrows indicate a new field of view.

Once steady state oscillations have been reached, the output sine wave for driving frequency ω is represented through the formula (Figure 3.1.8):

$$R_\omega(t) = y_0(\omega) + A(\omega)\sin\left(\omega t + \phi(\omega)\right)$$

This has two parameters to characterize the oscillations, A and ϕ, and one parameter which characterizes the baseline, y_0. A and ϕ are represented through the absolute value and phase of the complex number $\tilde{R}(\omega)$ respectively. This complex number is calculated from the Fourier coefficient of the experimental data, $R_\omega(t)$, taken for stimuli with period T_0 using the relation:

$$\tilde{R}(\omega) = 2 \int_{nT_0}^{(n+m)T_0} \frac{e^{-i\omega t} R_\omega(t)}{mT_0} dt .$$

The amplitude of the signal, is defined $A(\omega) = |\tilde{R}(\omega)|$, represents the half distance from the peak to the trough of the output sine wave. The phase parameter, $\phi(\omega)$, can be written implicitly as $\frac{\tilde{R}(\omega)}{|\tilde{R}(\omega)|} = e^{i(\phi(\omega)-\pi/2)}$. The parameter n is chosen so that the system is allowed to reach steady state before calculating $\tilde{R}(\omega)$. The parameter m, which represents the number of periods over which we calculate the Fourier transform, is set to be at least two for periods less than 64 minutes. For periods greater than or equal to 64 minutes, we find the first period to be a good representation of the steady state oscillations and calculate $\tilde{R}(\omega)$ over this period alone. Using this Fourier analysis, we approximate both the input and output signals as sine waves oscillating with a period $T_o = \frac{2\pi}{\omega}$. In principle, both the input and output signals also contain higher frequency components. Nevertheless, here we focus on the driving frequency alone in order to simplify analysis while obtaining significant information about the signal's strength and timing. We quantify the response by a magnitude $A(\omega)$ and phase shift $\phi(\omega)$, shown as Bode plots in Figure 3.1.9.

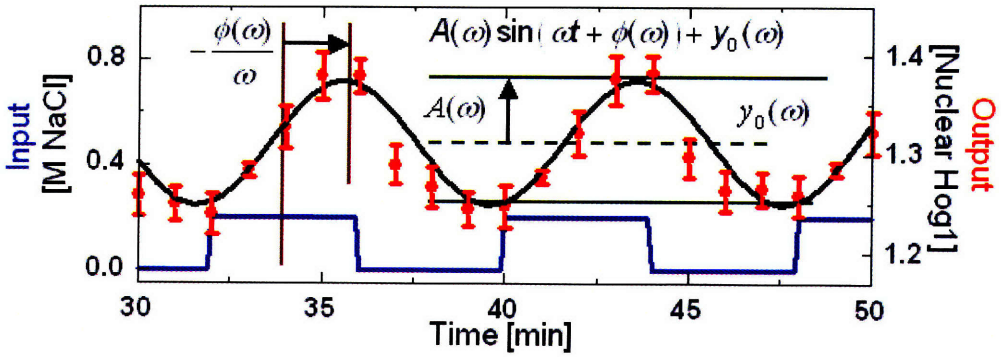


Figure 3.1.8 Illustration of the input (NaCl concentration, blue line), the network response (Hog1-YFP translocation, red circles), and the sine wave (black line) corresponding to the Fourier component of the response at the driving frequency ω. This Fourier component is described by three parameters: $A(\omega)$ (green) representing the amplitude of the oscillations, $\phi(\omega)$ (brown) representing the phase delay between the input and the response oscillations, and $y_0(\omega)$ representing the signal offset.

The main feature of the wild-type Bode plots is that around ω_c = 0.28 rad min^{-1} the slope of the $A(\omega)$ curve (green circles) rapidly changes from positive to negative on the log-log plot while the phase changes from -90 to +90 degrees (brown circles). This frequency, ω_c, separates two behavioral regimes: at high frequencies, Hog1 filters out the rapidly oscillating input, while for low frequencies, the Hog1 signal responds only to *changes* in the extracellular osmolyte rather than the concentration itself. Further, the magnitude of ω_c suggests that the relevant processes involved in the Hog1 response are occurring with timescales on the order of 10 min or less.

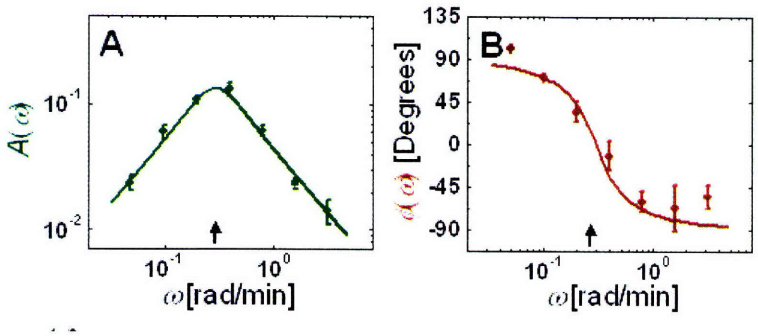


Figure 3.1.9 (A) Measurement of the Fourier amplitude $A(\omega)$ over a range of driving frequencies along with model fit to experimental data shows two different power law behaviors, ω and $1/\omega$. **(B)** Phase of the response measured relative to the driving signal along with model prediction of the phase. Error bars represent standard error.

3.1.5 Phenomenological Model

We use linear systems theory to develop a predictive model for the response to arbitrary osmotic input signals $u(t)$. We fit a general second order linear time invariant (LTI) model to the data in Figure 3.1.9 and use the parameters extracted along with a simple non-linear element to predict the response to a step input of 0.2 M NaCl. We write the linear input-output relationship in Fourier space as:

$$\tilde{Y}(\omega) = A_0 \frac{\prod_n (z_n + i\omega)}{\prod_n (p_n + i\omega)} \tilde{U}(\omega)$$

Here $\tilde{Y}(\omega)$ and $\tilde{U}(\omega)$ are the output and input Fourier spectra, respectively. The simplest such model from this class that accurately describes our data has three parameters $z_1 = 0$, and complex p_1 and p_2 yielding:

$$(p_1 + i\omega)(p_2 + i\omega)\tilde{Y}(\omega) = (i\omega)A_0\tilde{U}(\omega)$$

Applying the inverse Fourier transform produces a time derivative of u and y for each factor of $i\omega$ in the frequency-domain, resulting in the following relationship in the time-domain:

$$\ddot{y} + (p_1 + p_2)\dot{y} + (p_1 p_2)y = A_0\dot{u} \qquad \textbf{[1]}$$

A best fit analysis (Figure 3.1.9) yields values for the wild-type of A_o = 5.5×10^{-3} M^{-1}, $p_1 = (0.0026 + 0.0038i)\,\mathrm{s}^{-1}$ and $p_2 = (0.0026 - 0.0038i)\,\mathrm{s}^{-1}$, and for the mutant (Pbs2 underexpression) strain of A_o = 5.8×10^{-3} M^{-1}, p_1 = 1.2×10^{-3} s^{-1}, and p_2 = 1.6×10^{-2} s^{-1}

This LTI model predicts that the system will produce a negative output, $y(t)$, in response to a dramatic reduction in the osmotic stimulus (large and negative $\dot{u}$). However, we observe that the amount of Hog1-YFP in the nucleus does not significantly decrease below the basal level, R_0, found in the absence of stimulus. We can easily augment the model and increase its accuracy in this scenario by passing the output of this dynamic linear equation, $y(t)$, through a static non-linear element. This arrangement

of a dynamic linear system followed by a static non-linear element is commonly referred to as a LN (Linear-Nonlinear) or Wiener System (Westwick 2003). To determine the functional form of the non-linearity, we plot the output of the linear element $y(t)$ calculated from Equation 2 versus the measured data points $R(t)$ in Figure 3.1.10. The plot suggests that we relate the variable $y(t)$ to the observable $R(t)$ with a static transformation:

$$R_{\text{model}}(t) = f_{nl}(y(t)) + R_0$$

where R_0 = 1.237 is the basal response prior to osmotic stimuli, and the non-linear element $f_{nl}(y) = y(t) + |y(t)|$ acts as a rectifier to ensure that the response $R(t)$ is always greater than R_0. These functions, $y(t)$ and $|y(t)|$ are added with a slope of one because any other constant slope will be absorbed into the constant A_0. It is likely that this rectifying effect is due to the signaling elements upstream of Hog1, which only activate the downstream MAPK cascade in the presence of a positive osmotic pressure difference (hyper-osmotic shock).

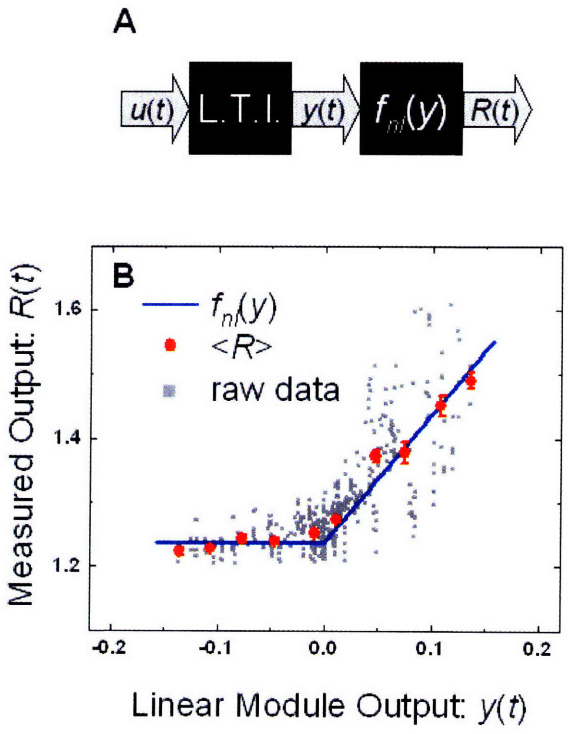


Figure 3.1.10 (A) Diagram of the Wiener or LN setup of where the quantitative fitting of our linear model is enhanced by adding a static non-linear transfer element. **(B)** Measured model output plotted versus linear model prediction. The scatter plot shows that when the linear model predicts

a negative response, the experimental system is near basal activity (gray boxes – individual points, red boxes - average). To account for this, we use a nonlinear element shown by the blue line.

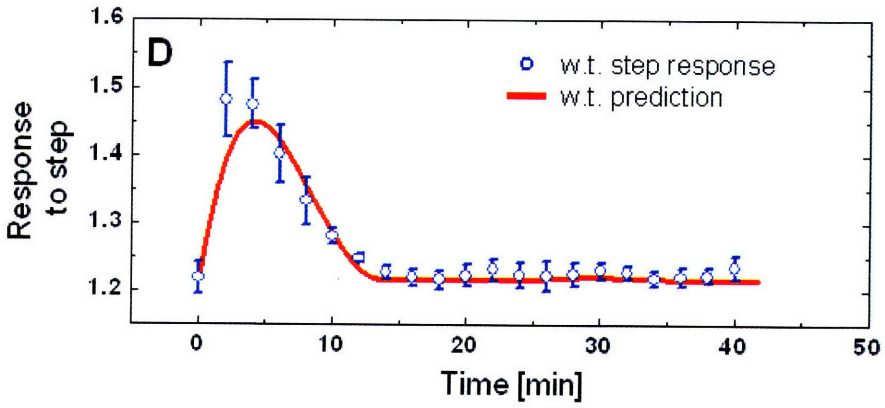


Figure 3.1.11 Response of the system to a step increase of 0.2 M NaCl compared to the step response predicted by the model.

3.1.6 Mechanistic Model

To compare our Bode-plot data with known biology, we write our second-order LTI model as two coupled first-order equations. We can write a general first order two-state model as a pair of differential equations with rate constants *a*, *b*, *c*, *d*, *e* and *f*.

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} u$$

We assume that one of the variables alone, namely *Y*, represents the observable Hog1 localization, *R*(*t*). This is similar to assuming that the dynamic variable Y is representing the activation and inactivation of the MAPK pathway, and that Hog1-YFP nuclear enrichment does not depend on other internal pathways or state variables. Therefore, *X* is the variable representing the hidden state, which does not directly influence the network output, and *u* represents the osmotic stimulus. When these equations are simplified to remove the hidden variable, *X*, we find a single second order differential equation in *Y*.

$$0 = (cb - ad)Y + (a + d)\dot{Y} - \ddot{Y} + (ce - af)u + f\dot{u}$$

When this equation is compared to the LTI model, we find that the relations $f = A_0$, $ce = aA_0$, and $c \neq 0$ must be true in order for the equations to be equivalent. Substituting these relations we obtain:

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = \begin{bmatrix} \frac{ce}{A_0} & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} e \\ A_0 \end{bmatrix} u = \begin{bmatrix} \frac{e}{A_0} & b \\ 1 & d \end{bmatrix} \begin{bmatrix} A_0 u + cX \\ Y \end{bmatrix}$$

Defining $y \equiv Y$, $x \equiv -cX$, $\alpha \equiv -\frac{ec}{A_0}$, $\beta \equiv -bc$, and $\gamma \equiv -d$ we can write these equations as:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ 1 & -\gamma \end{bmatrix} \begin{bmatrix} A_0 u - x \\ y \end{bmatrix}$$

Comparing this relation with the LTI model, we can equate coefficients to obtain the relations:

$$\alpha + \gamma \equiv p_1 + p_2$$
$$\alpha\gamma + \beta \equiv p_1 p_2$$

and equivalently:

$$p_1 = \frac{1}{2}\left((\alpha + \gamma) + \sqrt{(\alpha - \gamma)^2 - 4\beta}\right)$$
$$p_2 = \frac{1}{2}\left((\alpha + \gamma) - \sqrt{(\alpha - \gamma)^2 - 4\beta}\right)$$

Figure 3.1.12A illustrates a systems interpretation of this state-space model while Figure 3.1.12B shows the biochemical mechanisms through which osmotic adaptation is regulated. Our model contains two negative feedback loops, which act to reduce the difference, $(A_0 u - x)$, between the stimulus $A_0 u$ and the internal state variable x. One pathway increases x through the activity of our observable y (enrichment of phosphorylated Hog1), while a second pathway increases x independently of y. Based on the extensive Hog1 literature, we interpret the state variable, x, as the intracellular osmolyte concentration. Yeast cells are known to regulate their osmolyte concentration through two parallel mechanisms (Hohmann 2002; Klipp et al. 2005; Gennemark et al. 2006). First, the membrane protein Fps1 quickly (< 2 min) responds to changes in the osmotic pressure by modifying the rate at which it exports glycerol across the membrane (Luyten et al. 1995; Tamas et al. 1999). This causes cells to retain more (less) glycerol when there is more (less) osmolyte outside the cell. In addition, active Hog1 increases the expression of glycerol-producing proteins, thus increasing intracellular glycerol over longer timescales (> 30 min) in high osmolarity environments (Albertyn et al. 1994).

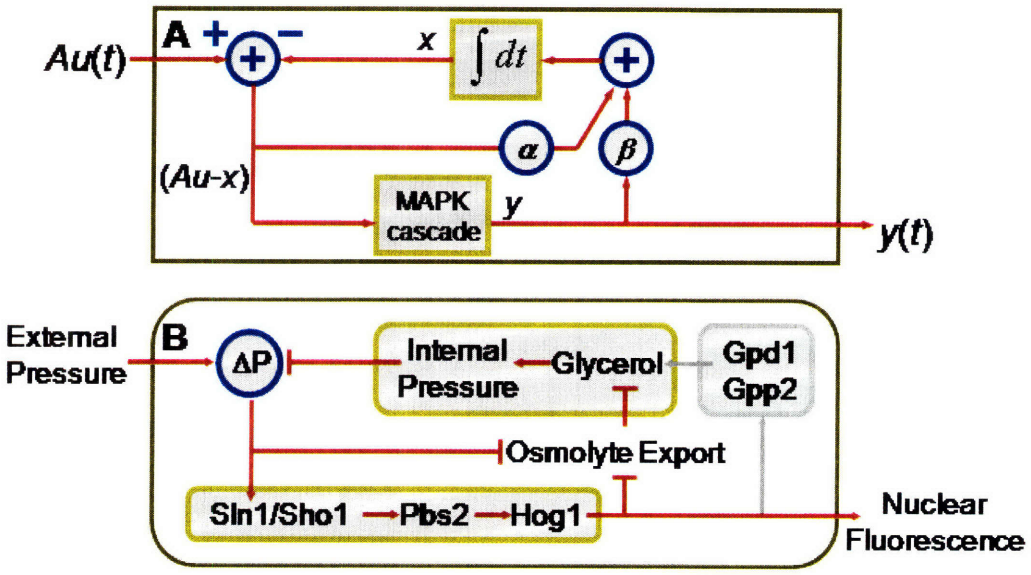


Figure 3.1.12 **(A)** Block diagram representation of the mechanistic model shows two linear negative feedbacks: one dependent on Hog1 activity with strength β, and a second independent of Hog1 activity with strength α. The intracellular osmotic pressure is modeled as an integrator, while the MAPK signal transduction pathway is modeled by the linear impulse response function. **(B)** The osmo-adaptation network structure. Upon osmotic stress, cells increase or decrease their export rate of glycerol through the trans-membrane protein Fps1, which is modified by Hog1 independent and Hog1 dependent mechanisms. In addition, under high osmotic stresses, active nuclear Hog1 is known to modify the expression of glycerol producing proteins over longer timescales.

Although our model contains a Hog1-mediated feedback loop, it is unlikely that this feedback involves protein expression because the model does not contain a delay between the activation of Hog1 and the accumulation of osmolyte that would occur if the cell first had to synthesize new proteins. Similarly, cells begin to adapt to the NaCl pulse within 5 minutes and are finished adapting within 15 minutes. Both of these time-scales are too fast for new proteins, synthesized ~15-60 minutes after Hog1 nuclear accumulation (Klipp et al. 2005), to contribute significantly to the adaptation. Furthermore, we observe that cells respond strongly and nearly identically to repeated stimuli. This would not be the case if protein production were an important factor in the observed dynamics, because proteins produced in the first pulse would facilitate the response and thus diminish Hog1 nuclear accumulation upon subsequent pulses. This suggests that both feedback loops are acting to control the rapid accumulation of glycerol, which is consistent with previous observations that activity of channels

responsible for glycerol export can be modulated by both Hog1-dependent and Hog1-independent mechanisms (Luyten et al. 1995; Tamas et al. 1999; Thorsen et al. 2006).

3.1.7 Frequency Response of Mutant Strain

To compare the gains, or relative strengths, of the Hog1 dependent and independent feedback channels, we apply the same Fourier technique to a mutant strain (DMY008) with reduced levels of Pbs2, the MAPK kinase that phosphorylates and activates Hog1 (Figure 3.1.12B). This strain is constructed by replacing the endogenous Pbs2 promoter with the P_{tet} promoter which can be induced by doxycycline in cells that are also expressing the tetracycline activator, rtTA. We find that when this strain is grown overnight in the absence of doxycycline, the response dynamics to a step input of 0.2 M NaCl are vastly different between the Pbs2 (DMY008) strain and the wild-type. However, when cells are grown in the presence of doxycycline, the response dynamics when stimulated with a 0.2 M NaCl step are indistinguishable from the wild-type response (Figure 3.1.13).

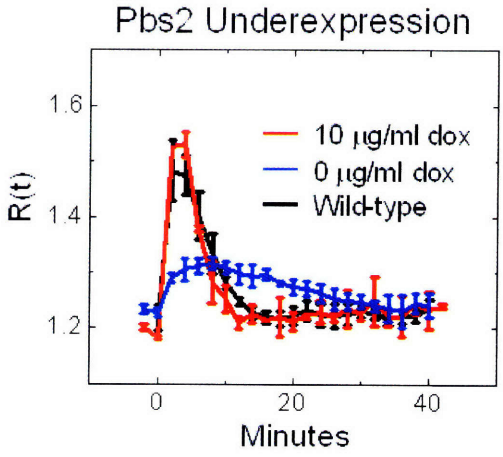


Figure 3.1.13 Pbs2 underexpression strain exhibits wild-type response dynamics to 0.2 M NaCl shock when induced with 10mg/ml doxycycline and slowed adaptation dynamics when fully uninduced. Error bars represent standard error.

Before probing the details of the signal propagation through the Pbs2 underexpression strain, we performed an additional experiment to determine which upstream branch of the signaling cascade is responsible for propagating the signal into to the MapK Hog1. Here, we compare the signaling dynamics in a strain (DMY017) where one of the membrane sensors upstream of Pbs2, namely Sho1, has been inactivated (Methods). In this strain the response dynamics are indistinguishable from the wild-type indicating that most of the signal is being propagated through the Sln1 branch. This is consistent with observations that the Sln1 branch is activated by small

osmotic shocks (0.1 M NaCl and up) whereas the Sho1 branch is only activated by much larger osmotic shocks (0.5 M NaCl) (Maeda et al. 1995).

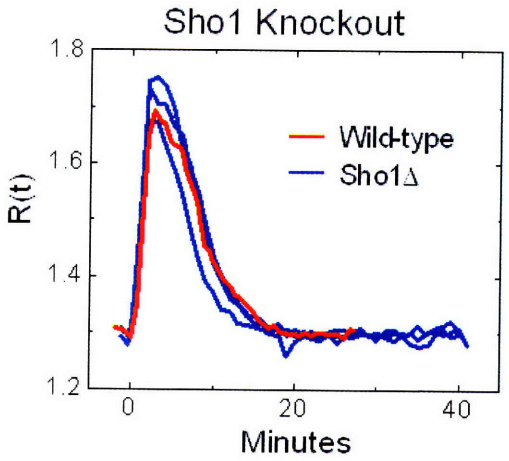


Figure 3.1.14 Response of cells with Sho1 sensor knocked out to a 0.2M NaCl shock is almost identical to wild-type. This suggests that the signal is being propagated through the Sln1 pathway.

As with the wild-type strain, we can now apply square-wave stimuli with various frequencies and measure the downstream response. These raw $R(t)$ data are shown in Figure 3.1.15 for stimuli with periods of 2 min to periods of 64 minutes. When compared to the data in Figure 3.1.7 it is clear that these response curves are dramatically different from wild-type. Namely, the maximal values of $R(t)$ are much smaller than in the wild-type experiments, while the output curves look much more like square waves with little transient response as seen with the wild-type.

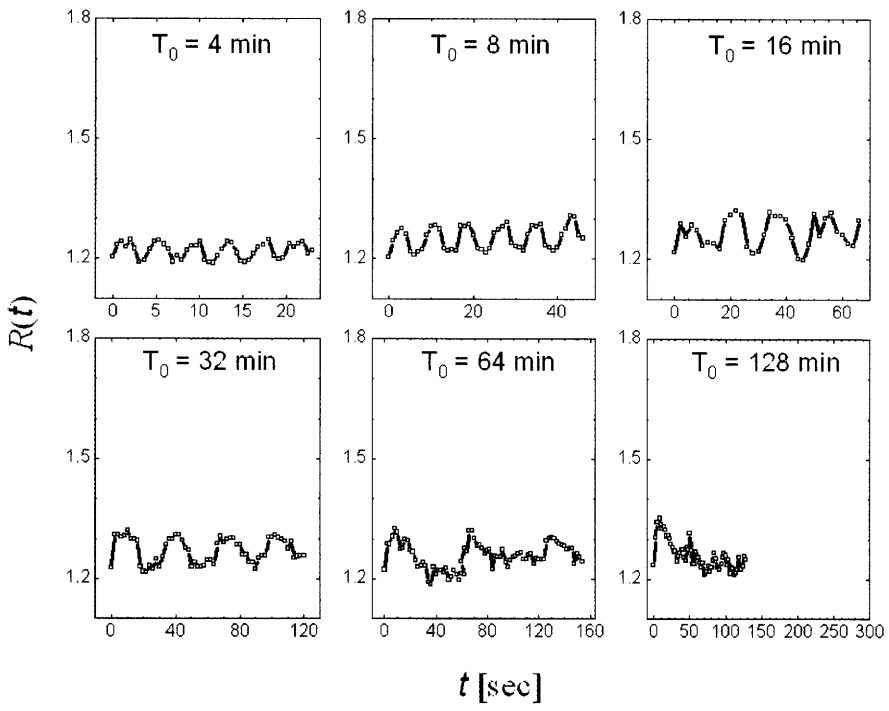


Figure 3.1.15 Example $R(t)$ traces for the Pbs2 underexpression strain (DMY008) used to calculate the $A(\omega)$ and $\phi(\omega)$ data shown in Figure 3.1.16. Cells are stimulated with 0.2M square-wave pulses of NaCl with periods, T_0, ranging from 4 min to 128 min.

Using this strain, we can effectively isolate Hog1-mediated feedback from Hog1-independent feedback by quantitatively comparing the response dynamics Figure 3.1.15 to those measured in the wild-type strain. As shown in Figure 3.1.16A-B, the Bode plots for this strain (gray squares) are notably different than for the wild-type strain. Fitting the LTI model to the $A(\omega)$ data alone yields a good fit (Figure 3.1.16A, black squares) as well as an accurate prediction for the $\phi(\omega)$ (Figure 3.1.16B, black squares) and $R(t)$ data (Figure 3.1.16C, black squares). Importantly, the $R(t)$ prediction (Figure 3.1.16C, gray line) reproduces both the reduced maximum response and the slower adaptation dynamics observed in this strain.

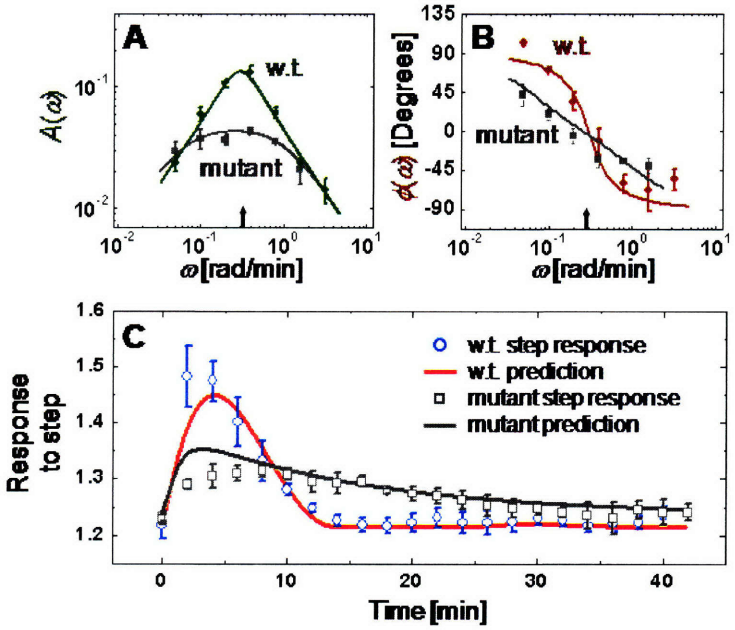


Figure 3.1.16 Response of the pbs2 underexpression mutant shows different dynamics than wild-type. **(A-B)** Mutant strain has a greatly reduced amplitude and poles shifted to the left and right of wild-type. **(C)** When exposed to 0.2 M NaCl step input, the mutant strain does not activate Hog1 nearly as strongly as wild-type, and it takes about three times longer for Hog1 activity to return to prestimulus levels.

The mechanistic model shown in Figure 3.1.12 suggests that the smaller response of the mutant strain results from a reduced activation of Hog1 compared to wild-type for the same osmotic shock. The model predicts that this will cause both the amplitude of the response, $R(t)$, and the feedback strength through Hog1 to be reduced by a similar fraction.

To model the mutant strain, we use Eqn. 1 with a simple modification, namely that the rate of activation of Hog1 is reduced due to the decreased concentration of the upstream kinase Pbs2:

$$\dot{y} = c(A_0 u - x) - \gamma y$$
$$\dot{x} = \alpha(A_0 u - x) + \beta y$$

In this equation, c represents the relative decrease in Hog1 activation rate compared to wild-type. We can define a new value of the y variable for the strain, y_{pbs2}:

$$y_{pbs2} = \frac{y}{c}$$

This allows us to rewrite the dynamic equations and the system output, R, as:

$$\dot{y}_{pbs2} = (A_0 u - x) - \gamma y_{pbs2}$$
$$\dot{x} = \alpha(A_0 u - x) + (c\beta) y_{pbs2}$$

$$R_{\text{model}}(t) = c f_{nl}(y_{pbs2}(t)) + R_0$$

From these equations, we can see two major differences from the wild-type cells. First, the parameter β is now multiplied by the factor c, reducing the feedback from y into x. Second, the response of the system is now scaled by c, reducing the amplitude of the signal output from the system. In other words, we can expect the mutant strain to exhibit a reduction in the output signal, characterized by A_0, as well as a reduction in MapK dependent feedback, characterized by the parameter β. Since both the amplitude and the time to adapt are changed roughly by a factor of three, this suggests that feedback through Hog1 is highly important in the response to increased osmolarity.

There is not enough information contained in the two dynamic parameters, p_1 and p_2, to determining α, β and γ for the wild-type or mutant strain alone. To resolve the ambiguity, we fit the LTI model defined by Equation 1 along with the static nonlinear element to the step function $R(t)$, the $A(\omega)$, and $\phi(\omega)$ data for both the wild-type and pbs2 underexpression strains simultaneously. Since the signaling cascade is operating with a decreased strength we assume that the parameters A_0 and β vary between the wild-type and mutant strains while α and γ remain the same for the two strains. We break the symmetry in the definition of α and γ by noting that HOG1 still reaches maximal activity on the order of 100 s in a strain whose FPS1 channels have been modified and can no longer adapt rapidly to osmotic shocks (Klipp et al. 2005). Table 1 summarizes the parameters resulting from this fitting procedure (Figure 3.1.17). We find that the mutant is characterized by an A_0 and β reduced from wild-type parameters by factors of (0.45 ± 0.03) and (0.24 ± 0.15) respectively.

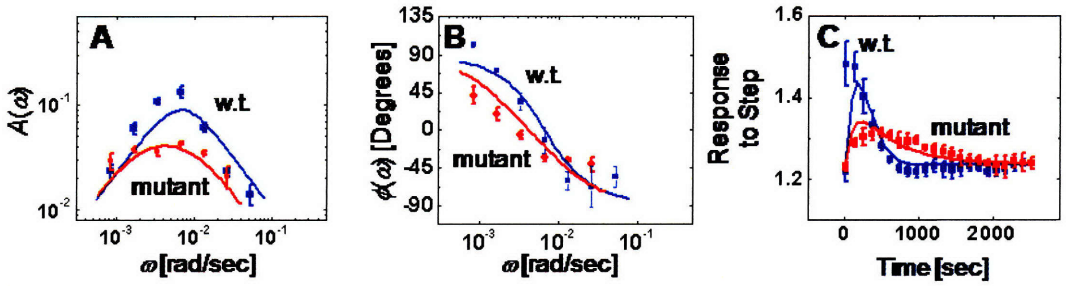


Figure 3.1.17 Results of minimizing the χ^2 cost function over the **(A)** A(ω) **(B)** ϕ(ω) and **(C)** 0.2 M NaCl step response for the wild-type (blue) and mutant (red).

Table 1. Estimates of Model Parameters

	wild type	*mutant*
A_0	$(8.0 \pm 0.3)\times10^{-3}$ (M NaCl)$^{-1}$	$(3.7 \pm 0.2)\times10^{-3}$ (M NaCl)$^{-1}$
β	$(3.4 \pm 0.5)\times10^{-5}$ s^{-1}	$(0.8 \pm 0.5)\times10^{-5}$ s^{-1}
γ	$(10.6 \pm 0.7)\times10^{-3}$ s^{-1}	$(10.6 \pm 0.7)\times10^{-3}$ s^{-1}
α	$(8.2 \pm 5.0)\times10^{-4}$ s^{-1}	$(8.2 \pm 5.0)\times10^{-4}$ s^{-1}

In addition to using the Pbs2 protein to modulate the signaling dynamics of the network, we also attempted to perturb the signal propagation at a second point, namely through the Ypd1 protein, which is part of the phosphorelay network upstream of Hog1 and downstream of the membrane sensor Sln1. In this strain, as with the Pbs2 strain, the endogenous Ypd1 promoter was replaced with the P_{tet} promoter allowing control of gene expression through the addition of extracellular doxycycline. Since we know from previous experiments (Figure 3.1.18) that the signal is propagated mainly through the Sln1 branch of the pathway, reducing the concentration of this protein should inhibit the pathway's ability to signal appropriately. We find that at low concentrations of Doxycyline (< 0.04 μg/ml) almost no signal propagates, while for higher concentrations (>0.05 μg/ml) we get almost full propagation of the signal. Unfortunately, the dramatically reduced signal magnitude prevents observing changes in dynamics.

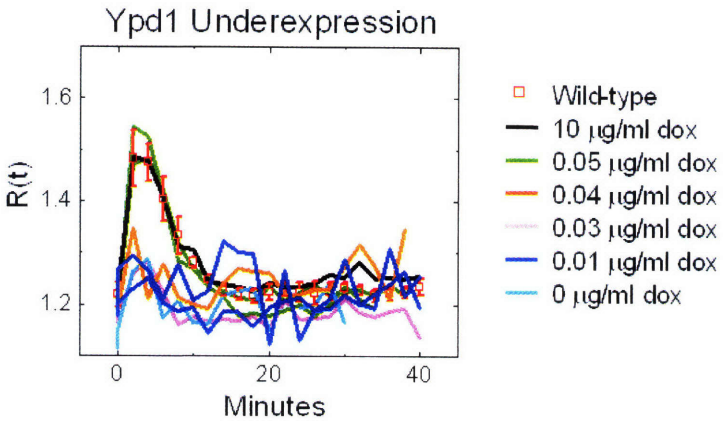


Figure 3.1.18 Response of a ypd1 inducible strain when subjected to a 0.2M shock of NaCl. Doxycyline concentrations above 0.05 μg/ml result in dynamics similar to the wild-type. Below this level of expression, however, almost no signal is transmitted through the network making it difficult to measure a significant change in signaling dynamics.

3.1.8 Network Dynamics with Cycloheximide

Although the time-scales associated with our data suggest a model where cells adapt through a Hog1-mediated feedback that does not require protein synthesis, previous studies have suggested that the creation of glycerol-producing proteins is important for osmoregulation (Albertyn et al. 1994). To further explore the role of gene expression in osmoadaptation, we measured the response of cells stimulated with periodic square wave pulses of 0.1 M, 0.2 M, 0.35 M, and 0.5 M NaCl (Figure 3.1.19). In each of these cases, cells are stimulated either in the absence (Figure 3.1.20A-D) or presence (Figure 3.1.19E-H) of cycloheximide, a small molecule that inhibits translational elongation and, therefore, protein synthesis. As predicted by the initial data, we find that cells adapt very similarly to an initial pulse of osmolyte regardless of their ability to synthesize new proteins (Figure 3.1.20, black lines). Nevertheless, we find that a longer time-scale component appears in these experiments that was absent from our earlier data. In particular, cells that are stimulated multiple times adapt to each subsequent pulse faster in the absence of cycloheximide and slower in the presence of cycloheximide.

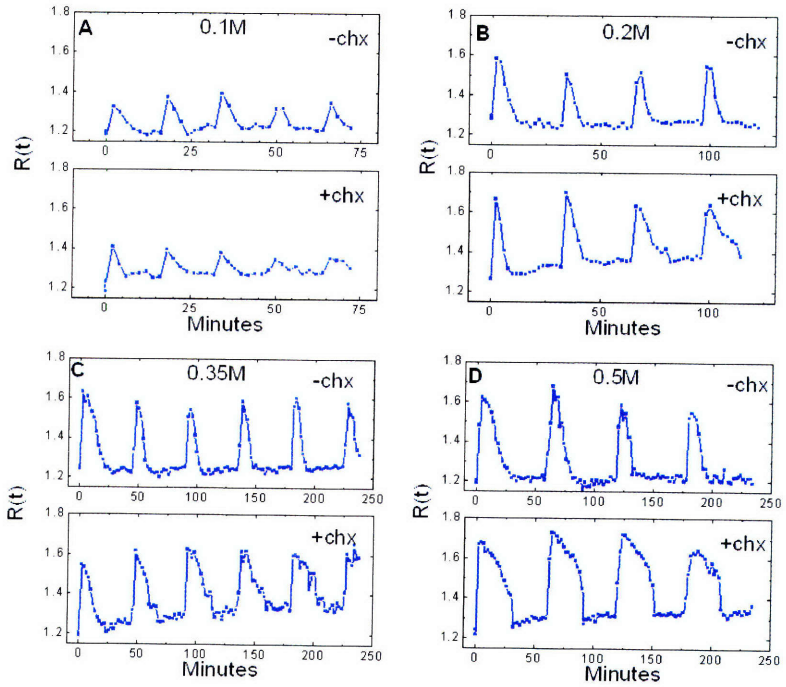


Figure 3.1.19 Raw *R*(*t*) data for square wave pulses of **(A)** 0.1 M NaCl with 16 min period **(B)** 0.2 M NaCl with 30 min period **(C)** 0.35 M NaCl with 45 min period and **(D)** 0.5 M NaCl with 60 min period. Cells were stimulated at each concentration in the presence (+chx) and absence (-chx) of 100 μg/ml cycloheximide.

These results support the hypothesis that when presented with an osmotic shock cells rapidly regulate their internal osmotic pressure through non-transcriptional feedback mechanisms such as modified glycerol transport (Klipp et al. 2005; Gennemark et al. 2006; Thorsen et al. 2006). Over longer time-scales and at larger osmotic shocks, cells increase their rate of glycerol production (Figure 3.1.21), which cause cells to adapt more rapidly to subsequent fluctuations in extracellular osmolyte. On the other hand, when the cells are stimulated in the presence of cycloheximide the glycerol production rate decreases slowly (Figure 3.1.21), impairing the cells' ability to rapidly increase the intracellular osmolyte in response to subsequent salt shocks.

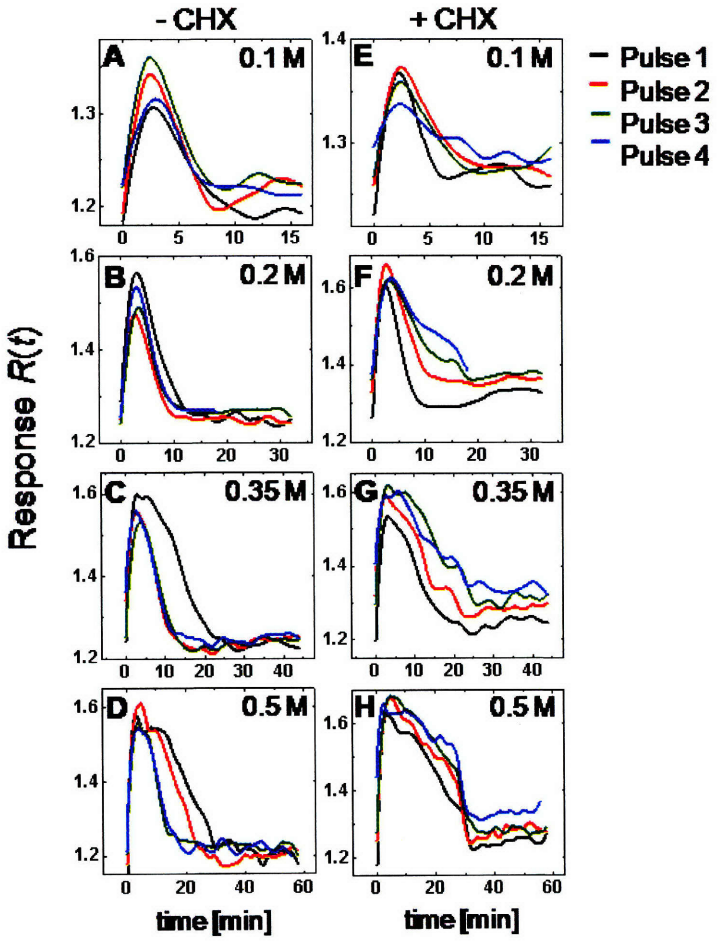


Figure 3.1.20 Cells exposed multiple times to **(A)** 0.1 M, **(B)** 0.2 M, **(C)** 0.35 M, and **(D)** 0.5 M NaCl show faster adaptation dynamics from pulse to pulse at higher osmotic shocks. **(E-H)** Cells exposed to cycloheximide during the osmotic cycling display slower adaptation rates to each subsequent pulse of osmolyte.

3.1.9 Glycerol Production Experiments

We measured the rate of glycerol production in wild-type cells when stimulated response to multiple pulses of 0.5 M NaCl (Similar to Figure 3.1.19D). These experiments were conducted either in the absence or presence of 100 μg/ml cycloheximide (Sigma). Cells were grown overnight to log phase (OD600 between 0.5 and 1.0), and at 0, 60, 120, and 180 min cells were washed and resuspended in fresh minimal media containing no NaCl. Similarly, at 30, 90, and 150 minutes cells were washed and resuspended in fresh minimal media containing 0.5 M NaCl. While suspended in 0 M NaCl media, aliquots were taken every ten-minutes, immediately boiled for ten minutes and then spun at 13,000rpm for 3 min. 20μL of this supernatant was added to 0.8mL of free glycerol reagent (Sigma). The absorbance at 540 nm was measured and normalized by the OD_{600} of the culture in order to obtain a total glycerol level relative to the number of cells present (Figure 3.1.21A-B). In this case the OD_{600} should be a good measure of relative cell numbers because cells were always measured after pulses when NaCl was not present in the media to change the size of cells. The change in total glycerol levels over 20 minutes were then used to calculate the relative glycerol production rates (Figure 3.1.21C). Cells stimulated in the absence of cycloheximide show an increase in glycerol production rate after each pulse, while cells stimulated in the presence of cycloheximide show the opposite trend.

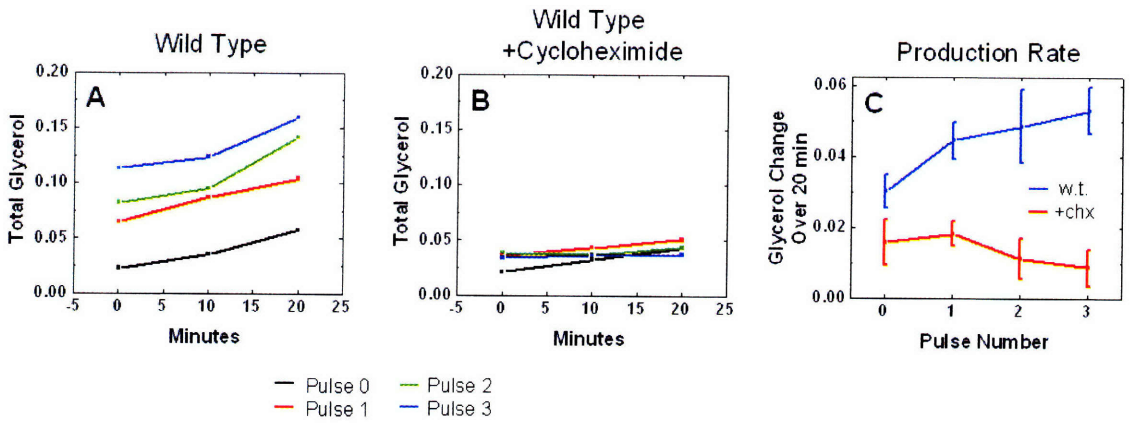


Figure 3.1.21 (A) Glycerol levels measured in a population of wild-type cells after they have experienced zero, one, two, or three 30 minute long pulses of 0.5 M NaCl. **(B)** Glycerol levels in a population of wild-type cells stimulated by 30 minute pulses of 0.5 M NaCl in the presence of cycloheximide. **(C)** Change in total glycerol levels over a 20 min observation window was used to measure the rate of glycerol production in the population for both wild-type (blue) and wild-type + cycloheximide (red). The wild-type displays an increasing rate of glycerol production after each pulse, while the cells stimulated in the presence of cycloheximide display a slowly decreasing rate of glycerol production.

3.1.10 Discussion

We used oscillatory inputs to identify the relationship between the structure and the dynamics of the yeast osmoadaptation network. This work provides further evidence for a possible role of Hog1 in rapidly and non-transcriptionally mediating the osmotic response, possibly through an interaction between Hog1 and trans-membrane channels (Proft et al. 2004; Thorsen et al. 2006). We demonstrate that cells adapt independently of new protein production over a wide range of osmotic shock strengths; however, for strong shocks, cells do produce new proteins that allow them to adapt more rapidly to subsequent osmotic changes.

By implementing a network architecture where glycerol transport is rapidly modified to adjust the intracellular osmotic pressure, cells likely gain several advantages over an architecture that relies exclusively on gene expression. First, a response mediated by extant proteins can occur more quickly than one that requires *de novo* protein synthesis. Second, negative feedback through protein production could lead to overshooting of the ideal intracellular osmolyte concentration, and such increased pressure could be dangerous for cells by causing rupturing of the cell wall. To avoid this potential danger, cells would have to rapidly degrade any new proteins produced which would waste resources by leading to a high turnover of glycerol producing proteins. Finally, upregulation of glycerol producing proteins could lead to a fundamental shift in glucose metabolism causing a possible decrease in metabolic efficiency.

3.1.11 Methods

Strain

To construct DMY007, which we refer to as "wild type", the endogenous copy of Hog1 in the haploid BY4741 background strain was C-terminally tagged with yECitrine (YFP) using standard PCR integration (Sheff et al. 2004). Colony PCR was used to confirm this and all subsequent integrations. A C-terminal GFP fusion to Hog1 was previously shown to be functional (Ferrigno et al. 1998), and the response of Hog1-YFP matches that of Hog1-GFP (data not shown). To visualize the nucleus, PCR integration was used again to fuse mRFP1.3 (kind gift from E. O'Shea and R. Tsien) to the C-terminus of Nrd1, a nuclear protein involved in transcriptional termination and mRNA processing. Finally, a plasmid bearing the tet-inducible transcriptional activator rtTA under control of the Myo2 promoter was integrated at the Leu2 locus, and PCR was used to confirm a single integration. DMY008, the pbs2 mutant strain used in this study, is identical to DMY007, except DMY008 expresses Pbs2 under the control of a TetO7 promoter. Specifically, we constructed a plasmid bearing two TetO7 promoters, one that drives CFP expression and the other that drives any gene of interest based on the primers used for PCR integration. For DMY008, this second TetO7 promoter drives Pbs2 expression. DMY028, the Ypd1 mutant strain used in this study, is constructed in an identical way to express Ypd1 under the control of a TetO7 promoter. DMY017 is identical to DMY007 except that the Sho1 protein is not expressed in this strain.

Time-series Data Acquisition

The flow chamber is constructed as shown in Figure 3.1.2. A channel is cut into an adhesive gasket with a thickness of 0.17 um (Grace Biolabs), which is then sandwiched between a microscope slide and glass coverslip. Cells in log-phase growth from an overnight low-osmolarity culture are spun down and immobilized on the coverslip with Concanavalin A. Cells are then visualized with an inverted microscope (Nikon TE2000) and 100x objective. Media is constantly removed from the flow cell with a peristaltic pump (Pharmacia) coupled to a 125mL media reservoir which dampens sudden changes in pressure. A computer-controlled valve selects the source of replacement media.

To measure the time the flow chamber takes to equilibrate, a small concentration of Rhodamine (Sigma) is added to the first media source. The average fluorescence of

a region in the middle of the chamber visualized with a 4x objective is shown in Figure 3.1.2. The rapid equilibration of the fluorescent signal indicates that the concentration of NaCl should also equilibrate in the chamber within 3-5 seconds.

In the experiments where new protein synthesis was inhibited, 100 μg/ml cycloheximide (Sigma) was added to both the low and high salt media sources of the flow chambers. After cells were attached to the coverslip, the 0 M NaCl media was allowed to flow over the cells for 5-10 minutes after which we began to apply the square wave pulses of NaCl.

Data and Image Analysis

Fluorescence image analysis for determining $R(t)$ is performed with a custom Matlab (Mathworks) script. First, cell nuclei are found by thresholding the RFP images. The phase and YFP images are then used to find the cell boundary and cytoplasm which correspond to each nuclei found in the RFP image. Finally, cells whose cytoplasmic or nuclear volume suggested incorrect segmentation were rejected. The reported $R(t)$ is the population average taken over the approximately 50-300 cells observed in the microscopes field of view. Data fitting to obtain poles and other parameters is performed in Matlab by using an optimization algorithm to minimize the χ^2 error. The parameters and error bars for parameters in Table 1 are calculated by propagating experimental error bars through the fitting procedure.

3.2 Gene Expression in Fluctuating Environments

3.2.1 Introduction

In order to cope with extracellular changes, cells often increase the expression level of key genes that function to either mitigate the effects of potential environmental stresses or to utilize new resources that have recently become available. For example, cells can produce heat shock proteins (Burdon 1986) in response to temperature increases that would otherwise adversely affect the cells ability to function. Given a particular environmental signal, cells must determine exactly how to modulate the level of key proteins responsible for enacting the cellular response. To behave in an optimal fashion, cells cannot produce too much or too little of the desired proteins, a fact further compounded by environmental fluctuations (noise) and delays associated with protein production. While much work has been done to understand the key players involved in the signal transduction and downstream gene expression in these systems, there is relatively little known about how the pathways process dynamic environmental cues in order to assure a rapid and robust response to environmental perturbations.

To address these questions, the osmotic response pathway in *Saccharomyces Cerevisiae* studied in Chapter 3.1 is used as a model system (Hohmann 2002). The network is probed under a variety of changing environmental conditions, while downstream gene expression is used as a metric instead of Hog1 localization. First, cells are analyzed in environments where extracellular osmotic pressure increases slowly. To capture the signal processing properties of the system over a wide range of timescales, the time for this increase is varied from minutes to hours. It is worth noting that this is a physiologically relevant environmental condition. It is likely that yeast cells are often present in droplets of water that slowly evaporate, leaving higher and higher concentrations of osmolytes to which the cells must continuously acclimate. By applying these slow and fast 'ramping' inputs it may be possible to learn how and why signals are processed the way that they are. Finally, by applying pulses of osmolyte that persist for various durations, some quantitative aspects of the timing of gene expression can begin to be explored. For example, it is not known whether it is the duration or the magnitude of a signal that determines gene expression levels, and these experiments begin to address this question directly.

3.2.2 Network Architecture

Gene expression in fluctuating environments is tested in the osmoadaptation system presented in Chapter 3.1, which has the advantage of being well studied (Hohmann 2002). The network diagram is shown in Figure 3.2.1, which has been modified from the previously studied strain, DMY007, by placement of a copy of GFP downstream of the Stl1 promoter. Hog1 activates the Stl1 promoter through the transcription factor Hot1 (Alepuz et al. 2003), and Stl1 is a glycerol transporting protein that is strongly but transiently induced upon osmotic shock (Ferreira, et al. Mol. Biol. Cell. 2005). Nevertheless, the expression of this gene (assayed by GFP levels) will provide a quantitative measure of the cell's transcriptional response to extracellular osmotic perturbations.

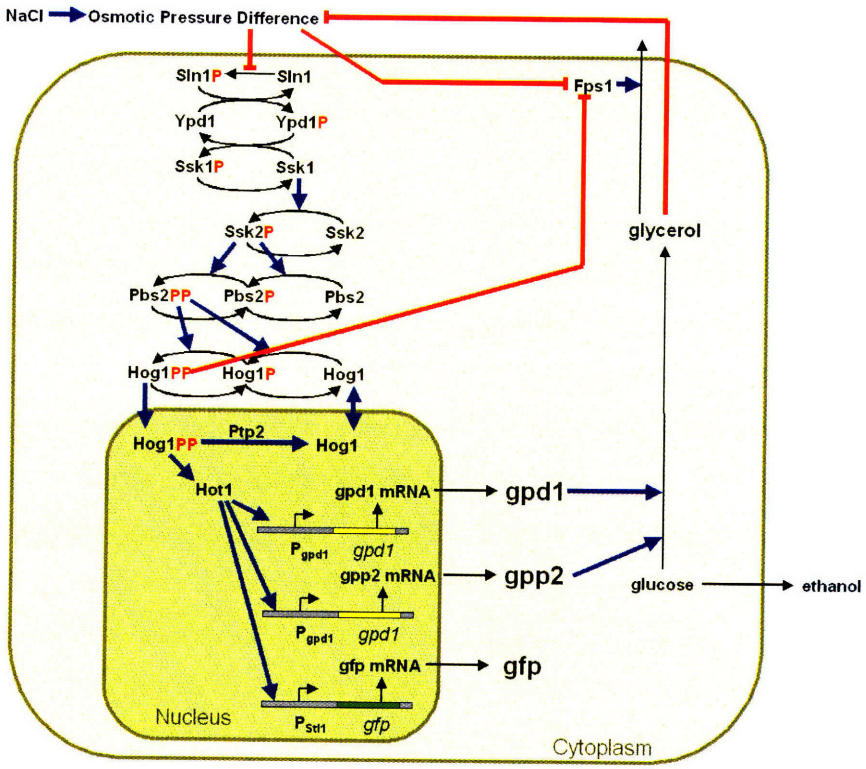


Figure 3.2.1 Network structure used for gene expression experiments in fluctuating environments. Gene expression is measured by way of GFP expressed under the control of the Stl1 promoter, which is controlled by Hog1 and Hot1 similar to Gpd1 and Gpp2 expression.

3.2.3 Experimental Setup

To measure changes in gene expression levels from cells when exposed to osmotic perturbations, it is reasonable to start with a time-series of fluorescence levels. To this end, cells were exposed to 0.5 M of NaCl at t = 0. Small aliquots of this culture were removed every 30 minutes over the course of five hours and immediately placed on ice. After 5 hours, the cellular fluorescence of each sample was assayed using flow cytometetry. Cooling the cells should prevent production of new proteins, allowing a time series of gene expression to be created (Figure 3.2.2). Here the GFP fluorescence starts around 55 units at t = 0 due to a combination of basal GFP production combined with crosstalk between the YFP and GFP channels. Note that YFP is already present in the cell as a fusion protein with Hog1. Nevertheless, YFP expression should remain constant during the experiment, and therefore any changes measured in this fluorescence level will be due to changes in GFP expression alone. The GFP fluorescence immediately begins to increase and continues to increase through the 30 minute and 60 minute time points. At 60 minutes, the GFP fluorescence peaks and then begins to decrease slowly, which is due to loss of GFP through degradation and dilution. Previous studies suggest that loss of GFP in yeast is mostly due to dilution, and the fact that the expression level is roughly 70% of the peak value after 2 hours suggests that it is being diluted since the doubling time of these cells is roughly 4 hours. The delay between induction and peak fluorescence is due to a large number of processes. First, cells must overcome the osmotic shock to a great enough degree that synthesis of new proteins can begin, the proteins must then be transcribed to mRNA, mRNA must be exported from the nucleus and transcribed in the cytoplasm. Finally, the newly transcribed proteins must correctly fold before the proteins will begin to fluoresce. Because Hog1 is only active for about 30 minutes in response to a 0.5 M NaCl shock (Klipp et al. 2005), the time dependent gene expression data suggests that it takes about 30 minutes before these processes are complete.

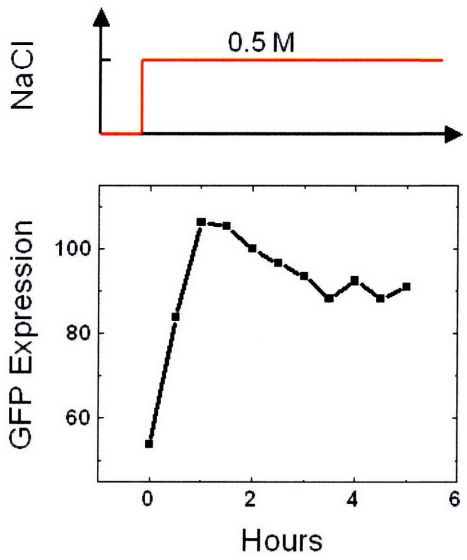


Figure 3.2.2 Gene expression measured over time. Cells were induced with 0.5M NaCl at t=0, and aliquots were put on ice every 30 minutes. After 5 hours the fluorescence of the iced samples was measured on the flow cytometer.

3.2.4 Ramping Gene Expression Data

While gene expression behaves as expected when cells are exposed to a step increase of osmolyte, it is still unknown how they will respond to more complex time varying input signals. To this end, cells are exposed to slowly increasing levels of extracellular osmotic pressure instead of simple step increases. To produce this input signal, a syringe pump was programmed to slowly add a predefined amount of NaCl to a cell culture over a long period of time (T_{ramp}) such that the final concentration was $NaCl_{Max}$. This setup is diagrammed in Figure 3.2.3.

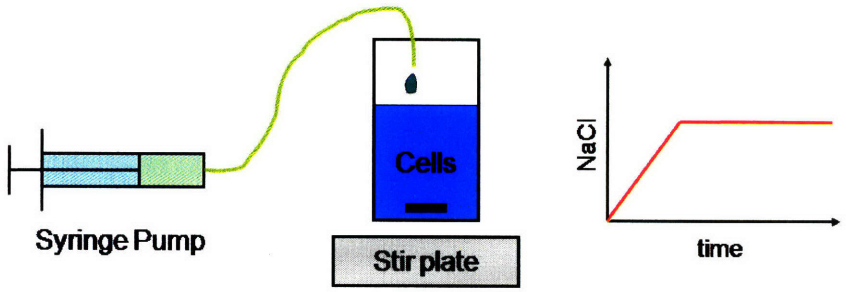


Figure 3.2.3 Experimental setup for ramping experiments. A syringe pump is programmed to add a predetermined amount of NaCl to a cell culture at a constant rate over a predefined period of time. A stir plate keeps the NaCl concentration homogeneous.

Cells are exposed to these ramping concentrations of NaCl for two different values of $NaCl_{Max}$, 0.25 M and 0.5 M. The parameter T_{ramp} is varied between 1 and 64 minutes, and aliquots of cells are taken and put on ice approximately 1, 2 and 3 hours after the beginning of the ramp. After 3 hours, the cells are taken to the flow cytometer at which point average fluorescence levels are measured from the data. These results are shown in Figure 3.2.4. As expected from the time series data presented in Figure 3.2.2, there is little change in fluorescence between the 1, 2 and 3 hour data points. The small fluctuations between comparable data points are expected as some of the data points were taken from different samples on different days. Two major features are shared by both sets of experiments. First, the expression level is relatively constant up to T_{ramp} = 10 min for $NaCl_{Max}$ = 0.25 M and T_{ramp} = 30 min for $NaCl_{Max}$ = 0.5 M. After this point, the expression levels begin to drop off and rapidly approach the basal GFP fluorescence measured for cells maintained in 0 M NaCl. The maximum level of

expression for $NaCl_{Max}$ = 0.25 M is about 20 units above background, while the maximum level of expression for $NaCl_{Max}$ = 0.5 M is about 50 units above background.

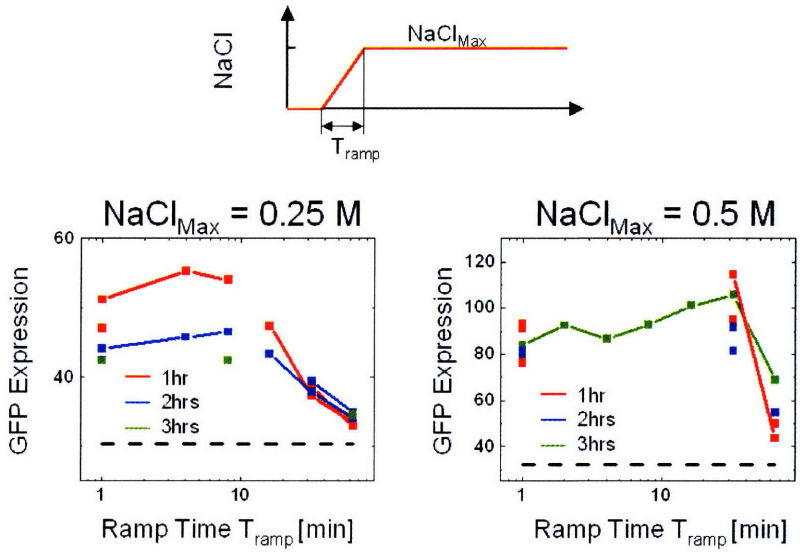


Figure 3.2.4 Gene expression in ramping environments. Cells were exposed to a linearly increasing osmotic pressure for T_{ramp} = 1, 2, 4, 8, 16, 32, and 64 minutes with a final concentration of 0.25 M and 0.5 M NaCl. At 1, 2, and 3 hours after t=0, samples of the culture were placed on ice and taken to the flow cytometer. The dashed black lines represent a control where cells remained in 0 M NaCl for the duration of the experiment indicating background fluorescence.

As a test of the flow cytometry technique used in the previous pair of experiments, cells were also measured on the microscope in order to determine if the fluorescence levels are being accurately assessed. Some of the data for cells stimulated with 0.5 M NaCl presented in Figure 3.2.4 are compared with their corresponding microscope obtained data and presented in Figure 3.2.5. The almost perfect linear correlation suggests that both measurement techniques are accurately reflecting the intracellular GFP fluorescence.

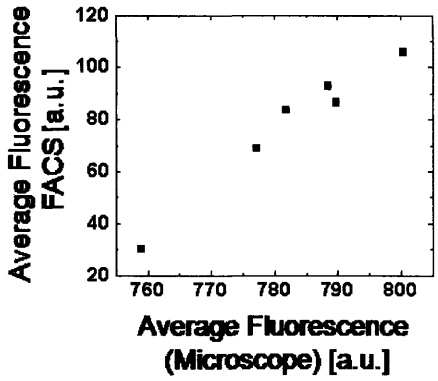


Figure 3.2.5 Flow cytometry fluorescence versus fluorescence measured on microscope shows almost perfect linear correlation between the two measures of gene expression.

3.2.5 Pulsing Gene Expression Data

To study the timing of gene expression, a second series of temporal inputs were presented to cells. Here cells were subjected to pulses of NaCl, which had variable duration (T_{step}) and variable heights. The pulses were generated by adding concentrated NaCl to cultures at t=0, and subsequently spinning down, washing, and resuspending cells in media without NaCl at time T_{step}. After two hours, cells were placed on ice and taken to the flow cytometer. The experimental method and the resulting data are presented in Figure 3.2.6. The GFP expression for all three experiments (0.1 M, 0.25 M, and 0.5 M NaCl) remain low for pulses with duration less than about 4 minutes, and then begin to increase linearly. This suggests that there is a minimum time for which the cell must be activated in order to produce new proteins. This could be due to a 1-2 min delay in propagating the signal down the MAPK cascade (Chapter 3.1) combined with delays resulting from activation of Hot1 by Hog1 and recruitment of RNA polymerase to the promoter site. When the GFP expression begins to increase linearly, all three pulse heights (0.1 M, 0.25 M, 0.5 M NaCl) seem to increase at approximately the same rate before saturating at different levels of GFP fluorescence. This suggests that in all three cases, once the production of proteins begins about 4 min after the beginning of the pulse, the production rate is roughly saturated. The difference between the different pulse heights is how long this saturated protein production lasts.

The bottom panel of Figure 3.2.6 demonstrates the average rate of GFP production over each window of observance. By taking the difference between data points in the middle panel of this figure, it is possible to ascertain how much GFP was produced due to the extra duration of the longer pulse, and therefore understand how much GFP was produced in that window of time. Here there is a striking similarity to the Hog1 phosphorylation (Hohmann, NBT) and nuclear concentration profiles (Chapter 3.1). The rate of GFP production in the early windows of time are all relatively similar for the three NaCl concentrations, however, the 0.1 M NaCl decays back to zero production after only a few minutes, while 0.2 M NaCl decays after 20 minutes, and 0.5 M NaCl decays after 30 minutes. This supports the saturated protein production hypothesis given above. By directly comparing these data with phosphorylation data, it may be possible to gain information about the signal processing properties of the transcription factor activation in addition to the MAPK activation pathway.

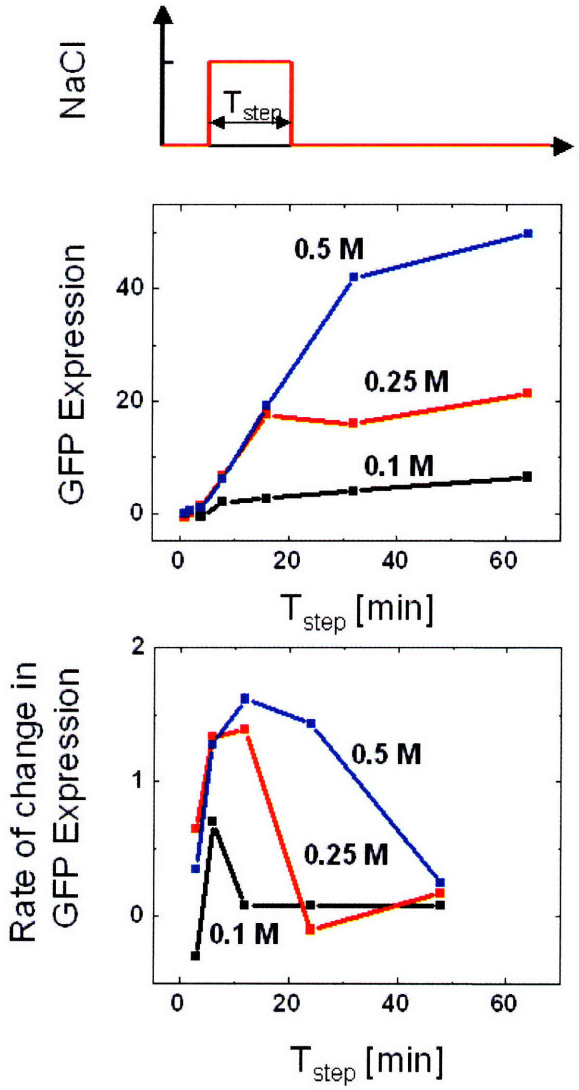


Figure 3.2.6 Gene expression resulting from pulses of NaCl of variable length. The top panel shows the input waveform with pulses of 0.1 M, 0.25 M, and 0.5 M NaCl for lengths of time 1, 2, 4, 8, 16, 32, and 64 minutes. Middle panel shows the increase in GFP expression as measured through flow cytometry two hours after the beginning of the NaCl pulse. Bottom panel demonstrates the average rate of change in GFP expression for each interval measured.

3.2.6 Processing the Signal

To understand how the signal transduction cascade processes the information in these ramping experiments one can begin by analyzing the theoretical model produced in Chapter 3.1. The basic equation for the dynamics of Hog1 translocation is given by:

$$\ddot{y} + (p_1 + p_2)\dot{y} + (p_1 p_2)y = A_0 \dot{u}$$

For slowly ramping signals, the Hog1 activity, y, will be in a quasi-steady state and therefore the equation can be approximated as:

$$y \cong \frac{A_0}{p_1 p_2}\dot{u}$$

This means that the activity of Hog1 will be approximately equal to the slope for slowly varying input signals. If the production of GFP proteins were to depend linearly on the activity of Hog1, then the number of proteins produced over the duration of the signal, P, would be proportional to:

$$P \propto \int_0^{T_{ramp}} \dot{u}(t)dt = \int_0^{T_{ramp}} \frac{NaCl_{Max}}{T_{ramp}} dt = NaCl_{Max}$$

While protein levels do seem to vary with the value of $NaCl_{Max}$ in the ramping experiments, this model predicts that the amount of protein produced is independent of T_{ramp} for slow ramp times, clearly in contrast with experimental evidence. This suggests that perhaps non-linear effects are coming into play. To see how non-linearities might affect gene expression, the rate of protein production is next assumed to be proportional to the square of the Hog1 activity:

$$P \propto \int_0^{T_{ramp}} \left(\dot{u}(t)\right)^2 dt = \int_0^{T_{ramp}} \left(\frac{NaCl_{Max}}{T_{ramp}}\right)^2 dt = \frac{NaCl_{Max}^2}{T_{ramp}}$$

Such an assumption would therefore explain both the dependence on NaCl, as well as the decrease in gene expression as T_{ramp} becomes very long. Because of the quasi-steady state assumption made earlier, this analysis cannot be used to study the

saturation observed in gene expression when T_{ramp} becomes small (Figure 3.2.4). Rather, to put this phenomenon in the context of the previously defined model, the entire dynamic model must be considered. Figure 3.2.7 shows the predicted Hog1 activity in response to sloping input signals, as well as the relative protein levels (in arbitrary units) produced by assuming protein production rates scale with the square of Hog1 activity. Without any fit parameters, this model reproduces the saturated protein expression for small T_{ramp}, as well as the sharp decline in protein levels for T_{ramp} > 10 minutes. The panel showing Hog1 activity suggests that this is occurring at short ramp times, because the Hog1 signaling pathway is filtering out the high frequency components of the signal. In other words, Hog1 activity levels cannot tell the difference between the input signal and a simple step response when T_{ramp} is less than 10 min.

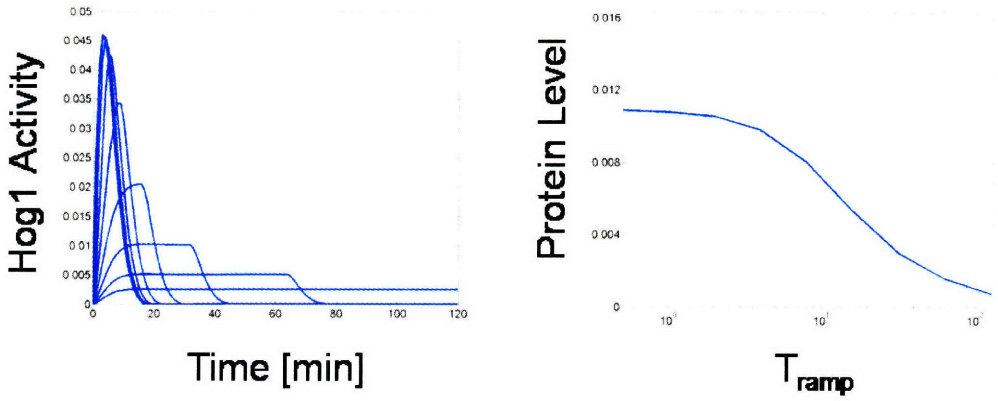


Figure 3.2.7 Calculated Hog1 activity from the model in Chapter 3.1 as well as the relative protein levels expected by assuming that gene expression is proportional to the square of Hog1 activity.

These analyses are simple theoretical explanations of what could be happening in the signal transduction pathway, and it is likely that other possible explanations also exist. Nevertheless, it may be possible to elucidate the exact role of the MAPK cascade and the transcription factors with further experiments that provide more accurate data combined with detailed measurements of the Hog1 activity in response to sloped inputs. Future work will hopefully address these questions from an experimental perspective.

3.2.7 Discussion

Although the preceding sections address how the GFP protein levels depend on temporally varying input signals, they say little if nothing about why the cell would choose to process the signals in this way. Nevertheless, there may be simple explanations that can assist in understanding the data presented in Figure 3.2.4 and Figure 3.2.6. First, a clue about the decay in protein levels with ramp times (Figure 3.2.4) may come from the time dependence of gene expression shown in Figure 3.2.2. This data suggests that some proteins may take at least 15-60 minutes following initiation of transcription to become fully functional. Although these data are for GFP fluorescence, similar delays may exist for the osmotic shock proteins. This means that the cell may constantly try to predict how many proteins will be needed at some time in the future, with this time being set by the delay between transcription initiation and protein functionality. Figure 3.2.8 illustrates this "prediction" process schematically. The cells do not know that the NaCl concentration will stop increasing as soon at it hits $NaCl_{Max}$, and therefore may assume that the concentration will keep rising. This means that they would be prone to producing higher concentrations of proteins when the concentration of NaCl rises rapidly than when it rises slowly.

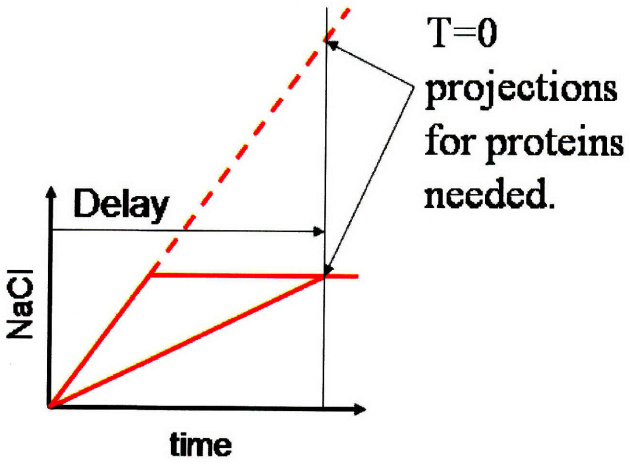


Figure 3.2.8 Cells may be projecting the amount of NaCl to expect at some time in the future. Because the cells do not know that the NaCl concentration does not increase beyond $NaCl_{Max}$, the network may be calculating what concentration to expect by the time the proteins which are begun to be produced at t = 0 become functional.

The delay in producing proteins demonstrated in Figure 3.2.6 may serve an alternate purpose. It is crucial that cells need to respond to large short-lived perturbations and long-lived but small perturbations, but ignore short-lived and small stimuli. If there are short-lived fluctuations in the environmental conditions, these should be dealt with by mechanisms other than protein production. This is because if a cell produced extra proteins every time the external environment produced a short-lived perturbation, there would be no way to remove these proteins when the perturbation ceased. Many of these perturbations in series would drive the protein levels higher than necessary for a given environment. The four-minute thresholding seen in this data may act as a type of low pass filter, ensuring that proteins are only produced when a perturbation persists for longer than a set amount of time.

3.2.8 Methods

For all experiments cells were grown in synthetic media for all experiments (CSM-his-ura-leu) containing 2% glucose. Cells were grown overnight at 30 degrees Celsius and harvested while still in early log phase growth (OD_{600}<1.0). The gene expression experiments were performed in 1 to 20mL volumes at room temperature (25 degrees Celsius). The GNY009 strain was constructed by inserting a copy of YFP driven by the Stl1 promoter into the genome. Cells were measured using flow cytometetry (FACScan; Becton Dickinson). Measurements were collected on populations of approximately 10,000 cells and population mean fluorescence was calculated on a population of about 1,000 of these cells at the peak of the forward scattering and side scattering distributions with a gate drawn by hand. Microscope measurements were performed using a Nikon TE2000 inverted microscope with 100x objective and X-Cite (Series 120; Exfo) illumination source.

4 Stochastic Effects on Cellular Decision Making

4.1 Discrete Decisions and Stochastic Transitions[1]

4.1.1 Introduction

Often biological systems do not exhibit a continuum response, but rather choose from a set of discrete pre-programmed responses to an environmental change. For example in a metabolic network the cell may need to activate a whole set of genes in order to metabolize a new molecule in the environment. Given a variety of potential metabolites in the environment, cells must choose which metabolic networks to activate in order to effectively utilize the most efficient sources of energy available. A classic example of this phenomenon occurs in the diauxic metabolism of the sugars lactose and glucose by the bacterium *Escherichia coli*, with the latter sugar being the more efficient of the two metabolites. When cells are grown in an environment containing both of these sugars, cells first metabolize only glucose until it is almost completely depleted, then they enter a lag phase at which point they 'turn on' genes responsible for lactose metabolism before finally beginning to digesting lactose. Even in the absence of glucose in the environment, cells can enter into an all-or-nothing response to lactose metabolism, producing high levels of lactose metabolizing genes in high levels of extracellular lactose or extremely low levels when lactose is absent in the environment.

In this work we study this decision making from a dynamic perspective in single cells allowing us to study the timing of activation as well as the microscopic processes responsible for driving transitions between the states. There are three main topics which we address using these experiments. First, we examine how cells change from one state to another during the decision process. Second, we measure the dominant sources of variability in a population, and assess these effects on the rates and nature of state-to-state transitions. Finally, we assess which biologically relevant parameters are necessary in order to effectively understand and quantify the decision making process.

One of the central goals of systems biology is to predict the dynamic behavior of a cell's genetic and metabolic networks. These predictions traditionally stem from models in which discrete molecular events, such as transcription and translation, are represented by continuous and deterministic differential equations. Such equations are valid when behavior of individual cells is very similar to the average behavior of the

[1] The work in this chapter can also be found in Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak and A. van Oudenaarden (2006). "Predicting stochastic gene expression dynamics in single cells." Proc Natl Acad Sci U S A **103**(19): 7304-9.

population. However, in many cases the inherently stochastic nature of biological systems leads to a significant cell-to-cell variability (Rao et al. 2002; Kaern et al. 2005; Raser et al. 2005), and previous studies indicate that individual cells often behave very differently from the population average in response to external stimuli. For example, studies of bacterial persistence indicate that the population survival rate can be fundamentally different from the average cell's survival rate in response to an environmental stress (Balaban et al. 2004; Kussell et al. 2005). The impact of noise-induced population heterogeneity is also relevant when studying cellular memory, where fluctuations in protein numbers can cause the stability of epigenetic memory to degrade by effectively causing cells to forget their original states (Hasty et al. 2000; Acar et al. 2005). In these cases, stochastic modeling techniques must be used to describe the large cell-to-cell variability.

Although deterministic models can describe dynamic network behavior and analytical stochastic models can faithfully predict steady-state population distributions (Pedraza et al. 2005), little work has been done to connect dynamic cellular behavior with noise models. It is important that stochastic models of biological systems correctly capture system dynamics since few biological systems ever reach steady state. Prior pioneering studies, aimed at predicting stochastic dynamics, have modeled stochastic effects by including high levels of microscopic detail (Arkin et al. 1998). Although this approach is correct in principle, it is often too complicated to have general applicability because the parameters required are usually unmeasured or difficult to acquire. Since these microscopic-based approaches, much has been learned about sources and propagation of noise in gene networks (Kepler et al. 2001; Thattai et al. 2001; Elowitz et al. 2002; Ozbudak et al. 2002; Swain et al. 2002; Blake et al. 2003; Paulsson 2004; Raser et al. 2004; Becskei et al. 2005; Colman-Lerner et al. 2005; Hooshangi et al. 2005; Pedraza et al. 2005; Rosenfeld et al. 2005), leading to comprehensive models of stochastic behavior. However, these previous models have lacked a sufficiently detailed set of dynamic data on which to test predictions of dynamic population distributions (Isaacs et al. 2003; Simpson et al. 2003; Vilar et al. 2003). Thus, it has still not been demonstrated that the current understanding of noise, which accurately describes distributions of protein concentrations in steady state, can be applied to predict dynamic distributions reflecting noise-induced behavior.

In this work we investigate the predictive power of stochastic dynamics by using an integrated experimental and computational approach in which we construct a

stochastic model of cellular dynamics. To test the model's predictions, we experimentally measure population distributions of protein levels over time in the lactose uptake network of *Escherichia coli* and then compare these data to the predicted population distributions. To construct our predictive model, we first build a deterministic model that incorporates relevant network components. Next, we use steady state measurements of this network to characterize the magnitude of the relevant noise sources. Finally, by combining these noise sources with the deterministic model we create a dynamic stochastic model, which is able to predict the dynamic behavior of distributions. Using this technique, we show that once macroscopic rates are known, we only need two additional parameters that characterize the noise in each gene to faithfully predict experimental dynamic population distributions without any fit parameters.

4.1.2 Network Structure

At the systems level, the lactose uptake network in a single *E. coli* cell displays an "all-or-nothing" response depending on the extracellular concentration of inducer (Novick et al. 1957). This ability to display different phenotypes at a single extracellular inducer concentration has been attributed to a positive feedback loop shown in Figure 4.1.1. LacY (purple) is a transmembrane protein involved in uptake of the inducer thiomethylgalactoside (TMG, orange). The synthesis of LacY is under the control of the *lac* promoter, which is repressed by the lac repressor, LacI (blue), in the absence of TMG. However, intracellular TMG molecules bind to LacI tetramers, causing their dissociation from DNA and an increase in *lacY* transcription. *lacY* transcription can be further activated by the cAMP receptor protein (CRP, gray), which upon association with cAMP binds to an activator site in the *lac* promoter and increases the probability of transcription. In summary, this positive feedback loop is composed of two negative and one positive connection: TMG inhibits LacI, LacI represses *lacY* transcription, and LacY increases the intracellular TMG concentration. In addition to this natural endogenous network, we constructed two fluorescent reporter systems to monitor the state of the network in single cells. The gene encoding for green fluorescent protein (*gfp*, green) under control of the *lac* promoter was integrated into the genome and reports the concentration of LacY. Additionally, we placed a red fluorescent protein gene (*rfp*, red) under the control of the *gat* promoter on a plasmid, which contains a CRP activation region and a gat repressor binding site. Because wild type K12 *E. coli* strains lack a functional gat repressor protein, GatR (Nobelmann et al. 1996), RFP is a faithful reporter for the activity of CRP (Ozbudak et al. 2004).

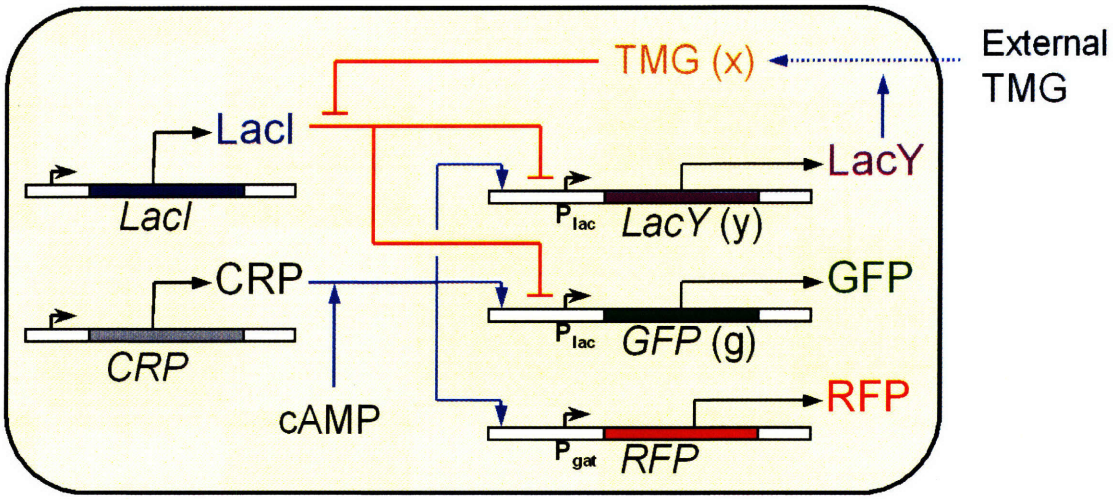


Figure 4.1.1 Diagram of the lactose utilization network. Blue arrows indicate positive interactions, red bars indicate negative interactions, and black arrows denote protein production. A positive feedback loop from LacY to TMG to LacI back to LacY creates the potential for multistability (high and low steady states). The fluorescent reporter GFP integrated in the genome is expressed in parallel with LacY under control of the lac promoter and reports the induction level of the cell. RFP under control of the *gat* promoter reports activity of the activator CRP.

4.1.3 Steady State Behavior

To model the stochastic dynamics of the lactose uptake network, it is first necessary to understand single-cell behavior. At steady state the positive feedback loop causes cells to be either in an ON (induced) state, where they maximally produce LacY and GFP, or in an OFF (uninduced) state where LacY and GFP are produced at a minimal, basal rate. In ON cells, LacY imports enough extracellular TMG to maximally produce LacY, while OFF cells do not have enough LacY or extracellular TMG to produce LacY molecules faster than they are lost. This means that cells do not generally contain intermediate concentrations of LacY and GFP when in steady state.

To observe this behavior, we prepare cells in either the ON or OFF state by growing them for 24 hours in media with 100 μM TMG or 0 μM TMG, respectively. We then remove the cells from this preparation media and subsequently grow them in fresh media containing an intermediate concentration of extracellular TMG. Cells are grown for 20 hours (about 7 cell generations), at which point the population distributions are no longer changing quickly. For cultures resuspended in very high or low concentrations of TMG, cells occupy either the ON or OFF state, respectively, independent of their induction history (Figure 4.1.2*a* and *c*). However, resuspension in intermediate concentrations of TMG maintains an ON or OFF population for extended periods of time in either the higher or lower peak, respectively (Figure 4.1.2b), indicative of hysteresis.

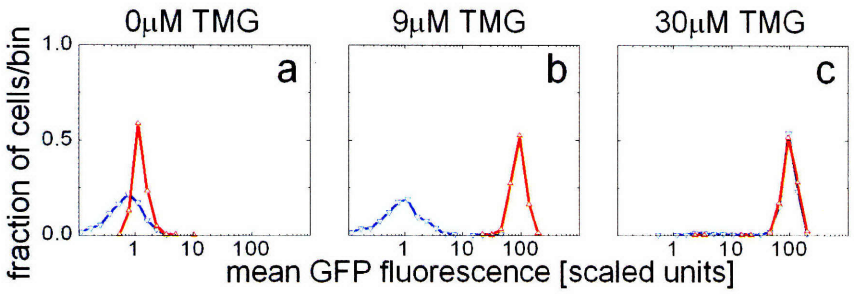


Figure 4.1.2 Demonstration of hysteresis. Histograms of mean GFP fluorescence are shown for cells with a fully induced (red) and fully uninduced (blue) history resuspended and grown for 20 hours in **(a)** 0 μM TMG **(b)** 9 μM TMG and **(c)** 30 μM TMG. Induced cells grown in 0 μM TMG for 20 hours still contain slightly higher quantities of GFP than uninduced cells, and this difference is roughly equivalent to that expected from exponential decay of fluorescence due to dilution of GFP during cell division.

4.1.4 Population Dynamics

Although we find that individual cells are either in the OFF or ON states after 20 hours, measurements at shorter time intervals must reveal transient, intermediate distributions that reflect the population morphing from its initial state (Figure 4.1.2*a*, blue curve) to its final state (Figure 4.1.2*c*, blue curve). To characterize the dynamics of these population distributions in response to changes in inducer concentration, we sample the population at various times and measure GFP levels in single cells. Fully induced or uninduced cells are washed and subsequently resuspended in media with an intermediate concentration of TMG. Next, the mean GFP fluorescence levels of individual cells are measured. Histograms are generated every one or two hours for several hours after resuspension in media with intermediate TMG concentration.

Two types of dynamic responses are observed: ballistic and stochastic. An example of a ballistic transition is shown in Figure 4.1.3a, where ON cells resuspended in 0μM TMG collectively switch OFF, drifting toward the new stable state. By contrast, Figure 4.1.3b illustrates stochastic transitions, where some initially OFF cells remain OFF while a subpopulation switches to the ON state. Characteristic of a stochastic response, the OFF peak decreases in magnitude exponentially with time.

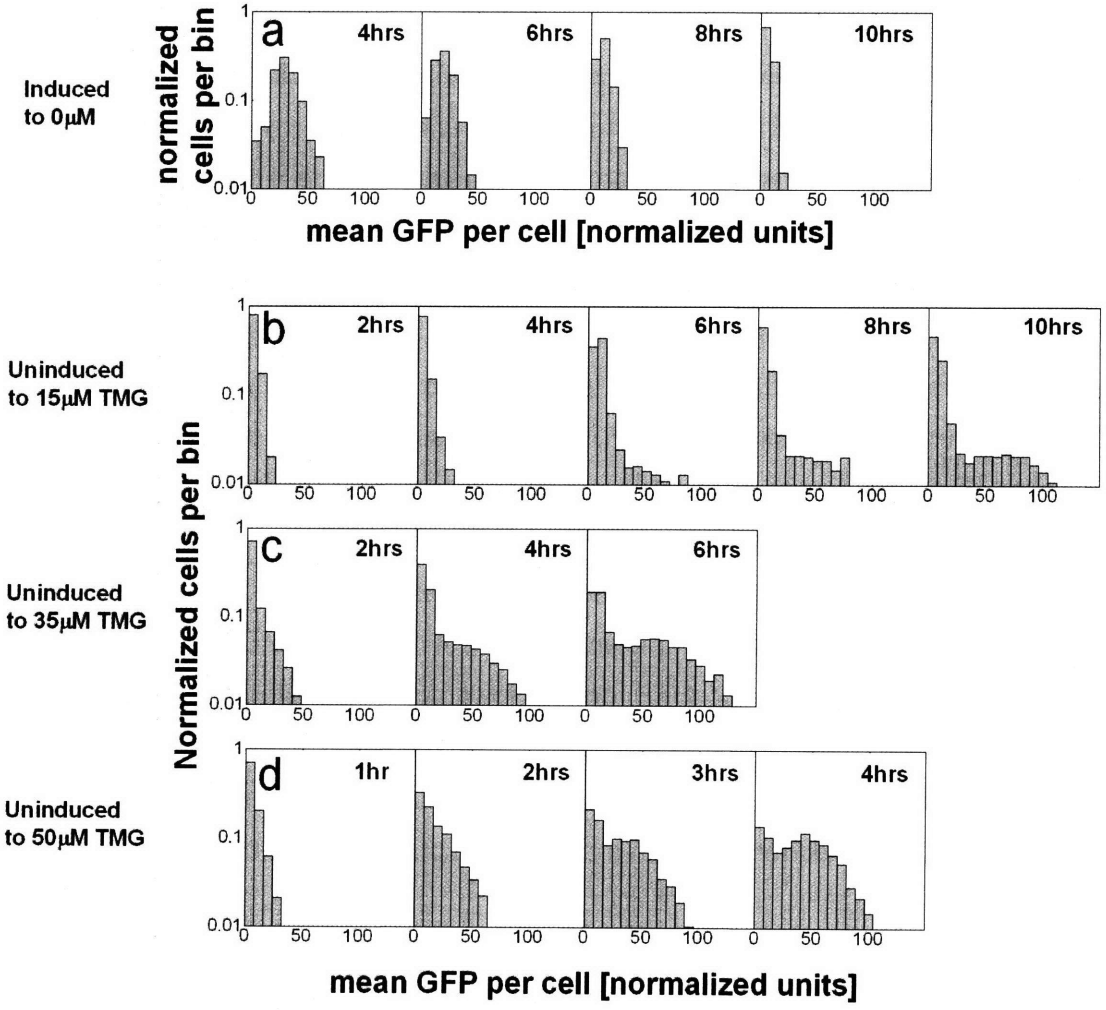


Figure 4.1.3 Transitions between OFF and ON steady states. Gray boxes are histograms of single cell GFP fluorescence for populations of 2,000-6,000 cells. **(a)** ON cells grown in 50 μM TMG and then placed in 0 μM TMG transition as a uniform population to the uninduced state with a single 'ballistic' peak. **(b-d)** OFF cells grown at 0 μM TMG are transferred to 15 μM TMG, 35 μM TMG, and 50 μM TMG. These three populations display stochastic switching behavior where cells randomly leave the uninduced state and move toward the induced state. Simulations predict the ballistic behavior associated with cells turning OFF and the stochastic behavior associated with cells turning ON.

4.1.5 Deterministic Model and Parameter Estimation

Several deterministic models have been used to explain the bistable behavior observed in the lactose utilization network (Chung et al. 1996; Wong et al. 1997; Yildirim et al. 2003; Ozbudak et al. 2004). We augment a model that has been employed to describe the strains analyzed in this study (Ozbudak et al. 2004):

$$\frac{R}{R_T} = \frac{1}{1 + \frac{X^2}{X_0^2}} \quad [1]$$

$$\frac{dY}{dt} = \frac{N_{LacY}/\tau_y}{1 + \frac{R}{R_0}} - \frac{Y}{\tau_y} \quad [2]$$

$$\frac{dG}{dt} = \frac{N_{GFP}/\tau_g}{1 + \frac{R}{R_0}} - \frac{G}{\tau_g} \quad [3]$$

$$\frac{dX}{dt} = B\,Y + \Lambda TMG_{external} - \frac{X}{\tau_x} \quad [4]$$

The fraction of active LacI tetramers (R) as a function of internal TMG (X) is modeled by a Hill function with a Hill coefficient of two. R_T is the total concentration of LacI tetramers, and X_0 represents the half-saturation point. The rate of production of LacY molecules (Y) and GFP molecules (G) in the presence of LacI (R) is also a Hill function with Hill coefficient of one. Here R_0 is the half-saturation concentration of active LacI (R) while N_{LacY} and N_{GFP} are the equilibrium number of LacY and GFP molecules, respectively, in fully induced cells. B represents the active uptake of TMG per LacY molecule, while Λ represents the passive, LacY-independent uptake of external TMG. Y, G, and X are all assumed to undergo first-order decay with time constants τ_y, τ_g and τ_x, respectively. To derive a simplified version of these equations, the following new parameters are defined:

$$g \equiv \frac{a}{N_{GFP}} G \quad [5]$$

$$y \equiv \frac{a}{N_{LacY}} Y \quad [6]$$

$$x \equiv \frac{X}{X_0} \quad [7]$$

$$\rho \equiv \frac{R_T}{R_0} + 1 \quad [8]$$

$$\beta \equiv \frac{N_{LacY} t_x}{a X_0} B \quad [9]$$

$$\lambda \equiv \frac{t_x}{X_0} \Lambda \quad [10]$$

Using these definitions and substituting for R, we rewrite our model with these three equations:

$$\tau_x \frac{dx}{dt} = \beta(TMG)y + \lambda(TMG) - x \quad [11]$$

$$\tau_y \frac{dy}{dt} = \alpha \frac{1+x^2}{\rho+x^2} - y \quad [12]$$

$$\tau_g \frac{dg}{dt} = \alpha \frac{1+x^2}{\rho+x^2} - g \quad [13]$$

Here x, y, and g represent intracellular concentrations of TMG, LacY, and GFP, respectively. *TMG* denotes the extracellular concentration of TMG. $1/\tau_x$ represents the rate of loss of intracellular TMG due to export, degradation, or dilution from cell division and growth. $1/\tau_y$ and $1/\tau_g$ represent the combined loss from dilution and degradation of LacY and GFP, respectively. The rate constant of active intake of TMG per LacY molecule is proportional to β and is a function of extracellular TMG, whereas λ represents passive uptake of TMG independent of LacY. We assume that GFP is transcribed at a rate identical to that of LacY because both are expressed under control of the *lac* promoter. Therefore, we set α as the maximal production rate of both LacY and GFP. ρ is the repression factor representing the ratio of transcription rates at the *lac* promoter between induced and uninduced cells. This factor accounts for the effect of fully activating all present LacI tetramers in the absence of intracellular TMG.

Using these equations, it has been shown that the system can have either a single stable steady state (monostable) or two stable steady states separated by an unstable steady state (bistable), depending on the concentration of extracellular TMG (Figure 4.1.4). Previously, the parameters α, β, and ρ have been determined by fitting the theoretical monostable-bistable boundaries (vertical arrows in Figure 4.1.4) to those measured experimentally (Ozbudak et al. 2004). Since the network used in our study is identical, we will use the parameters α, β, and ρ as determined in this previous study (Table 3). We set τ_y and τ_g equal to the dilution timescale due to cell growth, $\tau_{1/2} = \tau_{division} / \ln 2$, because we assume that the active degradation rate of GFP and LacY is much smaller than the dilution rate due to cell growth; here $\tau_{division}$ is the estimated average time between cell divisions. This leaves τ_x and λ as the only undetermined parameters in the deterministic model. It has been suggested that in order to model stochastic transitions in the lactose uptake network, one must include effects of TMG induction on doubling time (Vilar et al. 2003). We find that induction with TMG has no statistically significant change on growth rate; therefore, we ignore this effect in our model.

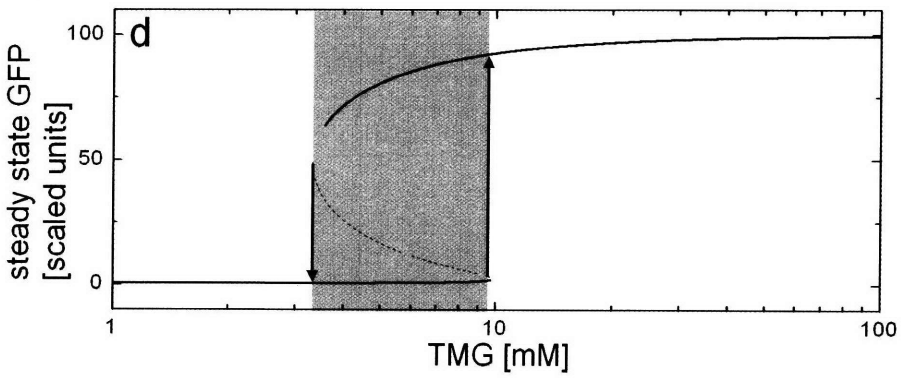


Figure 4.1.4 Steady state solutions of the deterministic model. The induced state is shown as the top dark line while the uninduced state is shown as the bottom dark line. The intermediate unstable steady state is shown as a dashed line in the shaded bistable region. Cells remain in either the induced or uninduced states until they are moved to a concentration of inducer where the previous state is unstable (vertical arrows).

We estimate the decay time of intracellular TMG, τ_x, by measuring how quickly the average GFP levels of a fully induced population decrease when placed in media without TMG (Figure 4.1.5*a*). In this scenario, LacY no longer affects the dynamics of the cells because there is no extracellular TMG to complete the feedback loop. To calculate

τ_x, the time constant of internal TMG decay, we examine a population of cells that has been grown for 24 hours in media containing 50 μM TMG, which is sufficient to force all cells into the fully induced state of high GFP expression. These cells were then transferred into media lacking TMG at t = 0. GFP measurements were made at t = 4, 6, 8 and 10 hours (Figure 4.1.5), and the average concentration of GFP was measured at each time point with an error estimated from the population standard deviation. Because there is no external TMG in the media, the cells will cease production of GFP when the internal concentration of TMG becomes sufficiently dilute. Since GFP is not actively degraded, its level should fall exponentially based on the doubling time. Therefore, we fit an exponential curve to the mean GFP concentration at each time point, using the least-squares method. The resulting best-fit curve was extrapolated to t = 0 and used to determine the approximate time at which GFP production ceased. Our extrapolation indicates that GFP production in induced cells begins to decay within 10-20 minutes following removal of external TMG, suggesting that $\tau_x << \tau_{1/2}$. Thus, we assume equation [11] is in equilibrium when compared to equations [12] and [13].

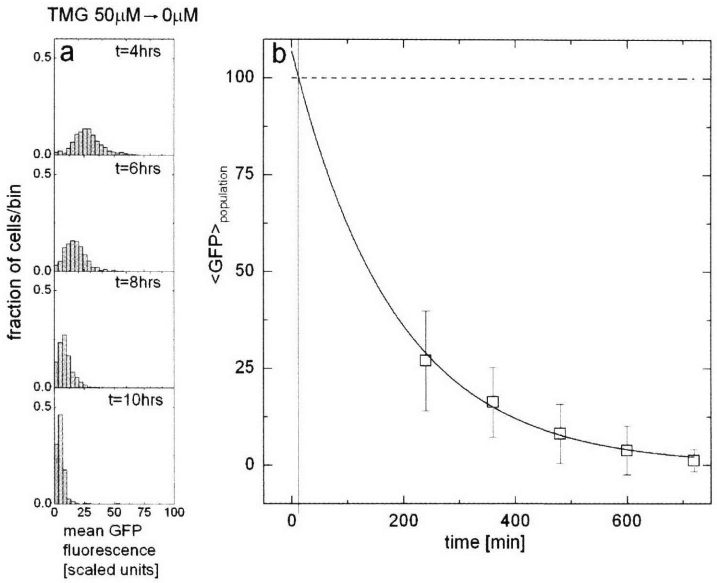


Figure 4.1.5 Estimation of τ_x through population average dynamics. **(a)** Cells in a fully induced state are placed in media without TMG and the population distribution is measured at t = 4, 6, 8 and 10 hours. **(b)** Population average of cells grown without TMG fit with an exponential. Because the intersection of the exponential (solid) with the fully induced mean (dashed) occurs within 10-20 minutes of TMG removal, τ_x is assumed to be much smaller than $\tau_{1/2}$.

Uninduced cells switched into media with a high TMG concentration are expected to transition quickly to the induced state, minimizing stochastic effects. In this situation the deterministic model should provide a reasonable estimate of this behavior at the population level. Therefore, we arrive at an estimate of λ, the rate of TMG leakage, by fitting the deterministic model to experimental data in which we observed cells transitioning from an uninduced state to a fully induced state. Cells initially grown in absence of TMG were switched into media containing 55 μM of TMG, and measurements were taken at t = 1, 2, 3 and 4 hours (Figure 4.1.6). The population average GFP level was determined from this distribution with error bars set by the standard deviation. The parameter λ was then varied from 0 to 1.0 in 0.0025 intervals, and the χ^2 between the predictions of the deterministic model and the population average was calculated at each point (Figure 4.1.6). The best-fit value was found to be λ = 0.06, with an 80% confidence range of (0.03, 0.12). Since α, β, and ρ were determined with λ = 0, a non-zero value of λ will change the position of the monostable-bistable transition and in principle require the refitting of all other parameters. However, we find that the lower monostable-bistable boundary shifts only from 3.5 to 3.4 μM TMG, which is smaller than the precision at which the boundary was originally determined. The position of the upper monostable-bistable boundary is changed as well, but the behavior of the stochastic model still matches the experiments in this region indicating that refitting is unnecessary.

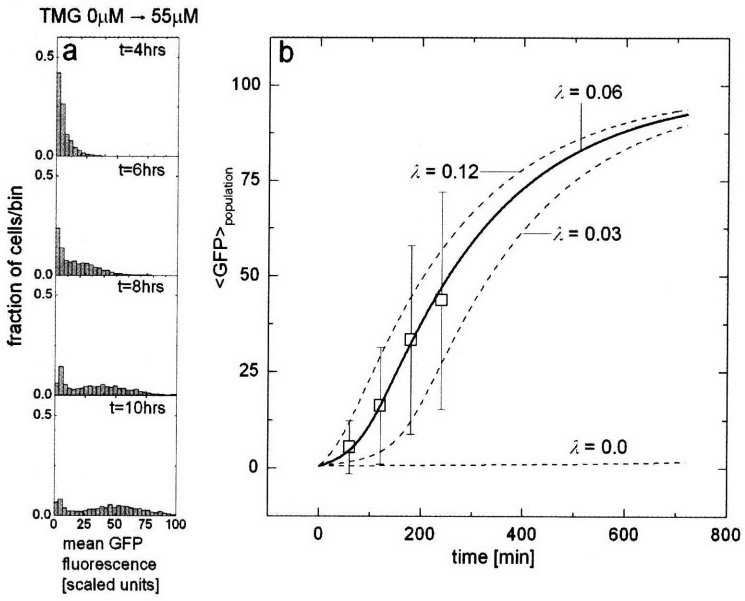


Figure 4.1.6 Estimation of λ through average population dynamics. **(a)** Cells in an uninduced state are place in media with a high concentration of TMG and measured at t = 1, 2, 3 and 4 hours. **(b)** The leakage rate λ is estimated by fitting the deterministic model to the population average of cells rapidly turning on.

We assume the parameter τ_x to be much smaller than $\tau_{1/2}$ because cells cease GFP production within 10-20 minutes after removal of extracellular TMG, and a fit of the deterministic model to the decreasing GFP levels yields $\tau_x = 0$ min. This could be due to rapid loss of intracellular TMG through efflux, which is known to occur for other inducers (Liu et al. 1999; Le et al. 2005). Thus we equilibrate the equation describing intracellular TMG in relation to the equations for LacY and GFP by setting: $\tau_x \frac{dx}{dt} = 0$.

4.1.6 Measurement of Intrinsic and Extrinsic Noise

In order to make predictions about switching transitions and dynamic population distributions, we must include the effect of stochastic fluctuations in our model. It has been shown (Elowitz et al. 2002; Ozbudak et al. 2002; Blake et al. 2003; Raser et al. 2004; Becskei et al. 2005; Colman-Lerner et al. 2005; Pedraza et al. 2005; Rosenfeld et al. 2005) that noise in protein levels comes mainly from discreteness of mRNA and protein numbers (intrinsic noise), as well as from global changes in intracellular environment that affect decay and production rates (global noise). To model intrinsic noise we must estimate two parameters for each gene: the average number of proteins produced from a single mRNA (burst size), and the conversion factor between absolute protein numbers and fluorescence counts.

Figure 4.1.7 contains a diagram indicating the generation and propagation of noise in the network. First, LacY, LacI, and GFP are sensitive to their corresponding intrinsic noise terms $\eta_{y\text{-}int}$, $\eta_{i\text{-}int}$, and $\eta_{g\text{-}int}$, respectively. These terms are due to the random creation and destruction events of *lacY*, *lacI*, and *gfp* mRNA and the corresponding proteins. The RFP reporter generates an intrinsic noise term, $\eta_{r\text{-}int}$ (Figure 4.1.1), which contains fluctuations due to RFP mRNA and RFP proteins as well as fluctuations in plasmid number. Noise generated by CRP and other factors such as RNA polymerase and ribosomes is combined into the term η_{global}, which we treat as a multiplicative factor on the production rates of LacY, GFP, and RFP.

Fluctuations in LacI numbers are propagated directly into LacY and GFP, and the strength of this transmission depends greatly on the intracellular TMG concentration since TMG binding decouples LacI from the production of LacY and GFP (Elowitz et al. 2002; Pedraza et al. 2005). Finally, noise in LacY will cause fluctuations in intracellular TMG concentration that affect the binding of LacI to the *lac* promoter. This effect causes LacY fluctuations to be transmitted into both LacY and GFP noise with a magnitude dependent on TMG concentration.

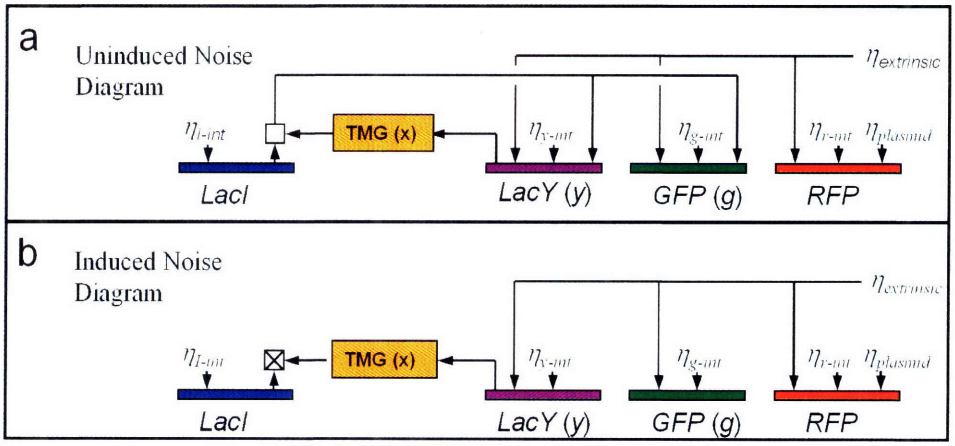


Figure 4.1.7 The noise network for the lactose utilization network. **(a)** Intrinsic noise is fed into each protein level and is propagated through the network. The square above LacI represents the combination and propagation of noise from total LacI and TMG through the active fraction of LacI tetramers, which is dependent on the concentration of intracellular TMG. **(b)** The effective noise network for induced cells with high levels of intracellular TMG, where LacI tetramers are highly inactivated. The crossed square represents the effective inactivation of this feedback by increased levels of intracellular TMG.

Extrinsic noise can be determined by examining correlations between levels of proteins influenced by the same upstream regulators. To extract the noise parameter, η_{global}, from the distribution of GFP concentrations, it would generally be necessary to solve a set of equations describing the propagation of noise through the entire network (Pedraza et al. 2005). However, the network can be simplified greatly by considering only fully induced cells, where TMG-bound and inactivated LacI no longer affect GFP and LacY production. In this case, fluctuations in LacY, GFP, and RFP expression levels are dependent only on extrinsic noise levels, η_{global}, as well as each protein's intrinsic noise level $\eta_{y\text{-}int}$, $\eta_{g\text{-}int}$, and $\eta_{r\text{-}int}$ (Figure 4.1.7). Since the term η_{global} is shared by GFP and RFP, it is possible to separate the total GFP noise into intrinsic and extrinsic components.

Each gene has a source of intrinsic noise, which is related to mRNA burst size, *b*, and protein number, *N*, by the relation $\eta_{\text{int}}^2 = \frac{b+1}{< N >}$. Because GFP expression and RFP expression share one source of noise, the correlation between the levels of these molecules will depend on this noise source and no others. To derive the relation between correlations and noise strengths, we can use the Langevin formalism in the same manner in which Pedraza *et al.* have applied it to noise propagation in a synthetic

gene cascade (Pedraza et al. 2005). Here we treat the rate of change of each molecule as having two components: continuous terms due to rates in the deterministic model and stochastic terms due to intrinsic noise. This is made more precise by writing the corresponding Langevin equations for the two extrinsic sources and the two reporters in induced cells, where LacI and LacY noise is not transmitted.

$$\tau\dot{E} = 1 - E + \mu_E$$
$$\tau\dot{G} = E\alpha_G - G + \mu_G$$
$$\tau\dot{R} = E\alpha_R - R + \mu_R + \mu_P \qquad \textbf{[14]}$$

Here E represents the collective effect of global noise factors (for example, CRP, ribosome and RNA polymerase concentrations) and is scaled to an equilibrium value of one. G and R are the number of GFP and RFP molecules, while α_G and α_R represent the equilibrium amount of GFP and RFP in induced cells. We assume that removal of these molecules results from cell division, so τ is the characteristic decay time due to dilution. The following relations define the noise terms accounting for intrinsic fluctuations of proteins.

$$< \mu_{G,R} >= 0$$
$$< \mu_{G,R}(t)\mu_{G,R}(t+\Delta t) >= 2\tau\alpha_{G,R}(b_{GFP,RFP}+1)\delta(\Delta t) \qquad \textbf{[15]}$$

We include a term μ_P to account for noise introduced by fluctuations in plasmid number, which we assume is uncorrelated with other sources of noise. Fluctuations of the global factors may be defined similarly to intrinsic protein noise, but we do not *a priori* know the magnitude of these fluctuations:

$$< \mu_E >= 0$$
$$< \mu_E(t)\mu_E(t+\Delta t) >=< \mu_E^2 > \delta(\Delta t) \qquad \textbf{[16]}$$

We want to use Eqn. [14] to derive the noise properties of our fluorescent reporters from the sources of intrinsic noise. This can be accomplished by calculating deviations from steady state values, Fourier transforming to relate fluctuations in molecules to sources of

intrinsic noise, squaring, and inverse Fourier transforming to give the relevant moments. First we subtract equilibrium values from the parameters yielding deviations from steady state: $\delta E \equiv E - 1$, $\delta G \equiv G - \alpha_G$, $\delta R \equiv R - \alpha_R$. Then we relate fluctuations in molecules to their intrinsic and extrinsic sources by Fourier transforming. Now, inverse Fourier transforming through the Wiener-Khintchine theorem yields the following result for the second moment of δG and the correlation between δG and δR:

$$< \delta G^2 >= \alpha_G (b_{GFP} + 1) + \frac{\alpha_G}{\alpha_R} < \delta G \delta R > \qquad \textbf{[17]}$$

Eqn. [17] relates the total fluctuations around steady state in GFP to a contribution from intrinsic fluctuations and a contribution from global fluctuations. This means that we can separate the total noise in GFP ($\eta_{g-total}$) into intrinsic ($\eta_{g-\mathrm{int}}$) and global (η_{global}) components by the relation:

$$\eta^2_{g-total} = \eta^2_{g-\mathrm{int}} + \eta^2_{global} \qquad \textbf{[18]}$$

Here the individual noise contributions are related to the reporter fluctuations:

$$\eta^2_{g-total} = \frac{< \delta G^2 >}{\alpha^2_G} = \frac{< G^2 > - < G >^2}{< G >^2}$$

$$\eta^2_{global} = \frac{< \delta G \delta R >}{\alpha_G \alpha_R} = \frac{< GR > - < G >< R >}{< G >< R >} \qquad \textbf{[19]}$$

The biochemical parameters α_G and b_{GFP} can be obtained from:

$$\eta^2_{g-\mathrm{int}} = \frac{b_{GFP} + 1}{\alpha_G} = \eta^2_{g-total} - \eta^2_{global} \qquad \textbf{[20]}$$

By assuming the mean fluorescence in a single cell is proportional to the concentration of molecules in that cell, we can measure total and extrinsic noise for a population using Eqn. [19], and the intrinsic noise is given by Eqn. [20]. We find that in fully induced cells the fluctuations in GFP are coming almost entirely from intrinsic noise. In this case, we could simplify our model by eliminating the extrinsic fluctuations without

greatly compromising accuracy. However, we leave this component in our model because extrinsic and global effects can often be quite large (Elowitz et al. 2002; Raser et al. 2004; Pedraza et al. 2005), and we want our model to be generally applicable without compromising ease of implementation.

GFP and RFP fluorescence levels are measured in individual cells on several populations induced with different concentrations of extracellular TMG. For 30 μM TMG this distribution is shown in Figure 4.1.8, indicating a weak correlation between GFP and RFP levels. For GFP, the total noise, $\eta_{g\text{-}total}$, and extrinsic noise, η_{global}, are measured directly from the GFP and RFP distributions in the induced population using the relations:

$$\eta^2_{g-total} = \frac{\langle \delta G^2 \rangle}{\langle G \rangle^2} = \frac{\langle G^2 \rangle - \langle G \rangle^2}{\langle G \rangle^2}$$

$$\eta^2_{global} = \frac{\langle \delta G \delta R \rangle}{\langle G \rangle \langle R \rangle} = \frac{\langle GR \rangle - \langle G \rangle \langle R \rangle}{\langle G \rangle \langle R \rangle} \qquad \textbf{[21]}$$

Again, the brackets $\langle ... \rangle$ represent the population average of fluorescent levels of single cells in the ON state only. The intrinsic GFP noise, $\eta^2_{g-\text{int}} = \eta^2_{g-total} - \eta^2_{global}$, is calculated for extracellular TMG ranging from 9 μM to 30 μM (Figure 4.1.8), and mean values of the noise measurements are $\eta_{g\text{-}total}$ = (0.25 ± 0.04), $\eta_{g\text{-}int}$ = (0.21 ± 0.03) and η_{global} = (0.14 ± 0.02). Error bars represent standard deviations over the eight measurements. These values remain constant over the range of measurement, indicating that the ON cells have similar noise characteristics regardless of the concentration of extracellular TMG.

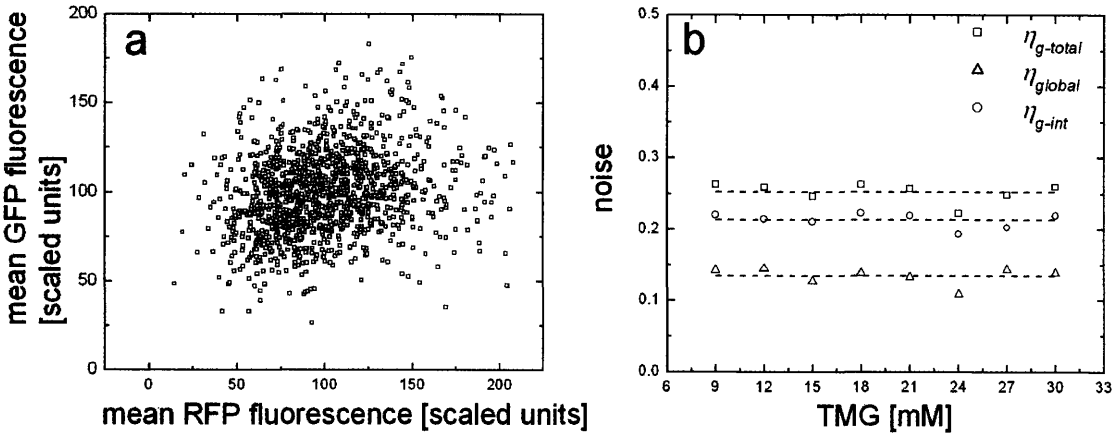


Figure 4.1.8 Measurement of intrinsic and extrinsic (global) noise is accomplished by comparing RFP and GFP concentrations in individual cells. **(a)** RFP and GFP concentrations for individual cells grown in media containing 30 μM TMG for 24 hours. A slight correlation between the two concentrations is present, indicating that the two share a weak source of global noise. **(b)** Measurement of this correlation allows division of total noise in GFP into intrinsic and extrinsic components, each of which are calculated from 9 μM to 30 μM TMG.

4.1.7 Measurement of Molecule Numbers

Decomposing the noise in a gene into intrinsic and extrinsic components still does not characterize fluctuations with sufficient detail to build a microscopic simulation; the numbers of relevant proteins in each cell is still needed. We estimate this number using a method where GFP fluorescence is compared between dividing cells. To calibrate between fluorescence and the absolute number of molecules it is sufficient to determine the number of GFP molecules in an induced cell, N (Rosenfeld et al. 2005). Rosenfeld *et al* showed that intensity fluctuations introduced by cell division vary with the number of fluorescent molecules in the cell. Upon cell division, each molecule will be independently and randomly partitioned into one of the two daughter cells. This process can be described by a binomial distribution, where the difference between the numbers of molecules in each daughter cell will, on average, be proportional to $N^{1/2}$. This means that as the number of molecules decreases (increases) the fractional asymmetry introduced by division, $N^{1/2}/N = N^{-1/2}$, will increase (decrease).

We arrive at this result by assuming each molecule of GFP is randomly and independently assigned to one of the two daughter cells. Letting N_1 and N_2 be the number of molecules in the first and second daughter, respectively, we assume that the following relations fully determine the statistics of this process: $N_{Pair} = N_1 + N_2$, $< N_1 >= \frac{N_{Pair}}{2}$, and $< N_1^2 >= \frac{N_{Pair}(1+N_{Pair})}{4}$, where averages are taken over the cell population. Further we assume that the mean fluorescence values, gfp_1 and gfp_2, in both daughter cells are directly proportional to the number of GFP molecules. Based on these assumptions, N_{GFP} can be estimated as shown below from the fluctuations of intensity between dividing cells without requiring details of photon flux or quantum efficiency.

$$\frac{< (gfp_1 - gfp_2)^2 >}{< gpf_1 >< gfp_2 >} = \frac{< (N_1 - N_2)^2 >}{< N_1 >< N_2 >} = \frac{4 < N_1^2 > -4N_{Pair} < N_1 > +N_{Pair}^2}{N_{Pair}^2/4} = \frac{4}{N_{Pair}} \qquad \textbf{[22]}$$

This equation relates the distribution of cellular fluorescence values in the daughter cells to the total number of molecules present in the mother. Because the fluorescence of an average cell is somewhere between the undivided cell (N_{Pair}) and the daughter cells ($N_{Pair}/2$), we assume that $N_{GFP} = 3/4N_{Pair}$. Finally we measure gfp_1 and gfp_2 for a

population of 70 pairs of fully induced cells in 30 μM TMG and find that N_{GFP}=790±210 molecules with error bars estimated by bootstrapping.

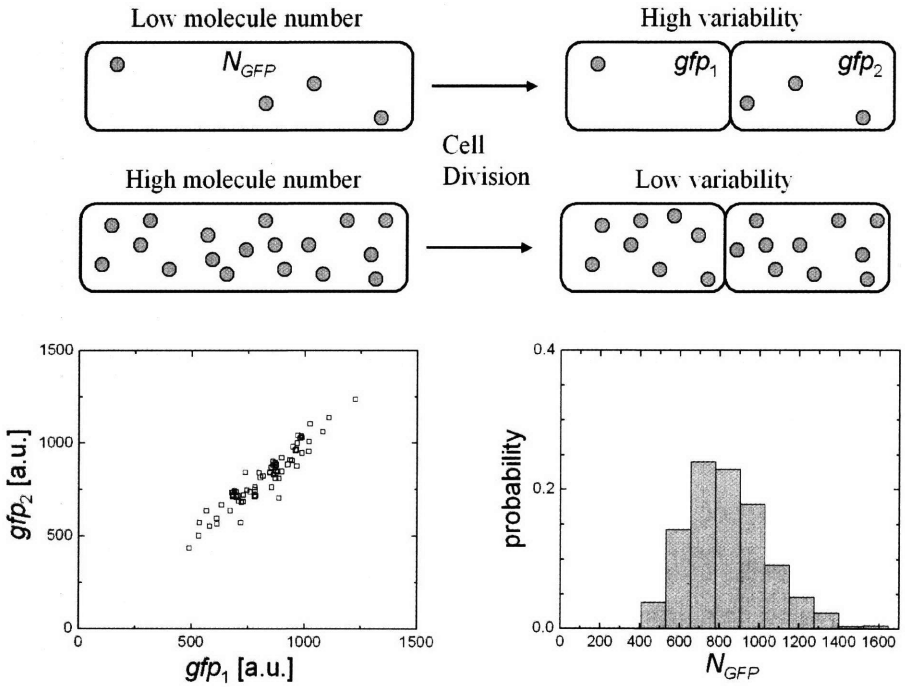


Figure 4.1.9 Procedure used to estimate the average number of GFP molecules per cell. Large amounts of proteins are more evenly divided than small numbers of proteins. In this way a high correlation between gfp_1 and gfp_2 indicates N_{gfp} is large. Using bootstrap analysis we calculate that N_{GFP} is approximately 790 ± 210 (stdev) molecules.

4.1.8 Estimation of Noise Parameters

At this point in the analysis we have measured the noise in GFP, split it between intrinsic and extrinsic noise, and measured the number of molecules in a cell using fluorescent division analysis. We next estimate the second missing noise parameter, burst size, which can be determined from the intrinsic noise, $\eta_{g\text{-}int}$, by using the relation $\langle \delta n^2 \rangle = \langle n \rangle (b+1)$ (Thattai et al. 2001; Ozbudak et al. 2002) where n is the number of protein molecules and b is the average number of proteins produced from a single mRNA. This relation stems from the fact that translation effectively amplifies the noise associated with low levels of mRNA. We find that the burst size for GFP is $b_{GFP} = 35.3 \pm 9.7$, consistent with burst size estimates of other genes in *E. coli* (Pedraza et al. 2005). Because the same promoter is regulating GFP and LacY expression, we assume that the production rate of mRNA should be similar for the two proteins. Thus we set $\frac{N_{LacY}}{b_{LacY}} = \frac{N_{GFP}}{b_{GFP}}$, where N_{LacY} is the number of LacY molecules in a fully induced cell, and b_{LacY} is the burst size of a LacY mRNA. We could reduce the ratio N_{LacY}/b_{LacY} to a single parameter analytically since the mRNA production rate for LacY is proportional to N_{LacY}/b_{LacY} and the burst size in units of fluorescence is proportional to $(N_{LacY}/b_{LacY})^{-1}$. However, in order to proceed later with simulations that model explicit molecular events, we need to assign values for both N_{LacY} and b_{LacY}. Therefore, without loss of generality, we choose values for these parameters that maintain the required ratio by setting them equal to those measured for GFP: $N_{LacY} = N_{GFP}$ and $b_{LacY} = b_{GFP}$.

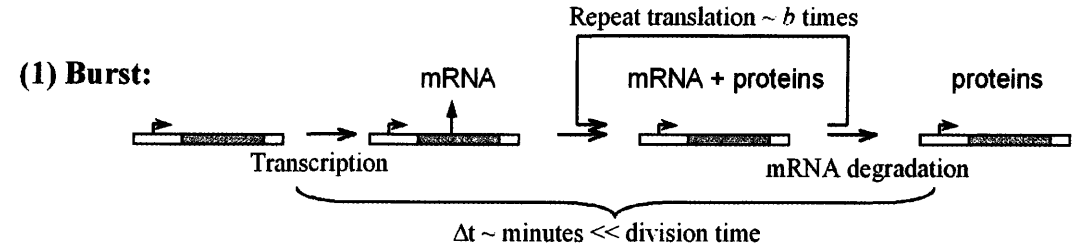
In principle one can also measure the noise from other extrinsic sources such as LacI, which propagates its total noise into the GFP and LacY protein levels. A similar measurement been performed in two previous studies (Elowitz *et al.*, 2002; Pedraza and van Oudenaarden 2005) in which there was no positive feedback loop from LacY. These works find that the amount of noise transmitted is dependent on the concentration of internal TMG. As the LacI level changes, the rate of change of LacY production determines how much noise is transmitted. A slope of zero will transmit no noise because small fluctuations in LacI level do not change the LacY production rate. As the slope increases the amount of fluctuations transmitted increases proportionately.

In our network the fully induced cells block noise from LacI, which means that transmitted noise from LacI is only observable in uninduced cells. However, at zero induction the noise in auto-fluorescence becomes comparable to the noise in the gene

and the measurement cannot be made. To get around this we use published estimates of the intrinsic noise determining parameters for LacI (Muller-Hill *et al.*, 1968). We assume the burst size in this estimate is not changed in our strain, and scale the mean number of molecules by the division time of the cell. We thus interpret these parameters as $N_{lacI} = 50$ and $b_{LacI} = 5$.

4.1.9 Stochastic Model

We now have a full set of parameters for the dynamic deterministic model, as well as all parameters necessary to describe the steady state noise properties of the lactose uptake network. In order to determine if these parameters can be used to predict the full dynamic distributions measured in the experimental section, we construct a dynamic stochastic model. In principle we have already fully defined the model, but for simulation purposes, we use a reduced model that contains only essential events. The model consists of three processes: (1) mRNA production followed immediately by a burst of protein production and mRNA degradation, (2) protein degradation, and (3) extrinsic or global noise. We therefore ignore specific events such as binding of LacI or CRP to operator sites and timing between productions of individual protein molecules.



(2) Protein Degradation:

$$\text{protein} \longrightarrow \emptyset$$

(3) Extrinsic Noise

$$\text{Extrinsic 'factor'} \longrightarrow \emptyset \qquad \text{and} \qquad \emptyset \longrightarrow \text{Extrinsic 'factor'}$$

Figure 4.1.10 Diagram depicting the three processes involved in the stochastic model.

We assume the number of proteins produced from an individual mRNA to be chosen from an exponential distribution by treating the decay of mRNA as a random Poisson process and assuming that the number of proteins translated is proportional to the lifetime. Because the lifetime of an mRNA is very short in relation to the timescales associated with fluctuations in protein level ($\tau_{1/2}$) we condense the three events (production, translation and degradation) into a single 'burst' event. We model this

process by production and immediate decay of an mRNA whose net effect was the addition of a random number of new proteins to the system. We quantify the rate at which these bursts occur by dividing the rate of protein production by the mean number of proteins produced from an mRNA.

We model the loss of protein levels by random Poisson decay of individual proteins at a rate commensurate with that caused by dilution. This adds noise to the system that is not inherent in cell growth and ignores noise due to the partitioning process. However, the noise difference between the decay and dilution processes should not be important because the noise from mRNA bursts is dominant in our experiments. This could become relevant for dynamics that are heavily influenced by decay processes when there are few if any mRNA burst events.

Global noise is included as a multiplicative term on the production of GFP and LacY. To include this effect, the random walk process *E(t)*, described in Eqns. [14] and [16], is simulated and included in the relevant mRNA production rates. Here the value $<\mu_E^2>$ is chosen so that the simulated value of η_{global} matches that measured experimentally. The statistics on the process are thus given by $\langle E \rangle = 1$ and $\langle \delta E(t)\delta E(t+\Delta t)\rangle = 2\eta_{global}^2 \exp(-\Delta t/\tau_{1/2})$. This is then numerically simulated by a discrete random walk, and the mRNA production rates of relevant proteins are multiplied by the resultant factor $E(t)$.

Table 2. Stochastic Model

Protein	*Event*	Reaction	*Rate*
GFP	Burst	$G \rightarrow G + B(b_{GFP})$	$f(x(Y), I)\, N_{GFP}/(b_{GFP}\, \tau_{1/2})\varepsilon$
	Decay	$G \rightarrow G - 1$	$G/\tau_{1/2}$
LacY	Burst	$Y \rightarrow Y + B(b_{LacY})$	$f(x(Y), I)\, N_{LacY}/(b_{LacY}\, \tau_{1/2})\, \varepsilon$
	Decay	$Y \rightarrow Y - 1$	$Y/\tau_{1/2}$
LacI	Burst	$I \rightarrow I + B(b_{LacI})$	$N_{LacY}/(b_{LacY}\, \tau_{1/2})$
	Decay	$I \rightarrow I - 1$	$I/\tau_{1/2}$

Where the following functions model the effect of LacI repression based on instantaneous values of *Y* and *I*:

$$f(x, I) = \left(\frac{(r-1)I}{N_{LacI}(1 + x^2)} + 1\right)^{-1}$$

$$x(Y) = \left(Y \frac{\alpha b}{N_{lacY}} + \lambda\, TMG_{external}\right) \qquad [23]$$

Here ε represents a correlated extrinsic noise factor with:

$$\langle \varepsilon \rangle = 1$$

$$\langle \delta\varepsilon(t)\, \delta\varepsilon(t+Dt) \rangle = 2\eta^2_{extrinsic}\, e^{-Dt/t} \qquad [24]$$

B(b) is an exponentially distributed random integer with mean equal to b. Capital variables are defined to be absolute molecule numbers of the respective lower case concentrations scaled to the deterministic model for example:

$$G \equiv g\, N_{GFP}/\alpha$$

$$Y \equiv y\, N_{GFP}/\alpha \qquad [25]$$

Final parameter values are listed in (1) –Ozbudak *et al.*, 2004; (2) – Muller-Hill *et al.*, 1968

Table 3:

Parameter	*Value*	*Error*	*Reference*
α	100	16	*(1)*
β	$0.123(TMG)^{0.6}$	15%	*(1)*
ρ	170	34	*(1)*
λ	0.06	0.04	
τ_x	0 min	20	
$\tau_{1/2}$	216 min	43	
N_{LacY}	790 proteins	210	
N_{GFP}	790 proteins	210	
N_{LacI}	50 proteins		*(2)*
b_{LacY}	34.8 proteins	10.1	
b_{GFP}	34.8 proteins	10.1	
b_{LacI}	5 proteins		*(2)*

(1) –Ozbudak *et al.*, 2004; (2) – Muller-Hill *et al.*, 1968

Table 3: Table of parameters used in the deterministic and stochastic models

To justify modeling mRNA production and decay as a single event, we compare simulations of two similar models. The first 'microscopic' model describes for the individual mRNA and protein production and decay events. The second 'bursting'

model, as described above, simulates mRNA bursts and protein decays only. Figure 4.1.11 shows the schematic for each model and the corresponding rates used, and the results of the simulations are shown in Figure 4.1.12. We see that although many of the processes are ignored in the 'burst' model, it still reproduces the equilibrium statistics of the 'microscopic' model. This agreement between the two models should still hold when multiple species are simulated. However, the approximation would break down if the lifetime of mRNA becomes nearly as long as the division time of the cell, but this is an unlikely scenario in our system.

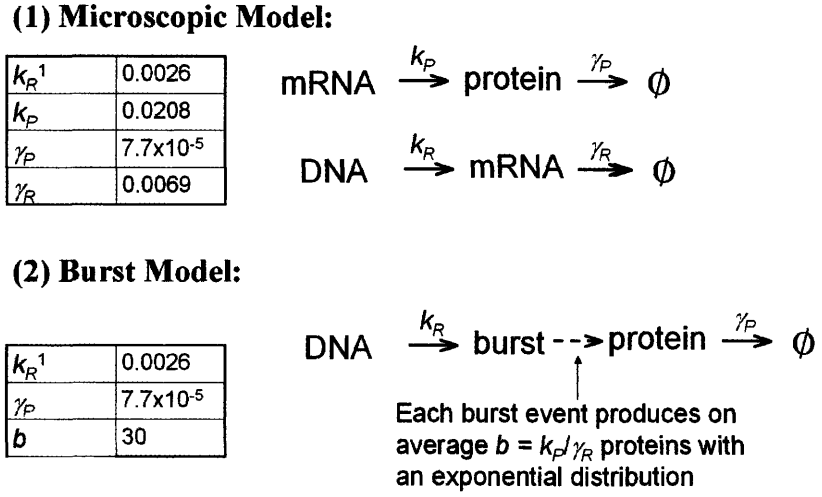


Figure 4.1.11 Comparison of detailed DNA to mRNA to protein model with bursting model where protein is produced in exponentially distributed bursts

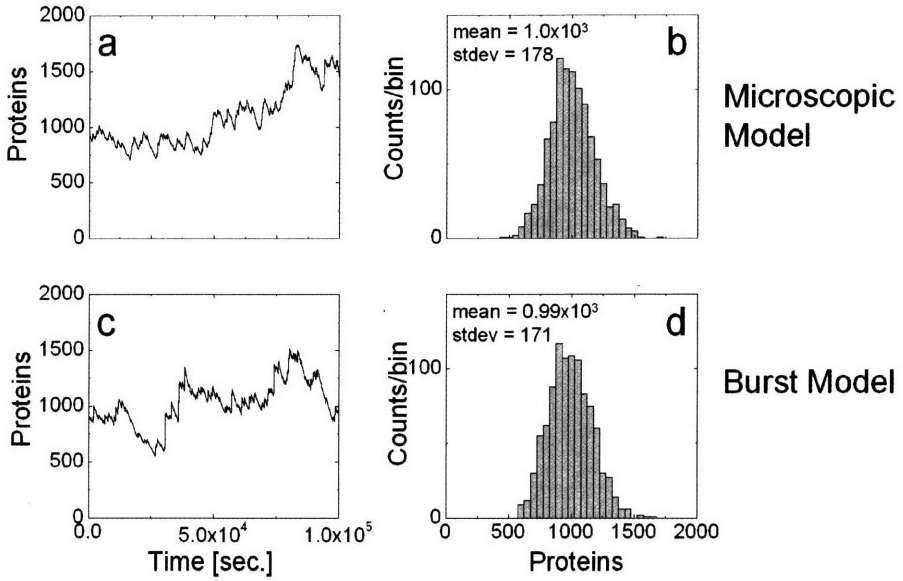


Figure 4.1.12 The 'microscopic' protein production model is relatively indistinguishable from the 'bursting' protein production model.

4.1.10 Stochastic Model Results

The full stochastic model can now be simulated using Gillespie's Monte-Carlo algorithm (Gillespie, 1977). This method treats each production or degradation event as an independent and random process (Poisson process), which gives an exponential waiting time until the next event. This time is estimated and the number of molecules and current time are updating according to whatever event (i.e. LacY burst, GFP degradation, etc.) occurs first. Between events the cell remains in exactly the same state with all molecule numbers preserved. Once the number of molecules is updated, the rates are changed to reflect the new state and another event is simulated. To initialize the system, cells are put into either the on or off state and simulated for 12 hours to allow the population to equilibrate.

Several simulation time series for this model are shown in Figure 4.1.13; here five single cells have been simulated in each of two external conditions. Panels a and b represent induced cells placed in 0 μM TMG, and panels c and d show uninduced cells in 35μM TMG. This figure shows that GFP is a good reporter for LacY level as both molecules show similar behavior in both circumstances. Both GFP and Lacy show similar levels of noise as their noise parameters were chosen identically, however there is a striking difference between the two transitions. Transitions from induced to uninduced (Figure 4.1.13a and b) show smooth curves with exponentially decaying protein levels. Also, most of the trajectories are very similar, differing only by about 100 molecules through the time series. The transitions from uninduced to induced, however, do not display this smooth behavior, and each cell acts very different from the other cells and the population mean. The differences in 'smoothness' can be attributed to the difference between noise in protein production and noise in protein decay. This can occur, for example, when the production of proteins ceases and the decay and dilution processes dominate the concentration kinetics. Because there is no protein production and therefore burst events, the noise in the induced to uninduced transition is largely controlled by the number of molecules rather than burst size.

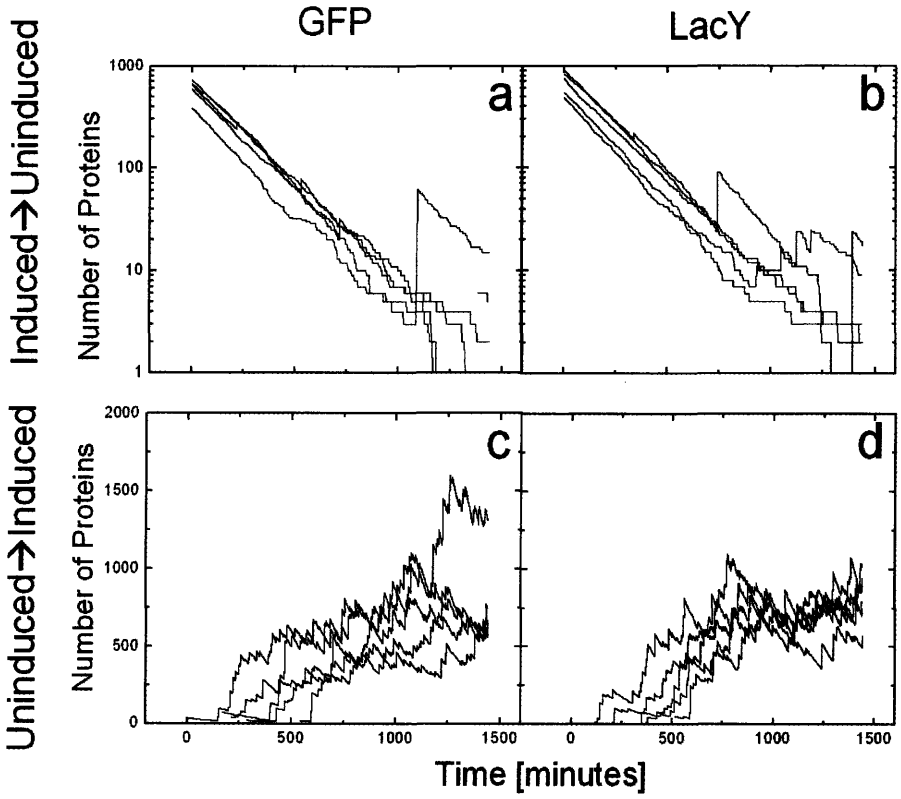


Figure 4.1.13 Examples of **(a)** GFP and **(b)** LacY trajectories from the ON to OFF states as well as **(c)** GFP and **(d)** LacY trajectories from the OFF to ON states. These trajectories serve to illustrate that protein production, which relies on mRNA production, is a highly noisy process. On the other hand, protein degradation, is a low noise process because of the large number of events involved.

Because it is not obvious that the stochastic model will preserve the relevant behaviors of the deterministic model, it is important to examine it in the steady-state regime where the deterministic model was originally compared to experimental data. To verify that the model is functioning properly, growth of cells from both ON and OFF histories is simulated for the equivalent of 20 hours in a range of external concentrations of TMG. As can be seen in Figure 4.1.14 the cells largely remain in their original states through the shaded region, while transitions between the two states occur more frequently past the edges of this region. The hysteretic effect is similar to published measurements of single cells (Ozbudak et al. 2004), indicating that the model is capturing many of the essential properties of the biological network. The greater than expected transition rate from OFF-to-ON in the deterministically bistable region could be due to increased sensitivity to parameter error in this region, or an overestimation of LacY noise.

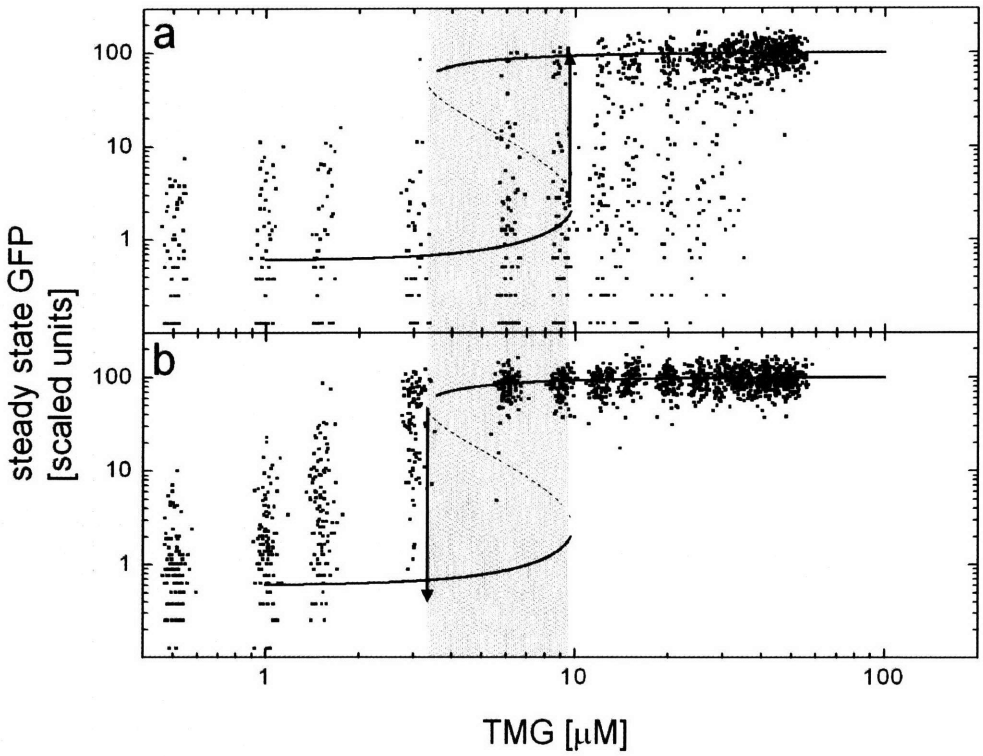


Figure 4.1.14 Hysteretic behavior of the stochastic model similar to that observed in experiments. **(a)** Simulated uninduced cells growing in various concentrations of external TMG for the equivalent of 20 hours. **(b)** Simulated induced cells growing for 20 hours in the same concentrations of TMG as in **(a)**. Each point represents a single cell in a population of 100 cells. Y-values represent GFP concentration, and x-values represent external TMG with a small spread introduced to reflect each cell's RFP concentration. The shaded region indicates the region of bistability as determined from the deterministic rate equations.

In Figure 4.1.15 we compare the model's steady state distributions to experimental data for cells of both an ON and OFF history grown in 6 μM and 9 μM TMG, which are both in the deterministically bistable region, and we find reasonable agreement between these data.

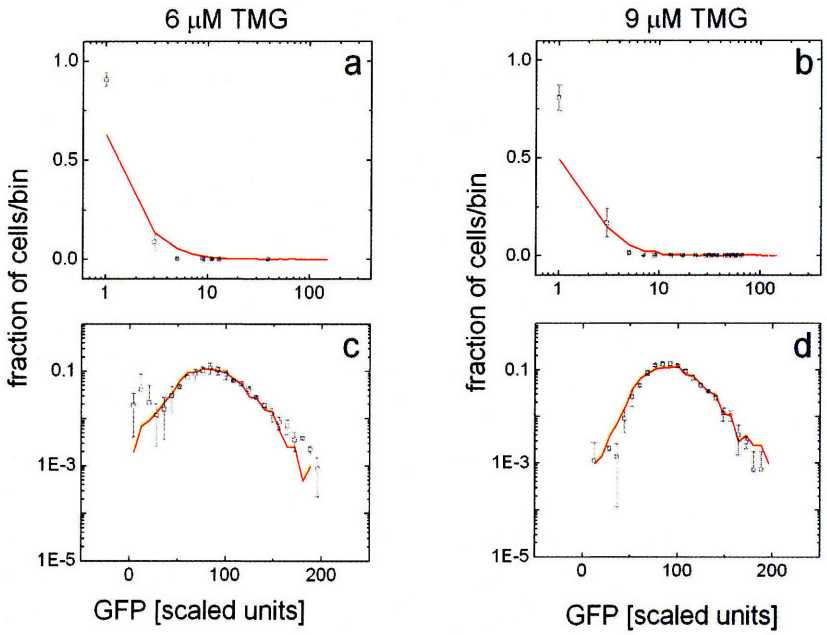


Figure 4.1.15 Comparison between model and experiments in bistable region. The gray boxes are experimental results from initially **(a,b)** OFF and **(c,d)** ON cells grown for 20 hours in **(a,c)** 6 μM TMG and **(b,d)** 9 μM TMG. The red lines are the result of simulating 10^4 cells in conditions equivalent to those used experimentally.

To further test our model in the bistable region, we calculate the OFF-to-ON transition rate when cells are switching stochastically. This can be done by fitting the equation $f_0 e^{-\gamma t}$ to the fraction of cells remaining in the OFF state at time t, $f_{OFF}(t)$. The results for the switching rate $\gamma(TMG)$ from this fitting are shown in Figure 4.1.16 for both experimental data and simulations in the range of 3 μM - 60 μM TMG. The transition rate for both curves is near or equal to zero until approximately 9 μM TMG, which coincides with the upper boundary of the deterministically bistable region. Above 9μM TMG, the model and experimental transition rates rise together as extracellular TMG increases.

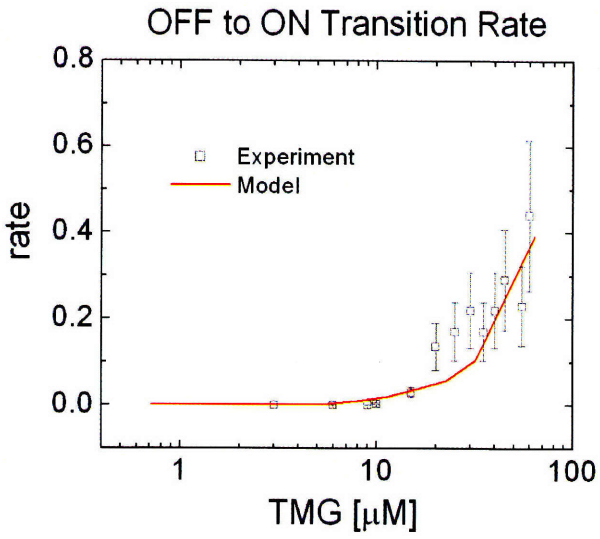


Figure 4.1.16 Comparison between switching rates for the experiments and model. The gray boxes represent the rate at which initially OFF cells are observed to transition to the ON state as a function of extracellular TMG concentration. The red line is the result of simulating 200 cells that were initialized in the OFF state.

The stochastic model reproduces both the ballistic and stochastic switching, whereas the deterministic model does not distinguish between these types of behavior. For the case shown in Figure 4.1.17*a*, the experiments and simulations both show that every cell moves like the average obtained from the deterministic model. In addition to demonstrating average behavior, the stochastic model also correctly predicts the width of the distributions. For the data shown in Figure 4.1.1*b*, the individual cells behave very differently from the average, and the stochastic simulation captures this behavior. In this case the peak at uninduced GFP levels slowly decays as a subpopulation of cells begins to transition to the induced state. In addition to demonstrating the general behavior, the model matches the rate of transitions out of the uninduced peak and predicts the shape of the population distribution over a wide range of TMG and time.

It is noteworthy that ballistic switching does not always occur when the initial state becomes unstable. For instance, even though the OFF state is no longer stable in media with 50 μM TMG, the timing of an OFF-to-ON switching event depends on a cell's rare production of LacY mRNA and subsequent protein creation. It is this rare, stochastic burst of LacY that ultimately triggers the positive feedback loop and drives the OFF-to-ON transition. Conversely, the ON-to-OFF transition is ballistic because it requires the dilution of intracellular TMG and GFP, both low-noise events involving high molecule numbers.

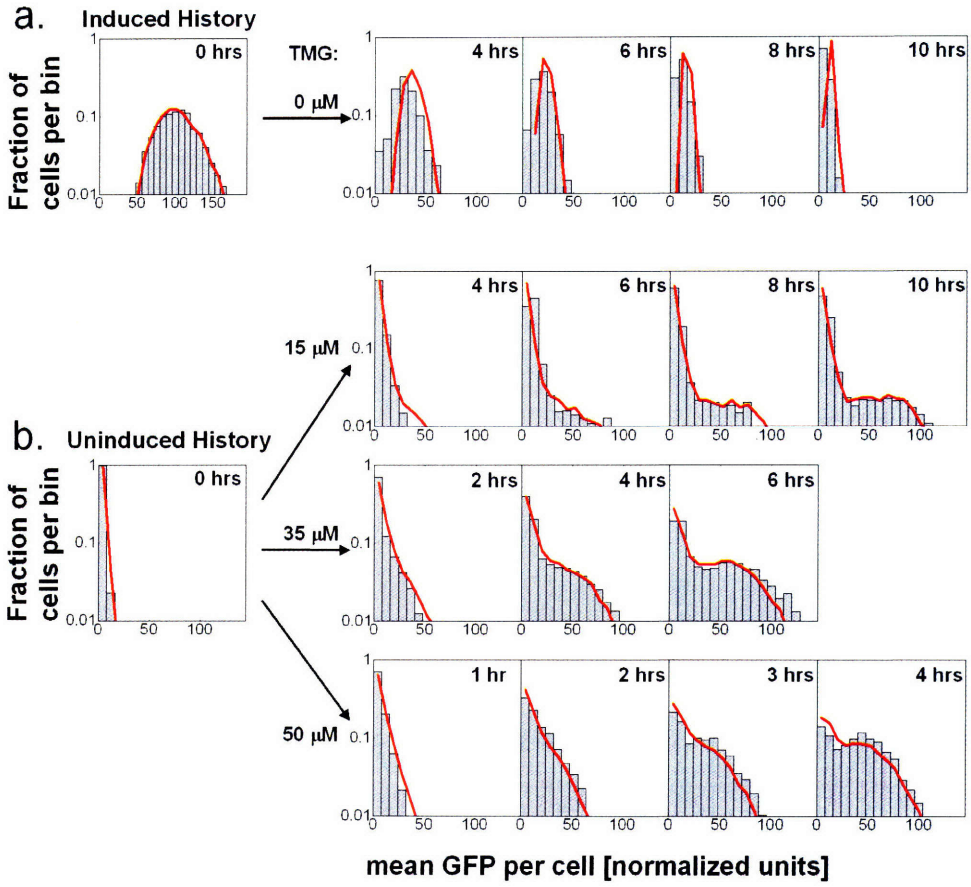


Figure 4.1.17 Comparison between model (red lines) and experimental data (gray boxes) when cells are transitioning from the **(a)** ON to OFF and **(b)** OFF to ON states. Model represents the results of 10,000 independently simulated cell trajectories.

4.1.11 Parameter Robustness

In order to quantify the sensitivity of the model's output to variations in the deterministic and noise parameters, we construct a cost function that compares two population distributions. The cost function should return small values for nearly identical distributions and large values for very different distributions. We set $H(G, T, t)$ to be the fraction of initially uninduced cells in the bin centered at G units of GFP fluorescence after t hours of growth in T μM TMG. This is similar to the curves shown in Figure 4.1.17. We define a cost function, X, similar to a χ^2 error, and evaluate it on the logarithmic-normalized histograms:

$$X = \Sigma_T \Sigma_t \Sigma_G \log^2(H_{experiment}(G,T,t)/ H_{model}(G,T,t)) \quad \textbf{[26]}$$

The sums are confined to the histograms shown in Figure 4.1.17, and any terms with $H = 0$ were ignored in the sum. To estimate the sensitivity of the model to parameter error, each parameter was individually varied, and the cost between the model predictions and experimental measurements was calculated. We vary $\alpha, \beta, \rho, \lambda, N_{GFP}$ and b_{GFP} by the calculated errors shown in Table 4, while N_{LacY}, b_{LacY}, N_{LacI} and b_{LacI} are each varied by a factor of two because these parameter values were not measured directly. The results of these robustness calculations, $X(Value+Error)$ and $X(Value-Error)$, are shown in Table S1. When all parameters are given by the values indicated in Table 2 the cost is found to be $X(Value) = 67.02$, which is close to the cost between two histograms generated from replicates of the same experiment: $X_{Experimental} = 45.50$.

Error on the parameters α, λ, and $\tau_{1/2}$ has the greatest impact on the model's predictions when compared to experiments. Even though the highest value of X is more than twice as large as the lowest, the qualitative predictions remain similar throughout the range of parameters. Variations in each parameter slightly change the shape of the simulated histograms. For example, increased (decreased) values of λ created a higher (lower) rate of OFF to ON transitions (Figure 4.1.18). While the shapes of each peak in the bimodal histogram of transitioning cells remains similar, the relative magnitudes of the peaks are changed. For changes in α, however, the peak of OFF cells decays with a rate similar to experiments, but the simulated transitioning cells either produce GFP too quickly or too slowly for the histogram to closely match experiments (Figure 4.1.19). The noise parameters N and b seem to affect prediction accuracy the least. Decreasing b_{LacY}

even causes better agreement than the value estimated through GFP noise, suggesting that noise in the LacY levels might have been over-estimated.

Parameter	*X(Value + Error)*	*X(Value – Error)*
α	141.96	120.15
β	82.95	81.22
ρ	85.78	94.37
λ	177.18	120.88
N_{LacY}	73.62	91.61
N_{GFP}	82.42	94.70
N_{LacI}	75.12	89.13
b_{LacY}	83.29	63.02
b_{GFP}	90.80	71.45
b_{LacI}	70.27	74.07
$\tau_{1/2}$	95.61	119.71

Table 4: Results of parameter robustness simulations where fractional error (X) is reported for changes in each individual parameter. For comparison: $X_{Experimental} = 45.50$ and $X(Value) = 67.02$

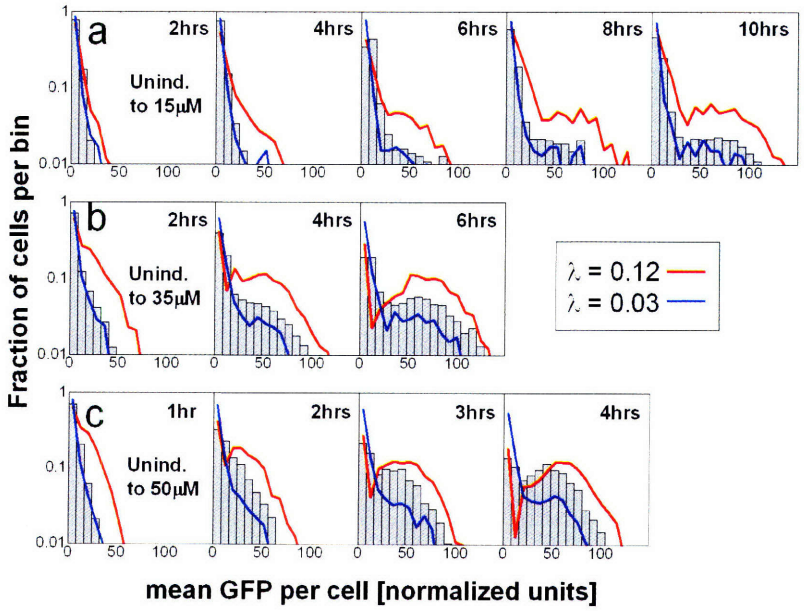


Figure 4.1.18 Model predictions for changes in λ both above and below measured value. The shapes of the curves are maintained while the escape rates are increased or decreased.

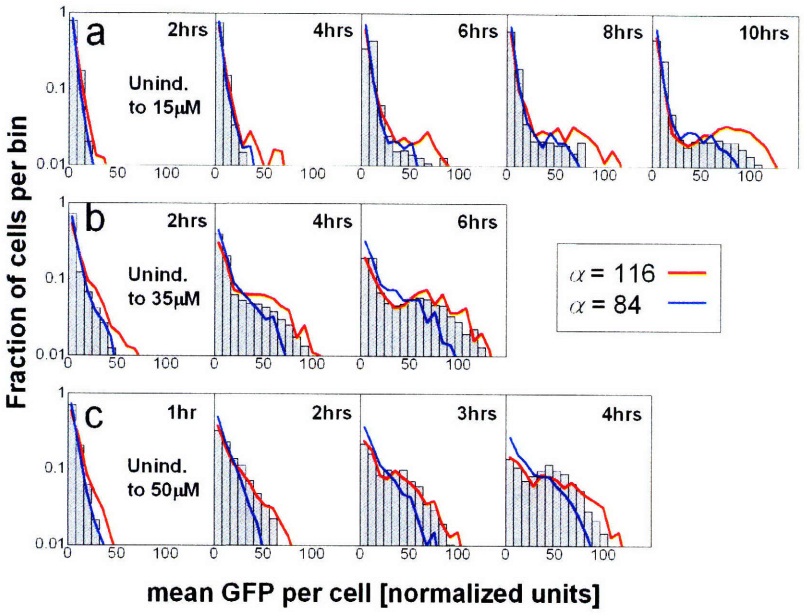


Figure 4.1.19 Model predictions for changes in α both above and below measured value. The escape rates are similar while the rate of GFP production is changed.

4.1.12 Discussion

We introduced a method that should have general applicability for the prediction of stochastic cellular dynamics. The first step includes characterization of a deterministic model that matches known steady-state behavior. Rate constants in this model can be estimated from published values, by fitting to steady state measurements, or through direct biochemical assays. Next, the magnitudes of noise sources are extracted from distributions of fluorescent counts and correlations between different expression reporters measured in steady state. The sources of intrinsic noise are then characterized by the discrete molecule numbers (N) and mRNA burst sizes (b). By combining these reaction rates and noise sources, a stochastic model is produced containing three important stochastic factors: mRNA production, protein degradation, and global noise.

The model is in good agreement with experimental data and can predict the type of response (ballistic versus stochastic), the escape rates from a state, and the distribution of reporter fluorescence values without any fit parameters. Furthermore, the predicted distributions proved to be robust against parameter variation in both magnitude and general behavior. The precision of our model's predictions could be improved by careful measurement of individual model parameters; however, we have shown that previously obtained parameter estimates are sufficient to provide interesting quantitative information about network behavior not available from deterministic equations alone.

4.1.13 Experimental Materials and Methods

Growth conditions and Media

Cells were grown at 37°C in M9 minimal media with succinate as the main carbon source. Overnight cultures were prepared in the absence of TMG or in the presence of 100 μM TMG to yield populations of cells uninduced or fully induced for *lac* expression, respectively. Cells from the overnight cultures were diluted into fresh media containing intermediate TMG levels and maintained at low density ($0.001<OD_{600}<0.005$) to prevent TMG depletion throughout the experiment. At specified time points, a portion of each culture was removed and prepared for imaging; fresh, pre-warmed media with the appropriate TMG level was added to dilute the remaining culture so that the cell density of each imaged sample was the same ($OD_{600} = 0.005$) for all time points. Samples were concentrated and prepared for imaging by filtration, centrifugation, resuspension in 1.25 μl of the appropriate growth media, and placement on a microscope slide.

Bacterial Strains

Dynamic population distributions of *lac* promoter activity were gathered using *E. coli* MUK21 (Ozbudak et al. 2004), in which the *gfp* gene is placed under the control of a wild-type *lac* promoter and chromosomally inserted. Published steady-state distributions from the ERT113 strain (Ozbudak et al. 2004) are further analyzed here for both noise measurements and partitioning analysis. Strain ERT113 was constructed by transforming MUK21 cells with a plasmid containing the red-fluorescent protein HcRed under the control of the *gat* promoter.

Fluorescence Microscopy

Measurements of GFP fluorescence in dynamic (pre-steady-state) cell populations were obtained using a Nikon TE300 microscope equipped with automatic stage and focus, and a Micromax 1024B CCD camera (Roper Scientific), all controlled by MetaMorph software (Universal Imaging). Steady-state measurements were gathered as previously described (Ozbudak et al. 2004). Fluorescence values for cells are corrected by subtracting background fluorescence measured in a region of the field of view devoid of cells. Cell boundaries were determined by auto-thresholding phase contrast images, and GFP intensity was averaged over this area. Mean fluorescence levels are assumed to be representative of reporter concentration and are calculated by dividing the total intensity

of the cell by the area in pixels of the cell in the phase contrast image. These numbers are then normalized so the induced population average of the mean fluorescence is 100 for both GFP and RFP.

4.2 Correlations in Switching Dynamics[2]

4.2.1 Introduction

Inheritance is more than the faithful copying and partitioning of genomic information. When cells divide, the mother passes numerous other cellular components to the newly born daughter, including nucleosomes, transcription factors, mitochondria, and substantial fractions of her proteome and transcriptome. In this way, an entire pattern of gene expression can be passed from mother to daughter, a phenomenon known as epigenetic or non-Mendelian inheritance. Classic examples abound in the literature and include the sex-ratio disorder in drosophila (Colaianne et al. 1970), the yellow-tip phenotype in melons (Ray et al. 1996), the telomere position effect in yeast (Gottschling et al. 1990) and mouse (Pedram et al. 2006), and prions such as Psi+ in yeast (Patino et al. 1996).

The timescale over which epigenetic phenotypes may persist spans many orders of magnitude and depends strongly on the physical mechanism employed by the cell (Rando et al. 2007). In general, however, epigenetic phenotypes are significantly less stable than chromosomally inherited ones (Rando et al. 2007), and can change reversibly in single cells (Gottschling et al. 1990; Halme et al. 2004; Ozbudak et al. 2004), during development (Muller 1930; Lyon 1961), or even in mature organisms (Tosh et al. 2002).

Beginning with landmark studies on the lac operon in the 1950s, positive transcriptional feedback loops have emerged as a means to store cellular memory (Novick et al. 1957; Maloney et al. 1973; Rao et al. 2002). Such epigenetic inheritance systems are frequently described as 'bistable', meaning that transcriptional activity of genes in the network tends to become fixed in single cells around one of two stable levels (ON and OFF), each of which is able to stably persist for many generations (Ozbudak et al. 2004; Acar et al. 2005; Suel et al. 2006). Stochastic fluctuations in the synthesis or decay of the proteins involved (Thattai et al. 2001; Elowitz et al. 2002; Ozbudak et al. 2002; Blake et al. 2003; Paulsson 2004; Raser et al. 2004; Becskei et al. 2005; Golding et al. 2005; Kaern et al. 2005; Pedraza et al. 2005; Raser et al. 2005; Rosenfeld et al. 2005; Bar-Even et al. 2006; Cai et al. 2006; Newman et al. 2006;

[2] The results in this chapter can be found in Kaufmann, B. B., Q. Yang, J. T. Mettetal and A. van Oudenaarden (2007). "Heritable stochastic switching revealed by single-cell genealogy." PLoS Biol **5**(9): e239.

Volfson et al. 2006; Kaufmann et al. 2007), or changes in external cues (e.g., a changing environment), are responsible for causing transitions between the two states (Novick et al. 1957; Ozbudak et al. 2004; Acar et al. 2005; Suel et al. 2006).

This flexible strategy, present in both prokaryotes and eukaryotes, allows genetically identical cells to diversify their population, possibly allowing them to exploit new environmental niches or survive in a fluctuating external environment. Feedback-based cellular memories show an exceptional range of stability; depending on the strength of the loops, cells may display memory of a previous expression state as short as a single generation to many thousands of generations (Acar et al. 2005). However, quantitative measurements of phenotype stability, switching, and heritability are rare, both because detailed genealogical relationships are challenging to produce in single cells (Axelrod et al. 1993) and because reporters indicating degree of inheritance are not always available.

To measure how a dynamic gene expression state is inherited, we focused on an engineered version of the galactose utilization (GAL) pathway in the yeast *Saccharomyces cerevisiae*. We disrupted the pathway's major negative feedback loop, and grew cells in conditions where only a single positive feedback loop was operational (see Methods). Under these conditions, cells stochastically transition between two distinct expression states even in the absence of an extracellular trigger. These infrequent switching events therefore likely arise from fluctuations in concentrations of regulatory proteins within the individual cells (Ramsey et al. 2006). We are able to monitor transitions between ON and OFF using a fluorescent reporter (see Methods). Because of these quantifiable stochastic transitions, our network is an ideal model system well suited to study the heritability of an entire dynamic gene expression state.

In this work, we find that not only is the epigenetic phenotype itself heritable, but that the stability of this phenotype is likewise a heritable quantity. In other words, when cells divide, the nascent daughter cell assumes both the expression state of the mother cell as well as her tendency to switch epigenetic states at a similar time in the future. This is surprising, especially considering that individual cells viewed outside their genealogical context appear to switch at random, consistent with simply a constant rate stochastic process. We resolve this apparent contradiction using a simple stochastic model.

4.2.2 Network Structure

Elaborate regulatory machinery has evolved to help cells focus on the carbon sources that maximize their growth rate in a particular environment. The much studied galactose utilization network (GAL) is a model for such decision/metabolic pathways. The network, comprising roughly a dozen genes, contains two positive and one negative transcriptionally-mediated feedback loops nested one within the next. The positive loops make possible a two-state 'all-or-nothing' arrangement, while the negative loop is thought to stabilize cells in one of these two states, called ON and OFF.

The ON state's defining characteristic is high activity of Gal4p, a transcriptional activator constitutively bound to the promoter of many GAL genes (Keegan et al. 1986). Gal4p activity occurs only in the absence of a dominant repressor Gal80p, which becomes sequestered to the cytoplasm in the presence of active Gal3p(Peng et al. 2002). *GAL3*, in turn, is positively regulated by the level of active Gal4p, closing the first positive feedback. Thus Gal3 presents something of a chicken-or-the-egg situation: high expression of Gal3p leads to the activation of the GAL network, and activation of the GAL network leads to high expression of Gal3p.

A membrane protein Gal2p, whose expression is also regulated by Gal4p, forms a second positive feedback loop by importing galactose into the cell and consequently activating Gal3p. Finally, *GAL80* negatively regulates its own production, again by suppressing the activity of Gal4p. For cells in the OFF state, the situation is reversed with low Gal4p activity, presence of repressor Gal80p at the GAL promoters, and a depletion of (active) Gal3p and Gal2p in the cytoplasm.

In wild-type cells, transitions between the OFF and ON states can be forced by changing levels of inducers (e.g., galactose) or repressors (e.g., glucose) in the surrounding environment. At intermediate levels of inducer, the OFF or ON status of the cell depends on the history of the media in which their ancestors grew, indicating significant hysteresis. This hysteretic behavior buffers against switching too rapidly between states, perhaps to avoid the metabolic cost incurred.

4.2.3 Experimental Setup and Network Behavior

We first set out to quantify the infrequent switching events that occur at random times using fluorescence microscopy. All experiments began with a single cell confined between a cover slip and a thick agar pad. Over a period of about 920 minutes (>15 hours) each cell grows and divides to form a small colony of between 50-100 cells. Throughout the measurement period, these cells diverge in behavior with some increasing in fluorescence and others decreasing. We repeated this process with more than 100 progenitor cells, so in sum our data represent many thousand single-cell trajectories.

We present two examples of the experimental procedure in Figure 4.2.1. In the top panels, an initially bright cell develops into a small colony with distinct subpopulations. The dim cells in the lower subpopulation continue to diminish in fluorescence with each successive cell division as the remaining molecules of GFP dilute. In the bottom panels, an initially faint cell likewise gives rise to a variegated colony with cells both dim and bright. Together, these two processes generate a broad bimodal steady-state distribution.

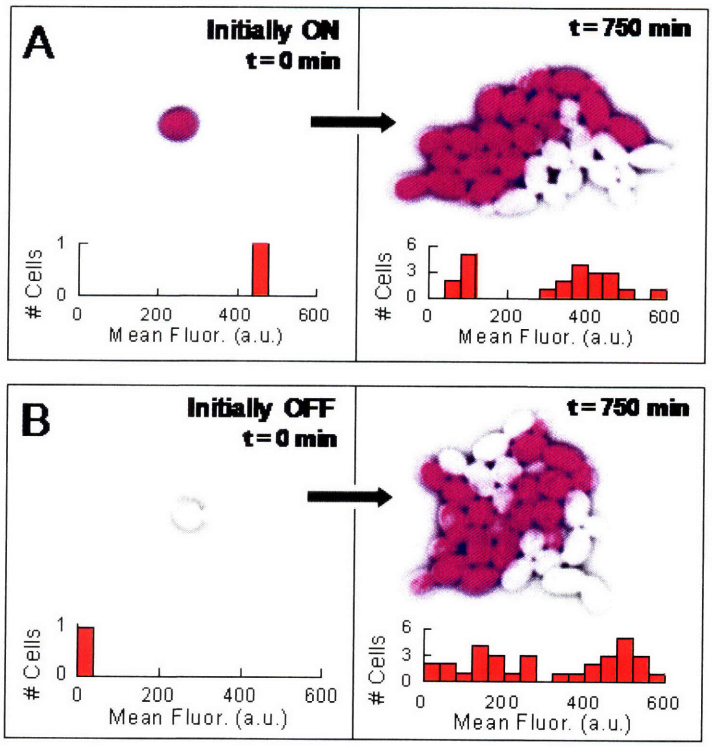


Figure 4.2.1 Cells switch between expressing and non-expressing states. Images are phase contrast micrographs (black and white) overlaid with background-subtracted fluorescent signal (purple). **(A)** Over 750 minutes, or between 4 and 5 generations, an initially ON cell of strain MA0188 develops into a small variegated colony with subpopulations of ON and OFF cells. **(B)** An initially OFF cell likewise grows into a mixed colony with both ON and OFF cells. The sharp interface between ON and OFF cells in both **(A,B)** indicates that cell-cell communication does not play a major role in defining cell expression state.

4.2.4 Individuals Switch in a Poisson Manner

Narrowing our focus to initially OFF progenitor cells, we allowed each to grow, divide, and give birth to other initially OFF cells. We then recorded instances when cells switched into the ON state (Figure 4.2.2). Because cellular auto-fluorescence is uniformly small throughout the population of OFF cells, these fluorescing events were generally distinguished unambiguously from background. Using these data we generated, for each colony, a family tree where the detailed genealogical relationships and gene-expression histories of corresponding family members are shown (Figure 4.2.2B).

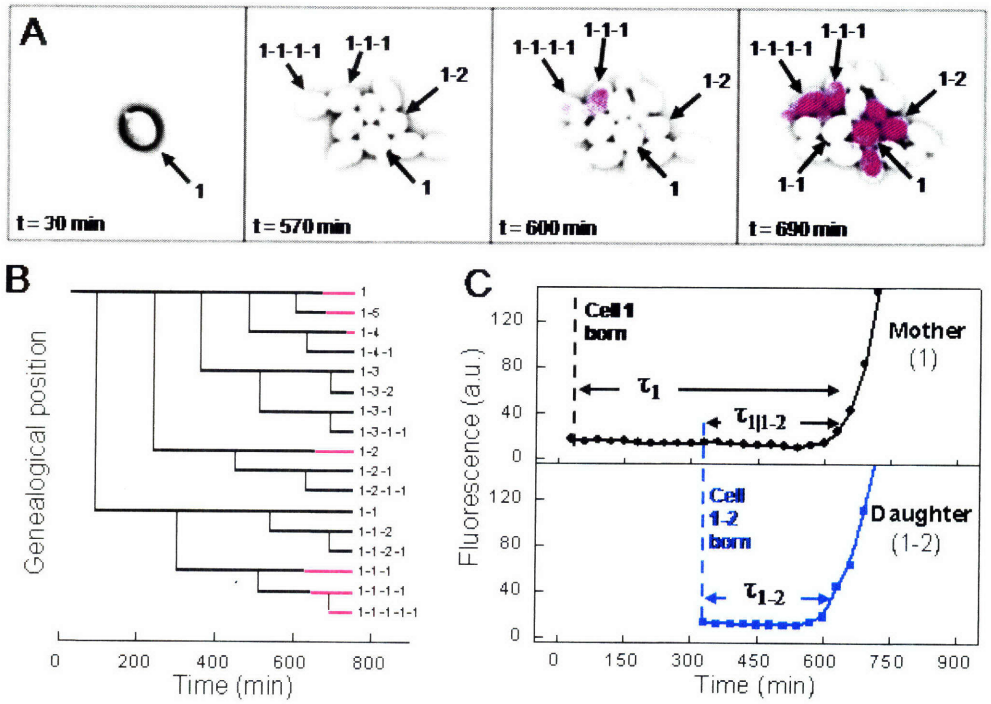


Figure 4.2.2 A genealogical switching history. We designate the first cell in each movie cell 1 and sequential daughters of that cell 1-1, 1-2, 1-3. These daughter cells bud in turn, giving rise to cells 1-1-1, 1-1-2, 1-2-1, etc. **(A)** As in Figure 4.2.1, an initially OFF cell grows into a variegated micro colony. Beginning at 600 minutes, or 4 generations, several cells fluoresce almost simultaneously. This includes the mother-daughter pairs (1,1-2) and (1-1-1,1-1-1-1). Conspicuously, cell 1-1 does not switch for the duration of our observation, even though its mother, daughter, and closest sibling all do. **(B)** The family tree for colony in (A). Black lines indicate cells in the OFF state, while pink lines represent cells after they have switched to the ON state. **(C)** Fluorescent time courses for mother cell 1 and her daughter 1-2, showing each as they switch into the ON state. The marginal switch times τ_1 and $\tau_{1\text{-}2}$, run from cell birth until the beginning of the increase in fluorescence and do not depend on any other cells. The period labeled $\tau_{1|1\text{-}2}$ runs between the birth of cell 1-2 and the fluorescence of cell 1 and is an example of a conditional switch time.

Because cells are continuously born throughout the experiment, we aligned them *in silico* so that their birth times were identical. In this context, it is natural to define the marginal switch time, τ_X, a parameter that describes the interval between the birth of a cell X and the moment it eventually becomes fluorescent (Figure 4.2.2). We normalized each measurement according to its expected likelihood of being observed to account for any biases caused by the cells' exponentially dividing throughout our measurement period. The resulting data fit well to an exponential curve with an effective transition rate of 0.12 switches per generation (Figure 4.2.3A, cyan line). The slight discrepancy between data and exponential fit is likely the result of some cells growing out of the focal plane. The reverse switching distribution, comprising ON cells switching into the OFF state, could not be obtained in this simple way due to the long life of the fluorescent proteins.

4.2.5 Heritability of Switching from Mother to Daughter

This exponentially distributed switching pattern applies to cells chosen at random without regard to genealogy. However, measuring cells instead on the basis of their family history paints a very different picture. To demonstrate this, we selected all daughter cells with marginal switch times below some value *T*. Out of this subset, we then asked what percent of their mothers had also switched at or before that time. The results, summarized in Figure 4.2.3B (open circles), show that when a daughter switches shortly after cell division, its mother cell is overwhelmingly likely to do the same. For example, of the daughters who switch within 400 minutes of cell division (about two generations), their mothers have approximately a 50 percent chance of switching in that same period. This represents a two-fold increase in the switching rate for a typical unrelated cell. As *T* grows to encompass an ever larger fraction of all daughter cells, the corresponding percent of switching mother cells asymptotically approaches the marginal switch distribution of Figure 4.2.3A (reproduced in black), which represents the limit of no genealogical information. As in the marginal switch case above, we are careful to weigh each of these mother-daughter pairs according to how likely we were to experimentally observe them.

To measure the underlying rates governing this process, we examined the possible switching events diagrammed in Figure 4.2.3A. In this simplified view, we assume cell pairs can either switch together into the ON state together at a rate $c(t)$, or independently of one another at a rate $r - c(t)$. In this way the total switch rate for any given cell sums to *r* at all times, as required by the marginal switch distribution. We assume that the correlations decay with a rate $c(t) \equiv r \cdot e^{-\frac{\max(t-20,0)}{Tc}}$, which is reminiscent of an Ornstein-Uhlenbeck process (Rosenfeld et al. 2005; Suel et al. 2006) (see Figure 4.2.3A). The fixed delay of 20 minutes is included to account for slow chromophore (YFP) maturation as observed in our data (daughters that switch within the first 20 minutes after cell division have mothers that always switch). This model includes two free parameters: *r*, the overall switch rate, and *Tc*, the characteristic time for the correlation to decay. A global least-squares fit to both curves (Figure 4.2.3B red and blue curves) simultaneously yields ($r = [7.0 \pm 0.5] \cdot 10^{-4}$ $\text{min}^{-1} = 0.12 \pm 0.01$ gen^{-1}) and ($Tc = 197 \pm 54$ min). This decorrelation rate is quite similar

to the average cell doubling time of 177 min, and similar connections between doubling time and decorrelation have been found in other protein regulatory networks (Rosenfeld et al. 2005).

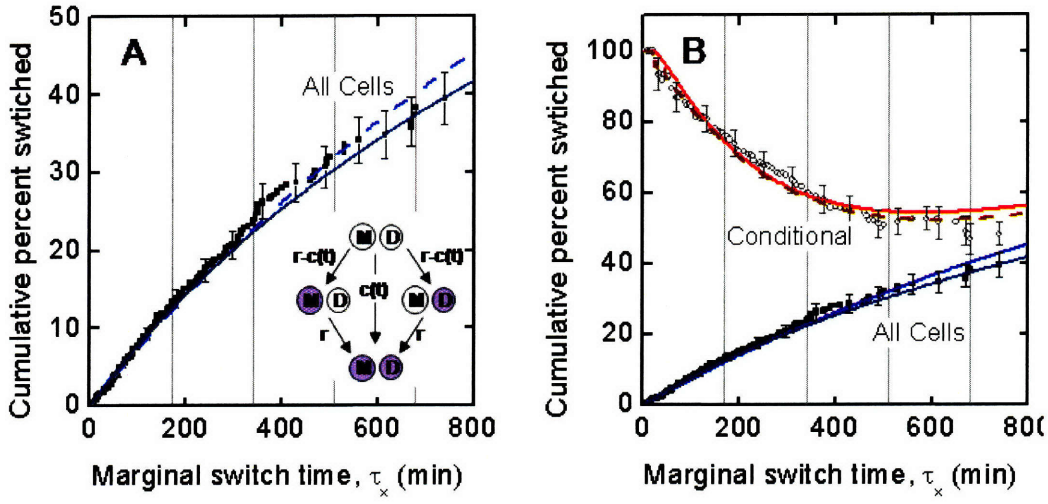


Figure 4.2.3 Single cell fate. **(A)** The cumulative percent of cells that have switched is plotted against their marginal switch time. The black squares represent 251 switching cells, and the blue line is an exponential fit. The cyan dashed line is a result of our stochastic simulation. Error bars are derived from a bootstrap analysis. The fits are consistent with the idea that a constant-rate process may underlie the network. The inset shows ways mother-daughter pairs may switch, either dependently via the center route or independently of one another via the outer routes. **(B)** Gray circles describe the likelihood that a mother cell has switched given that its daughter cell is known to have switched before this time. The solid red line describes a two-parameter least squares fit simultaneously to both curves with parameters described in the inset and main text. The dashed dark red line shows the fit resulting from the stochastic simulation. Black squares and blue lines are reproduced from (A) for comparison.

4.2.6 Correlations of Switch Times

The above analysis suggests that when cell pairs do switch they will do so in synchrony. To demonstrate that this is indeed the case, we turned our focus to the further subset of cell pairs where both cells are observed to switch during the experiment (and therefore ignoring cases where only one cell in a pair switches). More specifically, we concentrated on three cell relationships: mothers with daughters (henceforth M-D), grandmothers with granddaughters (GM-GD), and older siblings with younger siblings (S1-S2). Instead of marginal switching times, which are measured relative to each individual cell's time of birth, we chose instead to compute the switch times of both cells relative to the moment when their two respective branches of the family tree first broke apart. Put another way, this quantifies the amount of time between a switching event and the last moment that these cell lines shared cytoplasm. The purpose of this approach is to allow us to compare cells that were born at very different times on equal footing, ensuring that switching events are measured relative to the same point for both cells. For M-D pairs, the time we use is simply the birth of the daughter, but for GM-GD pairs it is the birth of the intervening daughter, and for S1-S2 pairs, it is the older sibling's birth. Formally we define the conditional switch time, $\tau_{X|Y}$, as the time elapsed between the fluorescing of cell X and the birth of cell Y. When X and Y both refer to the same cell, we recover the marginal switch time (i.e., $\tau_{X|X} = \tau_X$).

Comparing M-D conditional switch times (Figure 4.2.4A), we observe nearly synchronous switching that extends at least 300 minutes and yields a correlation coefficient of ρ_{MD}=0.87 ($p<10^{-45}$). GM-GD and S1-S2 pairs (Figure 4.2.4B and C) show somewhat lower correlation coefficients of ρ_{GMGD} = 0.74 ($p<10^{-9}$) and ρ_{SS} = 0.60 ($p<10^{-7}$) respectively, although the overall coefficient for all data combined remains a robust ρ_{TOT} = 0.8 ($p<10^{-62}$). The strength and duration of these correlations are surprising, and were not found in bacterial (Golding et al. 2005; Suel et al. 2006) and mammalian (Geva-Zatorsky et al. 2006) studies, except in context of morphological traits (Wakamoto et al. 2005). Like the marginal switch data, these scatter plots should be viewed in the context of finite experimental viewing times, giving weights to points that are inversely proportional to the number of experimental opportunities to have seen them.

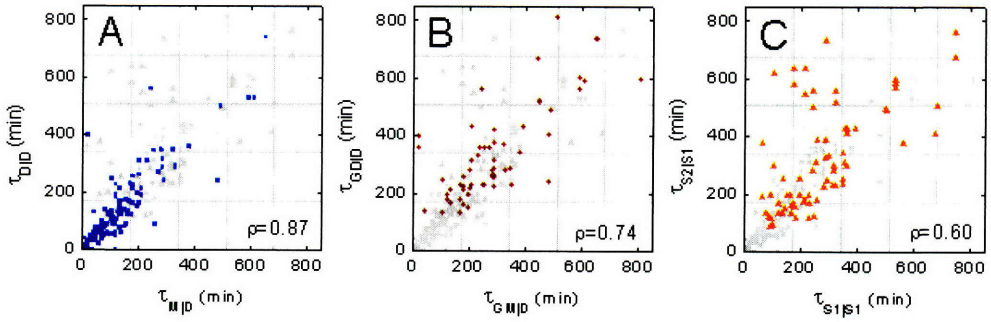


Figure 4.2.4 Cell pair behavior. The conditional switch times for closely related cells are compared. **(A)** The daughter switch time is compared to the mother switch times for 141 cell pairs. For times extending past 350 minutes (about two cell divisions), a strong correlation in times is observed. The other cell pair relationships, shown again in (b,c), are shadowed in grey. **(B and C)** The more distant relationships of grandmother-granddaughter (N=55) continue to show significant correlation, while the sibling-sibling relationships (N=74) shows somewhat less. The notable asymmetry of the sibling-sibling distribution reflects the tendency for older siblings to sometimes switch before the younger sibling is even born.

4.2.7 Persistence of Correlation in Switching Times

One dynamic measure for the randomness associated with the distribution is the average square difference of switch times for pairs of cells with comparable mean switch times (Figure 4.2.5D, blue curve). This curve rises rapidly at first, but at longer times it flattens out. This flattening is likely due at least in part to the limited duration of our experiments (on average 920 minutes), which constrains the scatter distribution to reside in the box shown in Figure 4.2.4A-C.

To generate the mean squared deviation in Figure 4.2.5D, we first combined all observed MD, GMGD, and SS switch pairs (Figure 4.2.5A-C) into a single data set. This was done to increase statistical power. These data sets consisted of two columns, t_1 and t_2, representing the older and younger cell switch times for 274 cell pairs. From this data set we used the bootstrap technique (draw with replacement) to generate 1000 alternate samples. Then, using each of these generated samples in turn, we transformed the data in the following way:

$$S = \frac{(t_1 + t_2)}{2}$$

$$D = | t_1 - t_2 |^2$$

We next binned the transformed data according to S, using bins of sizes 25, 50, 100, and 150 min, with centers staggered by 75 min. At each center point, we calculated for all bootstrap samples and all bin sizes the average of the quantity D. Thus for each bin center we had four thousand estimates for the mean squared deviation (4 bin sizes times 1000 bootstraps). The mean and standard deviation of these four thousand estimates gives the mean and error shown in Figure 4.2.5D.

To understand what this means, it is helpful to compare our results to those obtained using a stochastic Poisson model (Gillespie 1977) where closely related cells are assumed to switch independently of one another and with constant probability in time. If cells in this Poisson model were given an infinite time to switch and only mother-daughter pairs were considered, the mean squared deviation would simply scale linearly with the mean. To compare directly with our data, however, we run the simulation for the same duration as our experiment and include all cell pair relationships, giving the more complicated curve shown in Figure 4.2.5D (red curve).

The ratio of the data's mean square variation to that of the Poisson simulation (Figure 4.2.5E, green curve) is a measure for how correlated cells remain after a given period of time has passed. Points below the dashed line at unity would represent correlated switching behavior, whereas points above it would signify anti-correlated behavior. For over 600 minutes, the distribution remains distinctly sub-Poissonian. Only for the longest measured times are there indications that the cells become independent of their history, and even this is with large uncertainty. Put another way, pairs of cells often remain on approximately the same trajectory for several cell divisions, even though cell growth has diluted many of the relevant proteins to a fraction of their original level.

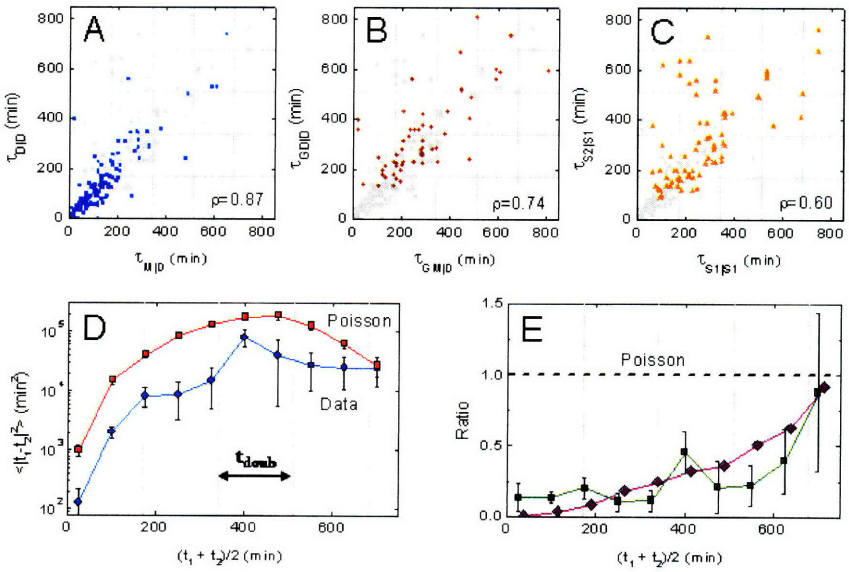


Figure 4.2.5 Cell pair behavior. The conditional switch times for closely related cells are compared. **(A)** The daughter switch time is compared to the mother switch times for 141 cell pairs. For times extending past 350 minutes (about two cell divisions), a strong correlation in times is observed. The other cell pair relationships, shown again in (B,C), are shadowed in grey. **(B and C)** The more distant relationships of grandmother-granddaughter (N=55) continue to show significant correlation, while the sibling-sibling relationships (N=74) shows somewhat less. The notable asymmetry of the sibling-sibling distribution reflects the tendency for older siblings to sometimes switch before the younger sibling is even born. **(D)** In blue, the mean squared difference of the switch times from the combined relationships in (A-C), binned according to their average switch time. In red, a computer-generated Poisson simulation sets a bound for switching correlation in the limit of correlation tends to zero. The mean cell doubling time is labeled t_{doub}. **(E)** Dark green squares show the ratio of the two curves in (D), demonstrating the persistence of a correlation for at least hundreds of minutes after cell division. In purple, the predicted fit from our stochastic simulation after fitting to the curves in Figure 4.2.3.

4.2.8 Stochastic Model

To examine our results at a microscopic level, we constructed a simple model that allows us to probe how the rich correlated switching dynamics arise from a simple regulatory network. Specifically, we asked whether the stochastic fluctuations of a single regulatory protein in our system could simultaneously explain the observed Poisson switching behavior expected for randomly selected individuals and subsequent long-timescale correlations. One key protein, Gal80p, functions to regulate the expression of all other genes in the network. When present in the nucleus, Gal80p binds in a highly cooperative manner to the transcription factor Gal4p and represses the expression of Gal2p, Gal3p, and YFP (Volfson *et al*, for example, assume a Hill number of 8 between Gal4p and transcription at the *GAL1* promoter (Volfson et al. 2006)).

Such high levels of cooperativity frequently give rise to steep transfer functions, which can result in switch-like behavior. This means that even a small decrease in the concentration of Gal80p can cause the transcription rate of downstream genes to increase dramatically from a very small basal rate to a large maximal rate. Once the downstream protein, Gal3p, begins to be produced, it will lead to sequestering of Gal80p to the cytoplasm, completing the feedback loop and causing the cell to completely switch from the OFF to the ON state.

We constructed a simple model that captures the essential properties of this process. In our cells, Gal80p is present in very low numbers, and we therefore account for the effects of stochastic production and degradation for this protein. Protein bursting invariably increases noise levels by amplifying rare events such as changes in promoter activation or mRNA creation and destruction (Thattai et al. 2001; Paulsson 2004) (Blake et al. 2006). We assumed that the burst-size distribution was exponential in shape with a mean consistent with the results of Bar-Even *et al.*, who found an average of 1200 proteins per burst (Bar-Even et al. 2006). We further assumed that the decay rate of the protein is dominated by dilution and therefore set by the division time of the cell. Finally, we included in our model a nonzero chromophore maturation time of 20 minutes, as observed in our data. To account for cooperativity between Gal80p and Gal4p, we assume that when Gal80p levels drop below a threshold value a cell rapidly activates gene expression and enters the ON fluorescent state.

In total, the model has only three parameters: 1) mean number of Gal80p molecules present per cell, 2) the switching threshold, and 3) the Gal80p burst size

estimated from literature. We estimated the first two of these parameters by fitting the model to the marginal and conditional switching distributions shown in Figure 4.2.3B. Once the theoretical switching rates were fit to the experimental data, we asked if the model explained the highly correlated switching times observed between related cells. Without any additional fitting parameters, we predicted the mother-daughter, $\tau_{M|D}$, and daughter-daughter, $\tau_{D|D}$, conditional switching times (Figure 4.2.7F, brown squares) as well as their mean squared deviation $\langle (t_1 - t_2)^2 \rangle$ (Figure 4.2.4E, purple diamonds). These predictions matched remarkably well with the experimental data (Figure 4.2.4E, green boxes; Figure 4.2.7F, grey circles). The model therefore predicts that related cells will remain highly correlated in their switching times even though switching events seem to occur in a Poisson manner. A robustness analysis (Figure 4.2.6) suggested a narrow range of possible values with an optimum centered around (average, threshold) ~ (2400 proteins, 670 proteins).

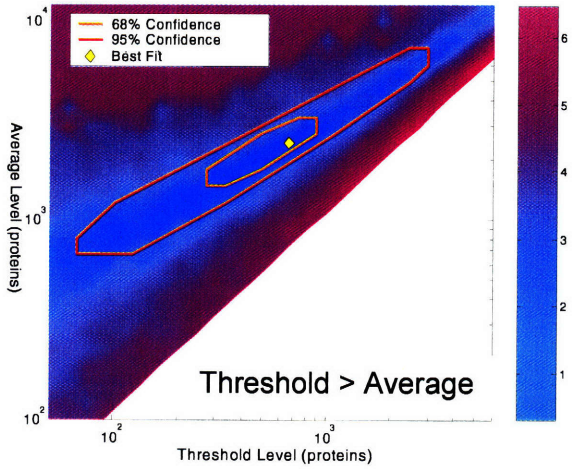


Figure 4.2.6 Monte-Carlo model confidence intervals. Chi-squared is shown in a heat map as a function of the two free parameters in our model, the average level of proteins and the switching threshold. Colorbar at right indicates chi-squared value on a logarithmic axis. Burst size is fixed to 1,200 proteins/mRNA, consistent with Newman *et al* (Newman et al. 2006). Yellow diamond indicates the best fit of all tested values and occurs at (average, threshold) ~ (2400 proteins, 670 proteins). Orange and red contours represent 68% and 95% confidence intervals, respectively. Points where the threshold exceeds the average have been excluded.

Bursting events in protein production are often associated with increases of noise in protein levels (Thattai et al. 2001; Paulsson 2004). A counter-intuitive aspect of our model is that the correlation observed in cell pairs comes as a consequence of

stochastic bursting. As the burst size is ratcheted up from 12 to the experimentally observed value of 1200, for example, keeping average protein level and switch rate constant, correlations begin to emerge in the cell-cell scatter plots (see Figure 4.2.7). The reason for this effect is that the periods between bursting events are dominated by dilution of proteins, a relatively low noise process. As the burst size is increased, the time between bursts must increase commensurately, leading to long periods of correlated behavior between cells. Two cells that start with the same amount of protein will therefore dilute that protein at a similar rate and switch ON (Figure 4.2.7C, black arrows) at similar times. Decorrelations can arise when one of the cells experiences a burst of new protein during this decay period. However, the cell experiencing the burst has a greatly reduced probability of switching ON in a short period of time. In this event, the cell will generally not be observed to switch at all over the duration of the experiment and consequently does not appear as a significantly decorrelated time-point in the τ_{MM}/τ_{MD} scatter plot.

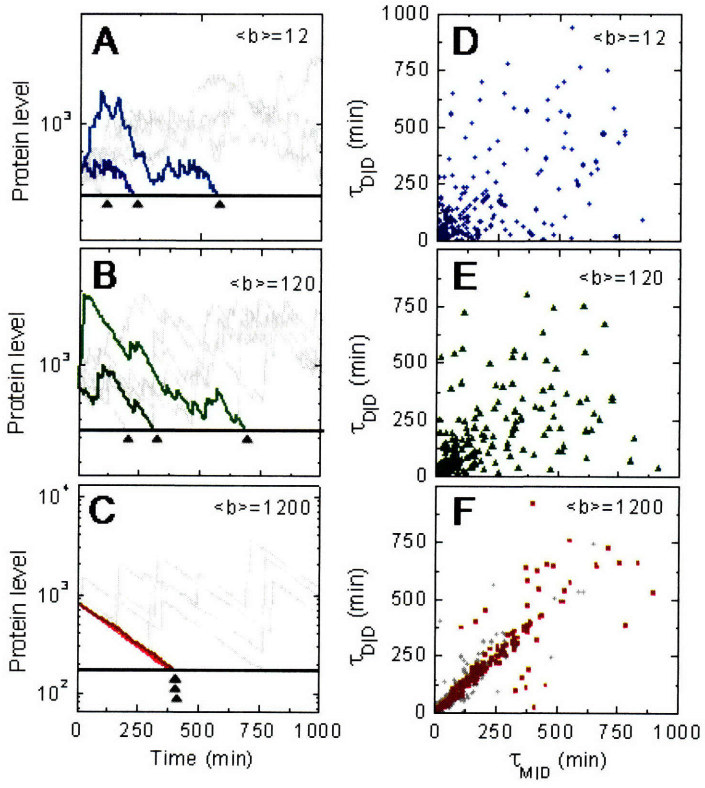


Figure 4.2.7 Burst-induced correlations. Results of our stochastic simulation. **(A-C)** Fluctuations of protein concentration as a function of time are shown (grey lines). Several selected realizations are highlighted for emphasis. Proteins are created in bursts of size <b> = 12, 120, and 1,200 respectively for the three panels, with the average protein level held fixed. Thresholds (solid black lines) are chosen to result in an average switching rate equal to our experimentally measured value. When protein levels drop below a threshold, that cell is considered to have switched. **(D-F)** The resulting mother-daughter scatter plots. As the burst size increases, the pattern becomes markedly more correlated. Gray circles in **(F)** are the experimental data reproduced from Figure 4.2.4A.

4.2.9 Discussion

In recent years, cells within isogenic populations have become increasingly scrutinized as individuals, each with its own original behaviors and gene expression patterns. What make single cells distinctive, however, are not only the stochastic chemical reactions taking place within them, but also their unique family histories. Here we have shown that a cell's decision to dramatically change expression states can hinge directly on this familial background. In doing so, we have separated what on its face appears to be an exponentially distributed random process into stochastic and genealogically-determined subcomponents. In addition, we show that protein or transcriptional bursting, processes which increase total noise across a population, can unexpectedly create correlated behavior between cells, a phenomenon hard to imagine occurring in deterministic descriptions.

In the engineered network we have used in this study, there is no reason to suppose that the correlations we observe provide an evolutionary advantage. However, we can speculate that cells might employ the mechanisms described as a simpler way to coordinate behavior between cells without relying on complex sensory machinery or physical proximity. In addition, cells might exploit these architectures to ensure that when a switching event does occur, that several other cells will do the same, thus achieving strength in numbers. For example, a group of infectious disease cells seeking to confront a host immune system might hypothetically choose to switch together from a slowly growing latent phenotype into an active virulent phenotype in a coordinated but randomly timed attack, thus enhancing their likelihood of sustaining an infection. It will be interesting to see how far such analyses can be taken in the future, with perhaps important features of a cell's future dynamical behavior being confidently predicted knowing only the fate of its immediate family.

4.2.10 Methods

Engineered Destabilization of the GAL Network

We used the well-characterized galactose utilization (GAL) network as our model genetic network. In wild-type cells, transitions between the ON (galactose metabolizing) and OFF (unable to metabolize galactose) states is largely determined levels of inducers (e.g., galactose) or repressors (e.g., glucose) in the surrounding environment. To generate a switching phenotype with large dynamic range, we destabilize this in two ways. First, we remove the negative feedback loop altogether by replacing the endogenous *GAL80* promoter with a weakly-expressing, tetracycline inducible one, P_{TETO2}. Second, we grow the cells in the absence of galactose, which fully eliminates the *GAL2* mediated positive feedback and weakens the *GAL3* feedback. Even in the absence of galactose, Gal3p has constitutive activity and in sufficient quantities can activate the network (Bhat et al. 1992). Considering the lower levels of Gal80p in our construct, this constitutive activity is likely a significant factor. Finally, the state of the network is read with P_{GAL1}-YFP, with fluorescing cells considered ON.

Cells engineered in this way transition between ON and OFF states in a seemingly stochastic fashion. Cells with this genotype exhibit an extremely broad steady-state expression histogram with fluorescence values that span more than two orders of magnitude and has peaks on both the high and low expression limits, suggesting a bistable system with relatively infrequent transitions between the two states.

GROWTH CONDITIONS

Prior to imaging, cells were grown at low optical density overnight in a 30°C shaker in synthetic dropout media with 2% raffinose as the sole carbon source. This neutral sugar is thought to neither actively repress nor induce the GAL genes (Johnston et al. 1994). We grew our cells in the absence of tetracycline so levels of Gal80p were determined by the basal expression level of P_{TETO2}. Approximately 12 hours later cells were harvested while still in exponential phase, spun down, and resuspended in SD media. Next cells were transferred to a chamber consisting of a thick agar pad (composed of the appropriate dropout media and 4% agarose) sandwiched between a cover glass and slide. The high agarose density constrains cells to grow largely in a two-dimensional plane.

MICROSCOPY

Fluorescent and phase contrast images of growing cells were taken at intervals of 20-35 minutes on 10 different days for over 100 initial progenitor cells. Image collection was performed at room temperature (22˚C) using a Nikon TE-2000E inverted microscope with an automated stage (Prior) and a cooled back-thinned CCD camera (Micromax, Roper Scientific). Acquisition was performed with Metamorph (Universal Imaging).

4.3 Transduction and Fidelity of Spatial Signals[3]

4.3.1 Introduction

The low number of molecules involved in biological systems can lead to large stochastic effects and population heterogeneity even within a genetically identical population (Rao et al. 2002; Kaern et al. 2005; Raser et al. 2005). For example, the swimming behavior of *Escherichia coli* cells varies greatly from cell-to-cell (Spudich et al. 1976), and recent studies have begun to link this variability in swimming behavior to concentration fluctuations in regulatory proteins (Levin et al. 1998; Levin 2003; Korobkova et al. 2004). It is an open question whether a similar variability can be observed in eukaryotic chemotactic cells, such as the slime mold *Dictyostelium discoideum*, which has the exquisite ability to sense and respond to shallow gradients of chemoattracants. In these spatially sensitive systems, signaling errors might be introduced in two different ways. First, the concentrations of intracellular signaling components might vary from cell-to-cell, and second, spatial inhomogeneities or asymmetries in the cellular distributions of molecules might influence the ability of cells to sense slight spatial differences in the extracellular environment.

To explore this question, we employ a quantitative approach to systematically study directional sensing in single *Dictyostelium* cells. Recent experiments have demonstrated that an extracellular signal induces spatial localization of several signaling proteins along the plasma membrane (Kimmel et al. 2003; Meili et al. 2003; Manahan et al. 2004; Van Haastert et al. 2004). The localization of these molecules at the membrane allows a cell to polarize and move in the direction of the external signal. We quantitatively monitored the spatial and temporal localization of one of the key signaling proteins fused to green fluorescent protein (GFP), which provides a convenient reporter of directional sensing at the single cell level.

Upon exposure to the same extracellular signal, the GFP localization varies greatly from cell-to-cell whereas a single cell will repeatedly give the same response. We find that the difference in response between the single cell and population is due to asymmetries internal to each cell. Furthermore, this asymmetry varies in magnitude from cell-to-cell causing some cells to accurately follow a moving stimulus while others do not.

[3] The results presented in this chapter can be found in Samadani, A., J. Mettetal and A. van Oudenaarden (2006). "Cellular asymmetry and individuality in directional sensing." Proc Natl Acad Sci U S A **103**(31): 11549-54.

Although we find that most cells are strongly asymmetric in their response, previous experimental and modeling studies have focused mainly on explaining the response of symmetric cells. Because the asymmetry can have a significant impact on the ability of cells to sense external gradients, we develop a model that combines intracellular asymmetries with extracellular signals. When combined with measured parameters, the model is able to accurately predict the observed population response.

4.3.2 Experimental Setup

Starved *Dictyostelium* cells were immobilized and seeded into an observation chamber containing a known concentration of caged cAMP. The response of cells to a short pulse of cAMP was quantified by monitoring the spatial and temporal localization of the cytosolic regulator of adenylyl cyclase (CRAC) fused to green fluorescent protein (GFP) (Kimmel et al. 2003; Meili et al. 2003; Van Haastert et al. 2004). The CRAC-GFP fusion retains the wild type activity. During stimulation, the pleckstrin homology (PH) domain of CRAC binds to the phospholipid phosphatidylinositol 3,4,5-triphosphate ($PI(3,4,5)P_3$) causing CRAC-GFP to translocate to the leading edge of the cell. The GFP fluorescence along the cell membrane therefore provides a reporter of directional sensing at the single cell level (Insall et al. 1994; Lilly et al. 1995; Parent et al. 1998; Meili et al. 1999; Funamoto et al. 2001). A spatio-temporal cAMP gradient was formed by uncaging a known concentration of cAMP using a circularly shaped UV beam (Figure 4.3.1). A major advantage of using a caged compound is the ability to reproduce exactly the same spatio-temporal gradient repeatedly. This allows us to measure the variability of the response of a single cell to multiple identical cAMP pulses.

4.3.3 Quantification of the cAMP pulse

We have quantified the spatial and temporal variation of the cAMP concentration by numerically solving the two dimensional diffusion equation. The diffusion in the observation chamber which is a semi-infinite reservoir is given by:

$$C(x,t) = \int_0^{\min(t_{pulse},t)} d\tau \tilde{C}(x,t-\tau), \qquad [1]$$

$$\tilde{C}(x,t) = \frac{C_0}{4\pi(Dt+\sigma_0)} \exp(-r^2/4(Dt+\sigma_0)), \qquad [2]$$

where r is the distance from the illumination region, D the diffusion coefficient, σ_0 the initial Gaussian width of the UV beam, C_0 the concentration of photoactivated molecules and t_{pulse} =2 s. $\tilde{C}$ is the concentration obtained from a delta function (infinitely short and infinitely bright) pulse of light. To justify the validity of our calculations, we visualize a similar gradient forming by photoactivation of caged fluorescein and fit the solution of the diffusion equation to the fluorescein gradient. The inset of Figure 4.3.1 shows the gradient formed by diffusion of photoactivated fluorescein from the illumination area, 6 seconds after the start of the pulse.

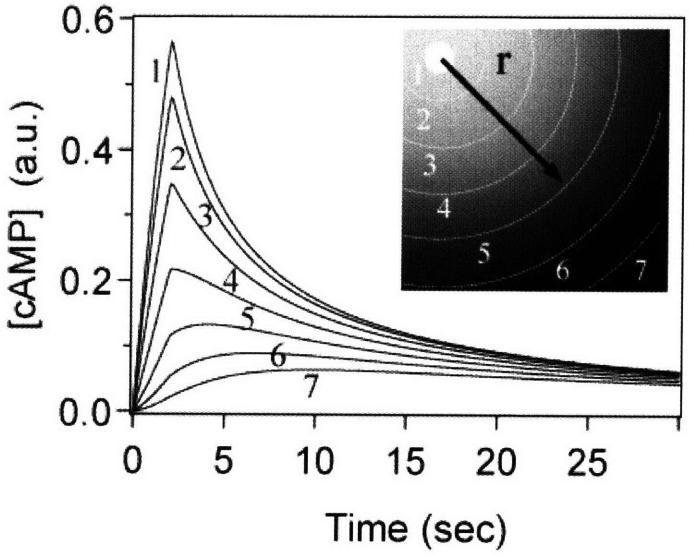


Figure 4.3.1 Quantitative measurement of the spatio-temporal variation of cAMP concentration as a function of time and distance from the UV spot. cAMP concentration is calculated by solving the diffusion equation and averaging the concentration over the area enclosed between $r \pm 1$ μm. The average cAMP concentration as a function of time is plotted for r = 10, 31, 52, 73, 94, 115, 136 ± 1 μm, which correspond to the regions 1-7 in the inset respectively. Inset: Image of the gradient forming by photoactivation of caged fluorescein after 2 seconds of UV illumination in a 15 μm circular region in the upper left hand corner of the image.

cAMP diffusion was estimated based on the Stokes-Einstein equation for spherical molecules, $D=k_BT/6\pi\eta a$ where k_B is the Boltzmann constant, T the temperature, η the viscosity, and a the radius of the molecule. Based on the molecular weight of cAMP, the equivalent spherical radius is about 0.516 nm. Therefore the Stokes-Einstein equation predicts a diffusion coefficient of $D = 4.3\times10^{-6}$ cm^2/s at 24 °C, consistent with previously measured value for cAMP diffusion constant (Dworkin et al. 1977).

While the diffusion equation gives us a relative measure of the cAMP concentration in the chamber, we must also estimate the uncaging efficiency χ in order to obtain an absolute measure. This efficiency is the number of photoactivated molecules divided by the total number of molecules. The quantum yield of uncaging Φ is defined as the number of photoactivated molecules divided by the number of absorbed photons n_{abs}. Following the Lambert-Beer law, the number of absorbed photons n_{abs} equals $n_o(1-\exp(-Ecl))$, in which n_0 is the photon flux, E is the molar extinction coefficient, l the length of the light path and c the concentration of caged molecules. For small Ecl, one can approximate the above formula by $n_{abs} = n_0Ecl$. The total number of molecules in the irradiation volume is the product of Avogadro's number N_A, c and V the volume of the exposed region. Therefore χ is given by:

$$\chi = \frac{\Phi n_o Ecl}{N_A cV} \qquad [3]$$

When irradiated with full intensity the photon flux density n_0/At is 4×10^{18} photons cm^{-2}s^{-1} as measured at the plane of focus, in which A is the area and t the time of exposure. Considering the extinction coefficient of E = 1500 cm^{-1}M^{-1}, at 400 nm and a quantum yield of 0.39 (Nerbonne et al. 1984; Corrie et al. 1993). It can be estimated that a 2 second pulse of light with full intensity will cleave about 1% of the NPE-caged cAMP molecules. The photon flux density will be slightly higher for the maximum wavelength of 390 nm that we use in this set up. Therefore we estimate χ to be slightly above 1%. Small χ is an advantage in our experiment, because caged compounds do not deplete during the 2 second illumination.

4.3.4 Quantifying the Cell's Response

Figure 4.3.2 panel b illustrates the dynamic translocation of CRAC-GFP to the membrane after stimulation with a 2 second UV pulse for a cell with a depolymerized actin cytoskeleton. Directional sensing does not require cell motility, morphological changes, or an intact cytoskeleton (Parent et al. 1998; Janetopoulos et al. 2004; Xu et al. 2005). Immobilized cells provide the advantage of studying directional sensing in the absence of the complex downstream responses such as changes in cell shape. The relative CRAC-GFP concentration in the membrane with respect to the pre-stimulus level was measured by subtracting the images taken after the release of the stimulus from the image taken just before the release of the stimulus (Figure 4.3.2c). We defined a response function $R(\theta,t)$, in order to quantify the relative CRAC-GFP concentration in the membrane. To determine $R(\theta,t)$, first the cell membrane was segmented in 20 sub-regions. Second, the difference between GFP intensity at time t and at $t = 0$ (just before uncaging) is computed for each of the 20 sub-regions. Finally we correct for the cell-to-cell variability in GFP levels by normalizing $R(\theta,t)$ by the average cytoplasmic GFP fluorescence at $t = 0$ available to each sub-region ($1/20^{th}$ of the total cytoplasmic GFP fluorescence). This ratio defines the response function $R(\theta,t)$. For example when $R(\theta,t)$ is 100% for all 20 segments this implies that every available CRAC-GFP molecule has been recruited uniformly to the membrane. However, if 10 adjacent sub-regions have $R(\theta,t) = 100\%$ and the remaining 10 sub-regions have $R(\theta,t) = 0\%$, this implies that 50% of the total cytoplasmic CRAC-molecules have been recruited to 10 sub-regions. Conversely, if $R(\theta,t)$ is 0% for a specific segment this means the CRAC-GFP concentration in the membrane is identical before and after cAMP uncaging. Figure 4.3.2d displays $R(\theta,t)$ for the cell depicted in Figure 4.3.2c. $R(\theta,t)$ shows a clear polarized response which is most pronounced about 8 seconds after uncaging. We have characterized $R(\theta,t)$ with three parameters: localization L, polarization P, and polarization angle ϕ. These three parameters are determined by fitting the experimentally obtained response function $R(\theta,t)$ with $R_{fit}(\theta,t) = L(t) + P(t)\cos\left[\theta - \phi(t)\right]$.

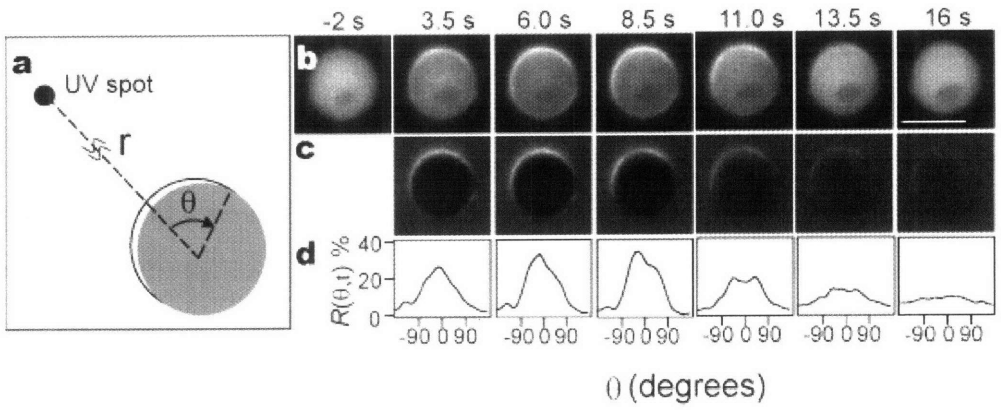


Figure 4.3.2 Dynamic translocation of CRAC-GFP at the plasma membrane after stimulation with a 2 second pulse of cAMP **(a)** The UV uncaging location is positioned a distance *r* away from the cell center. The angle θ defines the coordinate along the cell's periphery, where θ = 0 defines the position at the membrane that is closest to the uncaging location. **(b)** Unprocessed epi-fluorescence images displaying CRAC-GFP as a function of time. The scale bar denotes 10 μm and *r* = 70 μm. **(c)** Subtracted images illustrates the relative change of CRAC-GFP concentration in the membrane with respect to the pre-stimulus level (*t* = -2 s). **(d)** Response function, $R(\theta,t)$ as a function of time for images in c.

The localization $L(t)$ is defined as the average value of $R(\theta,t)$, $L(t) = \langle R(\theta,t) \rangle_\theta$, and therefore reflects the average recruitment of CRAC-GFP molecules to the membrane a time *t* after cAMP was uncaged. The brackets $\langle ... \rangle_\theta$ denote that the response function is averaged over the coordinate θ. The polarization $P(t)$ is a measure of the amplitude of the response function and therefore reflects half the difference between the maximum and minimum value of $R_{fit}(\theta,t)$ of a single cell at a time *t* after cAMP uncaging. Finally, the polarization angle $\phi(t)$ is defined as the angle for which the fit function $R_{fit}(\theta,t)$ reaches its maximum value, relative to the direction of the uncaging. Therefore, a value of $\phi = 0$ implies that the cell has recruited most of the CRAC-GFP molecules to the position at the membrane that is closest to the uncaging spot ($\phi = 0$, Figure 4.3.2a). Conversely, a value of $\phi = 180^\circ$ means that the cell has polarized in the opposite direction and recruits most of the CRAC-GFP molecules to the position in the membrane that is furthest away from the uncaging spot.

4.3.5 Response Dynamics

Figure 4.3.3 displays the dynamics of L, P and ϕ for a single cell (Figure 4.3.3 panels b, d, and f) and a population of 40 cells (Figure 4.3.3 panels c, e, and g). After stimulation with a pulse of cAMP, both L and P increase reaching a maximum value followed by a return to their prestimulus level in about 30 seconds. The time at which L reaches its maximum is defined as T_{max}. For a single cell, ϕ remains more or less constant during the response time (Figure 4.3.3f). The error bars on the single cell data reflect the variability from pulse-to-pulse, which is significantly smaller than the variability from cell-to-cell, denoted by the error bars on the population data. Taken together these data suggest that when a single cell is repeatedly stimulated with identical pulses, it responds in a highly reproducible manner. From pulse-to-pulse it recruits a very similar average CRAC-GFP concentration to the membrane (reflected in $L(t)$); it creates a very similar CRAC-GFP gradient at the membrane (reflected in $P(t)$); and this gradient is oriented in the same direction from pulse-to-pulse (reflected in $\phi(t)$). However from cell-to-cell a large variability is observed in these three parameters.

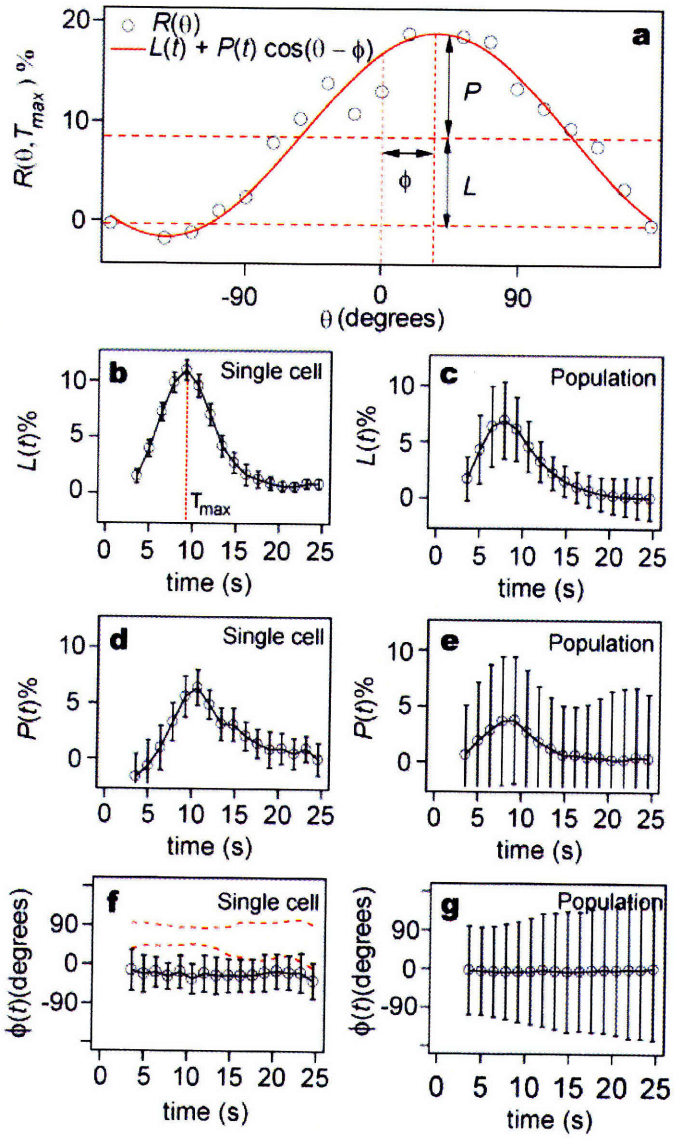


Figure 4.3.3 Definition of localization L, polarization P and polarization angle ϕ and comparison between the time dependence of these parameters for a single cell (which is stimulated 10 times) and a population of 40 cells (which are stimulated once). **(a)** The response function $R(\theta, T_{max})$ (circles), and the fitting function $R_{fit}(\theta, T_{max})$ (red line). **(b)** Time dependence of L for a single cell. **(c)** Time dependence of the average L for a population. **(d)** Time dependence of P for a single cell. **(e)** Time dependence of the average P for a population. **(f)** Time dependence of ϕ for a single cell. The two dashed red lines indicate the dynamics of ϕ for two other single cells. ϕ is very reproducible from pulse-to-pulse, even when $\phi \neq 0$. **(g)** Time dependence of the average ϕ for a population of 40 cells, which averages to zero. Error bars denote standard deviations.

4.3.6 Response Direction

In Figure 4.3.4, we further quantify the cell-to-cell, versus pulse-to-pulse variability of $R(\theta,T_{max})$. $R(\theta,T_{max})$ is highly reproducible from pulse-to-pulse, when a single cell is stimulated with 10 identical pulses of cAMP (Figure 4.3.4a). In contrast, $R(\theta,T_{max})$ for a population shows a large variability from cell-to-cell even though cells are stimulated with the same identical pulse (Figure 4.3.4b), consistent with the data presented in Figure 4.3.3 panels b-g.

Since the cell-to-cell variability is most pronounced for the polarization P and polarization angle ϕ it is convenient to present the data in a polar plot as shown in Figure 4.3.4c and Figure 4.3.4d. In these figures one data point represents data from a single cell at T_{max}. The distance from a data point to the origin of the polar plot equals the polarization $P(T_{max})$. The angle between the x-axis and the line which connects the data point to the origin of the polar plot is the polarization angle $\phi(T_{max})$. The x-axis is along the line which connects the center of each cell to the center of the uncaging spot ($\theta = 0$, Figure 4.3.2a) and the y-axis is perpendicular to that line ($\theta = 90^{\circ}$). In this representation the polarization component along the x-axis P_x equals $P(T_{\max})\cos[\phi(T_{\max})]$. The polarization component along the y-axis, P_y equals $P(T_{\max})\sin[\phi(T_{\max})]$. For example, cells responding along the 'right' direction ($\phi = 0$) are characterized by $P_x > 0$ and $P_y = 0$, whereas cells that respond in the 'wrong' direction ($\phi = 180^{\circ}$) are characterized by $P_x < 0$ and $P_y = 0$.

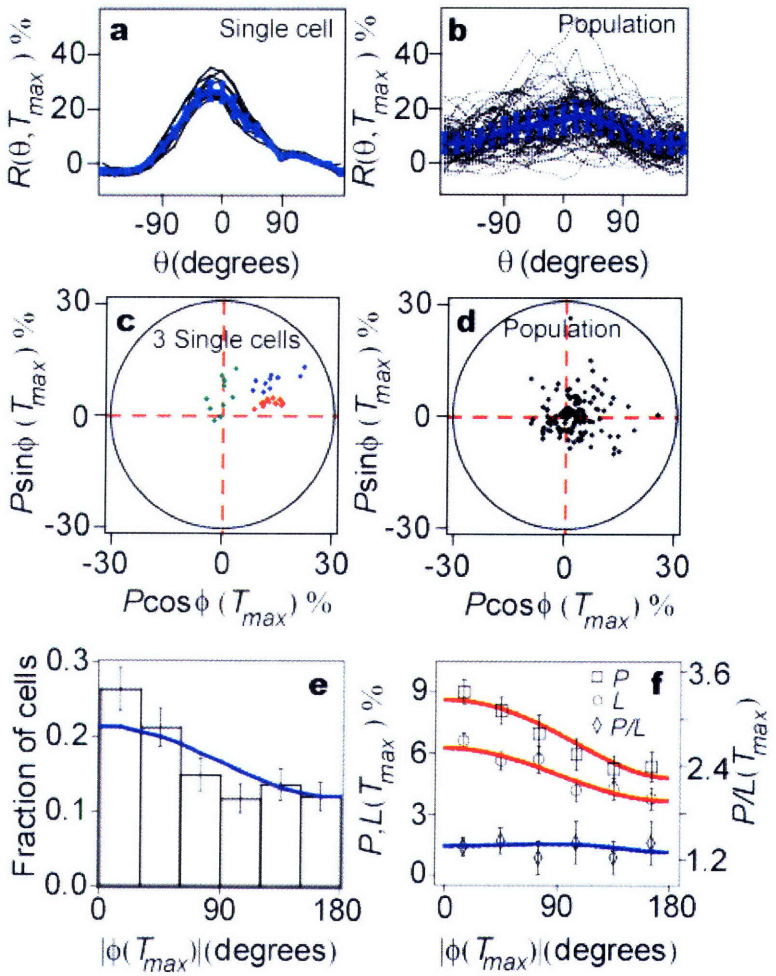


Figure 4.3.4 Cell-to-Cell versus pulse to pulse variability. **(a)** Comparison between $R(\theta, T_{max})$ for a single cell, which is stimulated 10 times and **(b)** a population of 40 cells, which are stimulated once. **(c)** Polar plot of the polarization at T_{max} for three single cells and **(d)** a population of 100 cells. In this representation one data point represents data from a single cell at T_{max}. The distance from a data point to the origin of the polar plot equals the polarization $P(T_{max})$. The angle between the x-axis and the line which connects the data point to the origin of the polar plot is the polarization angle $\phi(T_{max})$. **(e)** Population probability distribution of $|\phi(T_{max})|$ illustrating the fraction of cells displaying a particular polarization angle at T_{max}. f, Average of $L(T_{max})$ and $P(T_{max})$ (left ordinate) and the ratio of $P(T_{max})/L(T_{max})$ (right ordinate) as a function of $|\phi(T_{max})|$. Solid lines in panels e and f are predictions of the geometric model.

In Figure 4.3.4c and Figure 4.3.4d we present P and ϕ in a polar plot. For a single cell, ϕ is observed not to vary significantly from pulse-to-pulse (red dots in Figure 4.3.4c), even when $\phi \approx 90^{\circ}$ (green dots in Figure 4.3.4c). In contrast, a pronounced cell-to-cell variability of the polarization angle ϕ is observed in the population even though the cells are stimulated with the same pulse of cAMP (Figure 4.3.4d). The distribution of $|\phi(T_{max})|$ is plotted in Figure 4.3.4e, and shows a peak near $\phi \approx 0$. These experiments demonstrate that although there is a significant variability in $\phi(T_{max})$ from cell-to-cell, the

population faithfully detects the direction of the cAMP pulse. In contrast, the distribution of $|\phi(T_{max})|$ for uniformly stimulated cells is flat (Figure 4.3.5a). Here the uniform stimulation was created by uniformly uncaging cAMP by illuminating the entire field of view with UV light.

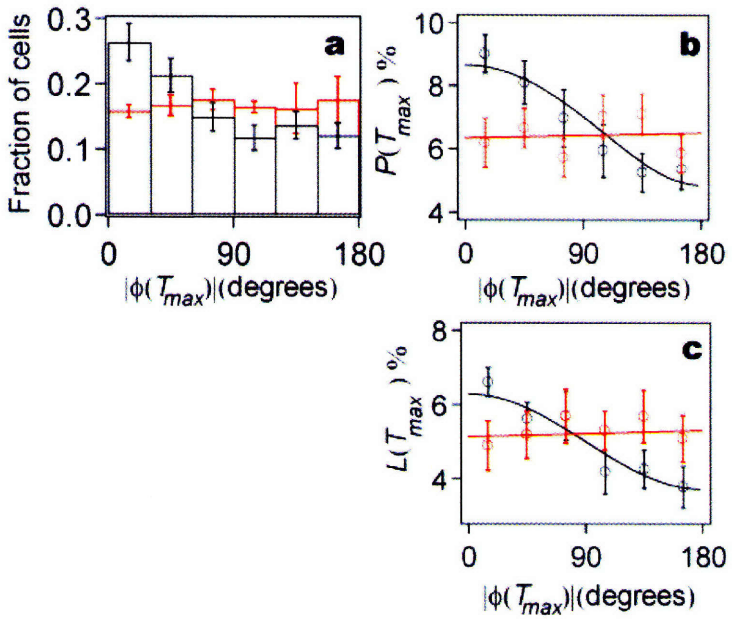


Figure 4.3.5 Comparison between cells that are stimulated with a directed pulse of cAMP (black) and a uniform pulse of cAMP (red). **(a)** Probability distribution of $\phi(T_{max})$. The distribution of the polarization angles ϕ is symmetric around $\theta = 0$ and therefore, to increase the statistics, the histogram of the $|\phi(T_{max})|$ was calculated. The histogram of $|\phi(T_{max})|$ then is normalized by the total number of cells. **(b)** The average polarization *P* as a function of polarization angle ϕ. **(c)** The average localization *L* as a function of polarization angle ϕ.

We find that both the localization and polarization show a strong correlation with the polarization angle as displayed in Figure 4.3.4f. Cells that correctly detect the direction of the pulse ($\phi \approx 0$) have, on average, a two-fold larger localization and polarization than cells that polarize in the opposite direction ($\phi \approx 180^{o}$). We will discuss this correlation after introduction of the geometric model. Note that this correlation is absent when cells are stimulated with a uniform cAMP pulse (Figure 4.3.5b and Figure 4.3.5c), and the ratio of polarization and localization *P/L* does not show a significant correlation with the polarization angle (Figure 4.3.4f).

4.3.7 Directional Variability is Internal to Cell

In order to verify whether the noise at the level cAMP-receptor binding is significant, we varied the extracellular concentration of cAMP over two orders of magnitude (Berg et al. 1977; Tranquillo et al. 1988; Bialek et al. 2005). $P\sin(\phi)$ versus $P\cos(\phi)$ at T_{max} for a population of 100 cells, which are stimulated with cAMP concentrations of 1 nM to 0.1 μM are plotted in Figure 4.3.6a. Probability distributions of $|\phi(T_{max})|$, when the extracellular concentration of cAMP is varied from 1 nM to 0.1 μM are plotted in Figure 4.3.6b. The standard deviations for the probability distributions are approximately 100° for all cAMP concentrations used. We observed that by increasing the extracellular concentration over two orders of magnitude, the probability of finding a cell with ϕ in the direction of the extracellular gradient (θ = 0) does not significantly increase. This indicates that an intracellular mechanism is responsible for the observed variability in P and ϕ. The same angular distribution is obtained over a wide concentration range of caged cAMP (Figure 4.3.6). This excludes the possibility that the variability is introduced at the level of the cAMP-receptor binding (Berg et al. 1977; Bialek et al. 2005).

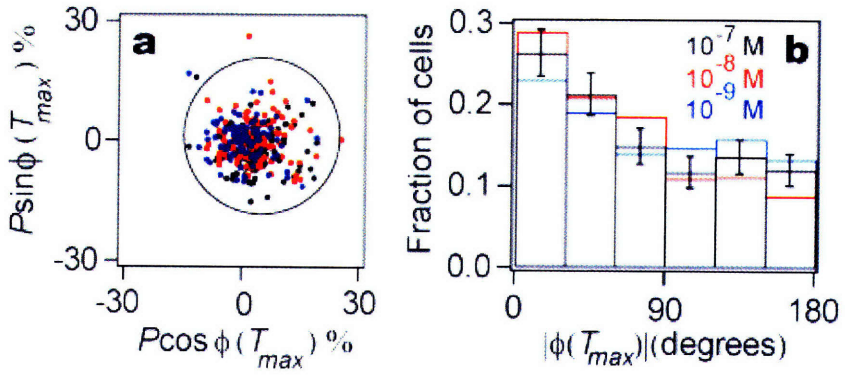


Figure 4.3.6 Dependence of the variability observed in P and ϕ on the extracellular concentration **(a)** Polar plots of P_y versus P_x at T_{max} for three different cAMP concentrations and for a population of 100 cells. cAMP concentration varied between10^{-9} M to 10^{-7}M. Similar to Figure 4.3.4d, the polarization components are P_x equals $P(T_{\max})\cos[\phi(T_{\max})]$ along the x-axis and P_y equals $P(T_{\max})\sin[\phi(T_{\max})]$ along the y-axis. Each dot represents data from a single cell. **(b)** Normalized histogram of ϕ as a function of θ when the extracellular concentration is increased over two orders of magnitude. By increasing the extracellular concentration, the probability of finding a cell with a smaller polarization angle does not increase significantly.

4.3.8 Geometric Model

A few stochastic models have been developed to describe random cell motility (Dallon et al. 1997; Arrieumerlou et al. 2005), however existing models of eukaryotic directional sensing are commonly constructed from spatially symmetric and deterministic systems of partial differential equations (Meinhardt 1999; Narang et al. 2001; Postma et al. 2001; Levchenko et al. 2002; Rappel et al. 2002; Janetopoulos et al. 2004; Gamba et al. 2005). Although the existence of an extracellular gradient can break the symmetry, the polarization by default will always occur along the direction of the extracellular gradient, yielding $\phi = 0$ for all cells. Therefore in order for a cell to polarize in a direction other than the direction of the extracellular gradient, an additional source of symmetry breaking must be present in the cell's gradient sensing network.

Rather than modifying existing models designed to capture the full dynamics of directional sensing, we introduce a simple linear model that allows for symmetry breaking and focuses on the cell-to-cell variability. To introduce a simple and general form of asymmetry, we assume that a cell has a static intracellular signal, $S_{int}(\theta)$. The intracellular signal might be caused, for example, by spatial inhomogenities in any of the signaling molecules in the gradient sensing pathway, or might include an inhomogeneous distribution of cortical factors that remain bound to the membrane after actin depolymerization. It is unlikely that the symmetry is broken by an inhomogeneous distribution of cAMP receptors since the receptors are essentially uniformly distributed along the membrane (Xiao et al. 1997). Our model accounts phenomenologically for this randomly oriented intracellular asymmetry without making statements regarding the molecular origins of the asymmetry.

This intracellular signal, in general could be a very complicated function of the angle θ. Because we observe that polarization is well approximated by the lowest frequency cosine term, the lowest frequency components of the intracellular signal are also assumed to be the most dominant terms in producing polarization. Similarly, the extracellular signal in our experiment can be approximated by a single cosine function, which describes how the extracellular cAMP concentration varies along cell membrane.

The main assertion of the model is that the extracellular cAMP signal $S_{ext}(\theta)$ is combined multiplicatively with the intracellular signal $S_{int}(\theta)$ to produce an effective signal $S_{eff}(\theta)$. The following equations can be used to describe the geometric model:

$$S_{ext} = S_0 + S_1 \cos\theta \quad [4]$$

$$S_{\text{int}} = 1 + \varepsilon \cos(\theta - \phi_\varepsilon) \quad [5]$$

$$S = S_{ext} \times S_{\text{int}} \approx \left(S_0 + \frac{\varepsilon}{2} S_1 \cos\phi_\varepsilon \right) + \left(S_1 + \varepsilon S_0 \cos(\phi_\varepsilon)\right)\cos\theta + \left(\varepsilon S_0 \sin(\phi_\varepsilon)\right)\sin\theta \quad [6]$$

The first of these equations characterizes the extracellular cAMP gradient around the cell periphery by an average concentration S_0 and an amplitude S_1. The coordinate θ defines the position along the membrane as defined in Figure 4.3.2a. Similarly, in the second equation the intracellular signal is characterized by a magnitude of the intracellular asymmetry ε and an angle ϕ_ε, which is the polar coordinate for which the intracellular signal is largest (Figure 4.3.7d). The final equation defines a total signal S as a multiplicative function of the extracellular and the intracellular signals, which can be approximated as:

$$S = L + P_x \cos(\theta) + P_y \sin(\theta), \text{ where} \quad [7]$$

$$L = S_0 + (\varepsilon/2) S_1 \cos\phi_\varepsilon \quad [8]$$

$$P_x = S_1 + \varepsilon S_0 \cos\phi_\varepsilon \quad [9]$$

$$P_y = \varepsilon S_0 \sin\phi_\varepsilon \quad [10]$$

$$\phi = \arctan\left[P_y / P_x\right] \quad [11]$$

This set of equations gives us a way to read out the localization (L) and polarization (P_x, P_y) when given the magnitude (ε) and direction (ϕ_ε) of the intracellular signal and the magnitude of the extracellular signal (S_0, S_1)

When cells are stimulated with a uniform extracellular stimulation, the effective signal retains the direction of the intracellular signal and the direction of the effective polarization simply matches the direction of the intracellular signal ϕ_ε (Figure 4.3.7a). Figure 4.3.7b illustrates the case where a cell is stimulated with a directed pulse of cAMP. In this case, an effective signal, whose polarization is biased by the direction of an intracellular asymmetry causes a cell to polarize in a direction different from either the intracellular or extracellular signal.

Equations [7]-[11] can be geometrically represented in a polar plot when P_x and P_y are used as the x and y coordinates of the polar plot, respectively (Figure 4.3.7c-d). The polarization vector of the cell (black arrow, Figure 4.3.7d) is proportional to the sum

of a vector with length S_1 (red arrow, Figure 4.3.7d) and a vector with length εS_o and angle ϕ_ε (blue arrow, Figure 4.3.7d). The extracellular cAMP signal is parameterized by $S_{ext}(\theta) = S_o + S_1 \cos\theta$, and the intracellular signal by $S_{\text{int}}(\theta) = 1 + \varepsilon \cos(\theta - \phi_\varepsilon)$, where the parameters S_o and S_1 reflect the average cAMP concentration and cAMP gradient, respectively. The parameters ϕ_ε and ε define the orientation and relative strength of the intracellular signal, respectively. This model predicts that for a uniform stimulation (S_1 = 0), the polarization angle ϕ equals ϕ_ε (Figure 4.3.7c). Experimentally we find that, when a population is exposed to a uniform stimulation, the polarization angles ϕ are uniformly distributed from cell-to-cell (Figure 4.3.7e). This implies that the orientation of the intracellular signal ϕ_ε follows the same distribution. For a directed pulse ($S_1 \neq 0$) ϕ generally not does equal ϕ_ε. In this case one would expect a non-uniform ϕ-distribution with a maximum at $\theta = 0$ as was experimentally observed (Figure 4.3.4e, Figure 4.3.7f). These results demonstrate that not only the direction of the intracellular asymmetry ϕ_ε, but also the magnitude ε varies from cell-to-cell.

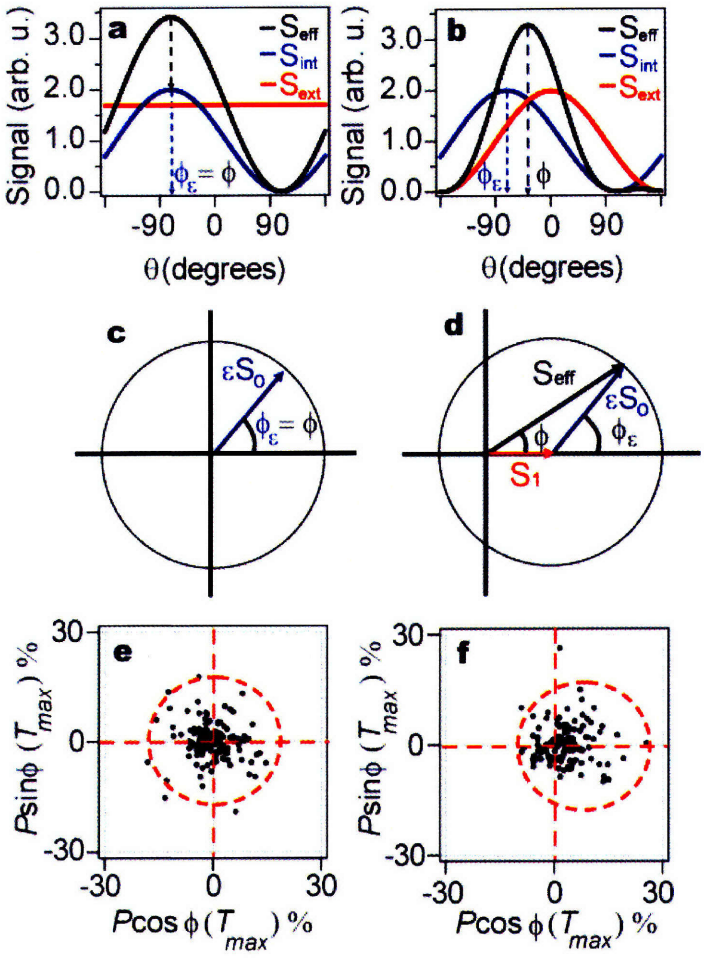


Figure 4.3.7 (a) Schematic illustration of the effective signal (black), which is a combination of the intracellular signal (blue) and the extracellular signal (red) for a uniform cAMP stimulus and **(b)** a directed pulse of cAMP. **(c)** Graphical representation of the geometric model and the polarization angle ϕ, when cells are stimulated with a uniform pulse of cAMP and **(d)** a directed pulse of cAMP. The effective polarization angle strongly depends on the direction of the intracellular signal ϕ_ε. **(e)** Experimentally measured polar plots for a uniform pulse of cAMP and **(f)** a directed pulse of cAMP.

The model also qualitatively explains the experimental correlation between the polarization *P* and polarization angle ϕ (Figure 4.3.4f). Because ϕ_ε is uniformly distributed (Figure 4.3.4e) the strongest polarization is expected for cells in which the intracellular and extracellular signal line up (Figure 4.3.4d, $\phi_\varepsilon = 0$). In contrast, for cells in which the intracellular and extracellular signal are oriented in opposite directions ($\phi_\varepsilon = 180^o$) the polarization *P* is expected to be the smallest possible. A similar argument can be made to explain the correlation between localization *L* and polarization angle ϕ. Equation 8 of the geometric model predicts that the localization *L* is the strongest, when the intracellular and extracellular signals line up ($\phi_\varepsilon = 0$) and smallest when the intracellular

and extracellular signals are oriented in opposite directions ($\phi_\varepsilon = 180^o$). While the model predicts that the ϕ distribution will change as a function of the extracellular signal strength, it is possible that the greater cAMP concentrations tested experimentally in Figure 4.3.6 are saturating the signaling pathway upstream of the point at which the intracellular and extracellular signals mix. That would means that in these experiments cells are seeing effectively the same strength extracellular signal even though the cAMP concentration has increased 100 fold, and would be consistent with the constant ϕ distribution observed there.

Each data point in Figure 4.3.7e (up to a constant S_0) contains the information about the magnitude and the direction of the intracellular signal for each cell. The mean of the εS_0–distribution for a population of 137 cells is (6.3 ± 0.4) with a standard deviation of (4.4 ± 0.3) leading to a coefficient of variation of (0.70 ± 0.04) indicating that the strength of intracellular signal is highly variable from cell-to-cell.

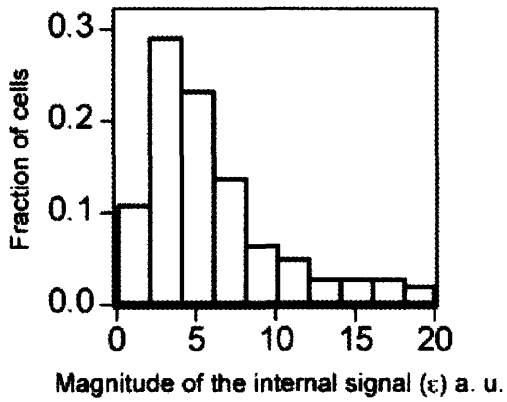


Figure 4.3.8 Normalized histogram of the magnitude of the internal signal (ε). The histogram of ε is calculated from data in Figure 4.3.7e, where cells are stimulated with a uniform cAMP concentration. The mean of the ε-distribution for a population of 137 cells is (6.3 ± 0.4) with a standard deviation of (4.4 ± 0.3) leading to a coefficient of variation of (0.70 ± 0.04) indicating that the strength of intracellular signal is highly variable from cell-to-cell.

4.3.9 Moving Source Experiment

To test the model more directly, we stimulated a single cell from multiple directions. We varied the stimulation angle θ_s, such that a cell is stimulated from 8 different locations separated by 45°, while the orientation of the cell and the direction of the intracellular signal remain fixed (Figure 4.3.9). Figure 4.3.10a schematically illustrates the geometric model in the frame of reference of a single cell, where the angle of the intracellular signal is fixed at ϕ_ε and the angle of the extracellular signal θ_s is rotated around the cell.

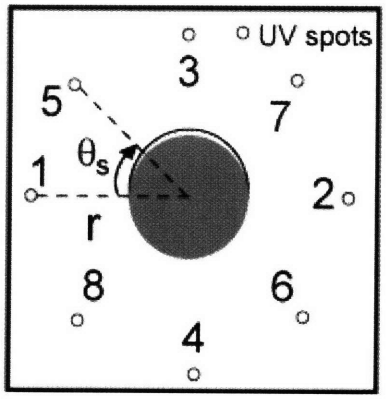


Figure 4.3.9 Schematic illustration of the setup for rotating pulse experiments. The location of the UV spot was kept at a distance *r* away from the cell's center. However the angle of stimulation, θ_s varied every 45 degrees, stimulating the cell from 8 different locations around the cells. Locations of the UV spots with respect to the cell are indicated by hollow circles. To avoid any effect of spatial or temporal memory of the previous pulse, the pulses were separated by 2 minutes and were given in the numerical order illustrated.

For a cell with small ε ($<< S_1/S_0$), we expect the contribution of the intracellular signal to the effective signal to be minimal causing the effective signal to follow the extracellular signal exactly ($\phi \approx \theta_s$, Figure 4.3.10 panels b and c). In contrast, for a cell with large ε ($>> S_1/S_0$), the contribution of the extracellular signal to effective signal is minimal and therefore the effective signal will be predominately in the direction of the intracellular signal ($\phi \approx \phi_\varepsilon$,Figure 4.3.10e). For a cell with $\varepsilon \approx S_1/S_0$, the intracellular and extracellular signals have equal strengths resulting in an effective signal which is neither in the direction of the intracellular nor extracellular signal (Figure 4.3.10d).

The plots of ϕ as a function of θ_s in Figure 4.3.10 panels c, d and e were fitted to the geometric model described by the following function:

$$\phi = \arctan\left[\frac{\alpha \sin(\phi_\varepsilon) + \sin(\theta_s)}{\alpha \cos(\phi_\varepsilon) + \cos(\theta_s)}\right], \quad \text{where} \qquad \textbf{[12]}$$

$$\alpha = \frac{\varepsilon S_0}{S_1}, \qquad \textbf{[13]}$$

with α and ϕ_ε as the fitting parameters. The ratio of S_1/S_0 was independently calculated to be 0.38 by integrating the extracellular gradient from $t = 0$ to T_{max} using Eq. [1] and Eq. [2]. Our calculation for S_1/S_0, in addition was experimentally verified by using a caged fluorescent marker. The fits to the geometric model using the two parameters $\alpha=\varepsilon S_0/S_1$ and ϕ_ε are in good agreement with our experiments (red lines in Figure 4.3.10 panels c, d, and e).

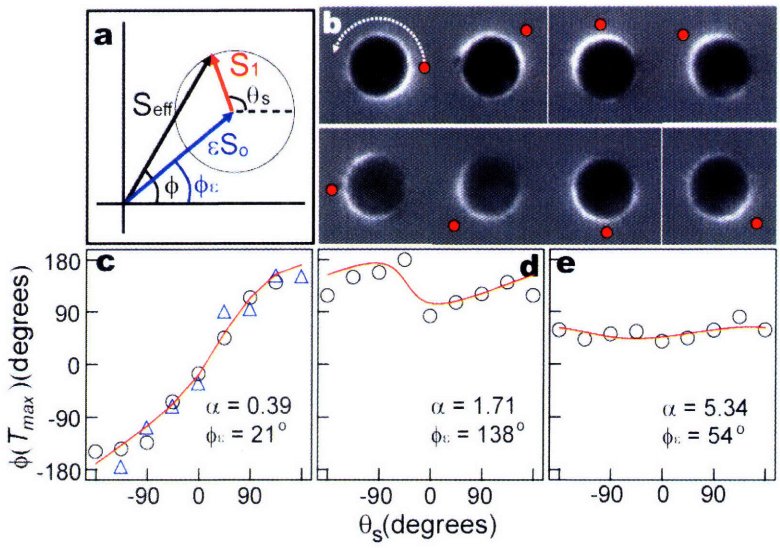


Figure 4.3.10 Experimentally measured relation between the polarization angle $\phi(T_{max})$ and the extracellular signal θ_s, when the direction of the extracellular signal is varied relative to the intracellular signal ϕ_ε and comparison to the geometric model **(a)** Schematic illustration of the geometric model in the frame of reference of a cell with a fixed ϕ_ε. **(b)** The difference image for a cell with a small $\alpha << 1$. Red dots on the images indicate the direction of extracellular stimulation. **(c)** $\phi(T_{max})$ versus θ_s for a cell with small $\alpha << 1$. The triangles and circles denote two independent experiments demonstrating the reproducibility of this assay. **(d)** $\phi(T_{max})$ versus θ_s for a cell with an intermediate $\alpha \approx 1$ and **(e)** a large $\alpha >> 1$. The red lines represent fits to the geometric model with fitting parameters α and ϕ_ε.

4.3.10 Model Predicts Angular Dependence

To further challenge the predictive power of the geometric model, we tested if the model combined with the single cell data in Figure 4.3.10, could predict the population experiments summarized in Figure 4.3.4e-f. We experimentally find α for 20 cells (using the method outlined in Figure 4.3.10) and find an average value of α = 3.6. Using this experimental value for α the model successfully predicts the ϕ-distribution and the ratio *P/L* without any fit parameters (Figure 4.3.4e-f, blue lines).

Equations [7]–[11] can be simplified to demonstrate that the only relevant parameters in calculating the fraction of cells as a function of $|\phi(T_{max})|$ (Figure 4.3.4e) are α and ϕ_ε:

$$\phi = \arctan\left[P_y / P_x\right] = \arctan\left[\frac{\sin\phi_\varepsilon}{(1/\alpha + \cos\phi_\varepsilon)}\right] \qquad \textbf{[14]}$$

In order to obtain a mean value for α (and therefore ε), we fitted the experimental curves of ϕ as a function of θ_s obtained from 20 individual cells to Eq. [12] (as described in previous section). From these fits, the mean value of α was measured to be 3.6 (giving a mean value of 1.3 for ε). The distribution of ϕ_ε was also experimentally measured to be uniformly distributed between [0, 2π] (Figure 4.3.4). Therefore both α and ϕ_ε (also ε) are experimentally determined and therefore the red line in Figure 4.3.4e is the prediction of the geometric model with no fitting parameters.

Similarly, we demonstrate that the correlation between $L(T_{max})$ and $P(T_{max})$, and $|\phi(T_{max})|$, illustrated in Figure 4.3.4f, depend on ϕ_ε, α and ε (Eq. [15] and [16]). However these also depend on an undetermined constant, which is the ratio between the cAMP concentration and the experimentally measured response function $R(\theta)$. The two red lines in Figure 4.3.4f, are the geometric model fit to the curves $L,P(T_{max})$ versus $|\phi(T_{max})|$ using only one fit parameter.

$$L = S_0(1+\frac{\varepsilon^2}{2\alpha}\cos\phi_\varepsilon) \quad [15]$$

$$P = \varepsilon S_0((1/\alpha + \cos\phi_\varepsilon)^2 + \sin^2\phi_\varepsilon)^{1/2} \quad [16]$$

$$P/L = \varepsilon \frac{\left[(1/\alpha + \cos\phi_\varepsilon)^2 + \sin^2\phi_\varepsilon)\right]^{1/2}}{(1+\frac{\varepsilon^2}{2\alpha}\cos\phi_\varepsilon)} \quad [17]$$

The ratio of P and L at T_{max} as a function of $|\phi(T_{max})|$ illustrated in Figure 4.3.4f is a function of ϕ_ε, α and ε given by Eq. [17]. All these three parameters are experimentally determined and therefore the blue line in Figure 4.3.4f is the prediction of the geometric model with no fitting parameters.

To predict how L and P vary with ϕ (Figure 4.3.6f), we require one undetermined proportionality constant reflecting the ratio between the cAMP concentration and the experimentally measured response function $R(\theta)$. When this factor is used as the sole fit parameter, the experimental data closely match the model predictions (Figure 4.3.4f, red lines). It is encouraging that this simple linear model correctly captures the key properties of the observed stochasticity in directional sensing.

4.3.11 Discussion

Our results show that cellular asymmetries and cell-to-cell variability in the size of these asymmetries can have a significant impact on the fidelity of directional sensing. Whereas some individual cells correctly detect the extracellular cue, most cells display a significant deviation from this direction due to an intracellular asymmetry. However, this does not hinder a population of many cells from accurately detecting the direction of the extracellular cue (Figure 4.3.4e). The geometric model provides an intuitive explanation for this. The model suggests that the effective signal is the product of a randomly oriented intracellular signal and the extracellular cue. Individual cells that, by chance, have the intracellular signal aligned with the extracellular cue will have a large effective signal and will therefore display a larger polarization than cells in which the intracellular and extracellular signal are counteracting. This mechanism biases the net polarization of the population towards the extracellular cue. This provides a heterogeneous population with a simple, yet effective, response strategy that provides an isotropically sensitive direction sensor even in the presence of large cell-to-cell variability.

4.3.12 Methods

Materials. Adenosine 3',5'- cyclic Monophosphate, p1-(2-Nitrophenyl)ethyl Ester (NPE-caged cAMP)(Calbiochem-Novabiochem, San Diego, CA) (32, 33) is used to create a controlled release and measurable gradient of cAMP. CMNB-caged fluorescein (fluorescein bis-(5-carboxymethoxy-2-nitrobenzyl) ether, dipotassium salt) (Molecular Probes, Eugene, OR) is used to verify the two dimensional diffusion calculations and visualize the gradient. Latrunculin A (Molecular Probes, Eugene, OR) is used to depolymerize actin.

Cell culture. A *Dictyostelium discodideum* cell line expressing the CRAC-GFP was constructed by electroporating plasmid pWF1 (generous gift of C. Parent) into wild type AX3 cells. CRAC-GFP cells were cultured and selected in HL5 medium with 20 μg/ml G418 and grown a density of 5×10^6 cells/ml. *Dictyostelium* cells harvested by centrifugation were suspended in development buffer (DB, 10 mM phosphate buffer, 2 mM $MgSO_4$, 0.2 mM $CaCl_2$). Cells are starved for 5 hours in DB by repeated pulses of 75 nM cAMP every 6 minutes (12). Subsequently, starved cells were harvested and diluted 10^3 fold in DB to reach the density of 10^3 cells/ml and treated with 0.5 μM Latrunculin A, 10 minutes prior to observation. Cells were then seeded into the observation chamber, a round well with inner diameter of 20 mm and 1 mm depth adhered to a microscope slide (FW20 well, Grace bio-labs, Inc.). The observation chamber was covered with a cover slip and mounted on an inverted Nikon TE2000 microscope.

Data analysis. In our image analysis, we treat the cells as quasi two dimensional objects. Three dimensional deconvolution analysis demonstrated that latrunculin treated cells resemble spread-out droplets, with a maximum height of about ¼ of the cell's diameter. The diameter of a typical cell is about 10 μm and therefore the height is about 2.5 μm. Cell were imaged with a 60× objective with a depth of field of about 1 μm. Therefore a small contribution of the variability in the magnitude of the intracellular signal may be due to the out-of-focus fluorescence. Cells and their edges were determined using nearest neighbor cluster finding algorithm on auto-thresholded images. Abnormally shaped cells or cells that were touching were discarded for analysis. All curve fittings are done using the least squares method using Matlab.

Photoactivation of cAMP and gradient formation. A cAMP gradient was formed by photoactivation of a known concentration of NPE-caged cAMP. An inverted Nikon TE2000 microscope equipped with a 60× oil immersion objective and a 100 W mercury lamp was used. A UV-GFP (370-410 nm) filter was used to remove the visible component of the mercury lamp. The filtered light was focused through an objective into the observation chamber. To restrict the area of photoactivation, we place a 1 mm diameter pinhole in the light path before the objective. The light passing through the pinhole was focused by the objective to a 17 μm diameter spot on to the field of view. In all of our experiments the UV exposure time was 2 seconds. The setup was equipped with a CCD camera to permit time lapsed microscopy. Two types of short pulses were used in the experiments to stimulate cells: a directed pulse, which forms a spatio-temporal gradient of cAMP around the cell, and a pulse of spatially uniform cAMP concentration. A spatially uniform increase of cAMP concentration around the cell membrane was generated by illuminating the entire field of view with a low dose of UV exposure.

5 Populations in Fluctuating Environments

5.1 Bet Hedging as a Strategy for Surviving Fluctuating Environments

5.1.1 Introduction

Unlike controlled laboratory environments, cells in the wild have to face and surmount the challenges raised by random fluctuations in extracellular conditions (Cohen 1966; Levins 1968; Schaffer 1974; Stearns 1976). Signal transduction pathways allow cells to actively sense and respond to particular environmental changes (for example see Chapter 3.1). However, given the myriad of environmental conditions and the limited number of signal transduction pathways, it is likely that cells use complementary, more passive methods to anticipate changes in the environment. Without the need to sense the environment, cells can 'blindly' anticipate and survive environmental changes by randomly switching between multiple phenotypes, each fit to a particular environment. Following this bet-hedging strategy (Schaffer 1974; Slatkin 1974; Seger 1988), a population consisting of a variety of phenotypes enhances its fitness by ensuring that, at any given time, at least some of its members are prepared for an unforeseen environmental fluctuation.

This random bet-hedging is naturally employed by many microbial systems in order to cope with environmental uncertainty (Wolf et al. 2005). One prominent example known as phase variation (van der Woude et al. 2004) is defined by two distinct phenotypes between which cells reversibly transition. Phase variation in the expression of surface pili of *Escherichia coli* during infection of the urinary tract allows single cells to have two phenotypes: a free-floating phenotype and a phenotype that adheres to the urinary tract surface (Hernday et al. 2002; Wolf et al. 2002). Similarly, the methylation state of the *agn43* promoter of *E. coli* defines two phenotypes: a individual planktonic life-style or community life as part of a biofilm (Danese et al. 2000). More examples include the transition to competence in *Bacillus subtilis* (Grossman 1995; Suel et al. 2006), bistability in metabolic networks (Novick et al. 1957; Ozbudak et al. 2004), and antibiotic tolerance in many prokaryotes (Dhar et al. 2007). Transitions between the two distinct phenotypes are thought to be random and rare, meaning that the transition frequency is much smaller than the cell division rate. Therefore an important question

which arises in this context is how the inter-phenotype transition rates relate to the fitness of the population.

Several recent theoretical models, developed to address this topic, have proposed that an optimum bet-hedging strategy is achieved when the rate of stochastic transitions between phenotypic states is tuned to match the rate of fluctuations in the environment (Lachmann et al. 1996; Thattai et al. 2004; Kussell et al. 2005; Kussell et al. 2005; Wolf et al. 2005). Here we experimentally test this hypothesis *in vivo* using a re-engineered version of the galactose utilization network of the budding yeast *Saccharomyces cerevisiae*. We use this particular re-engineered network because it allows a more quantitative control over both the inter-phenotype switching rates and phenotypic growth rates than would be possible using one of the natural systems where bet-hedging is thought to play an important role. Stochastic fluctuations in gene expression (Paulsson 2004; Acar et al. 2005; Kaern et al. 2005; Raser et al. 2005; Kaufmann et al. 2007) in the re-engineered network allow a single cell to randomly switch back-and-forth between two distinct phenotypes. Each phenotype is designed to confer a growth advantage over the other phenotype in a certain environment. We experimentally demonstrate that in order to optimize population growth, cells have to match the inter-phenotype switching rate to the frequency of environmental changes. Our experiments demonstrate that random bet-hedging can constitute a simple, yet effective, survival strategy to cope with fluctuating environments without the need to actively track environmental conditions.

5.1.2 Discrete Model

In general, cells are faced with a continuous spectrum of dynamically changing environmental conditions, including fluctuations in temperature, pH, and concentrations of nutrients and toxins. To cope with these diverse environmental conditions, cells could potentially explore a multitude of phenotypes. Here we focus on a simpler, discrete system with two environments and two phenotypic states (Figure 5.1.1) that captures the important properties of bet-hedging strategies (Lachmann et al. 1996; Thattai et al. 2004; Wolf et al. 2005).

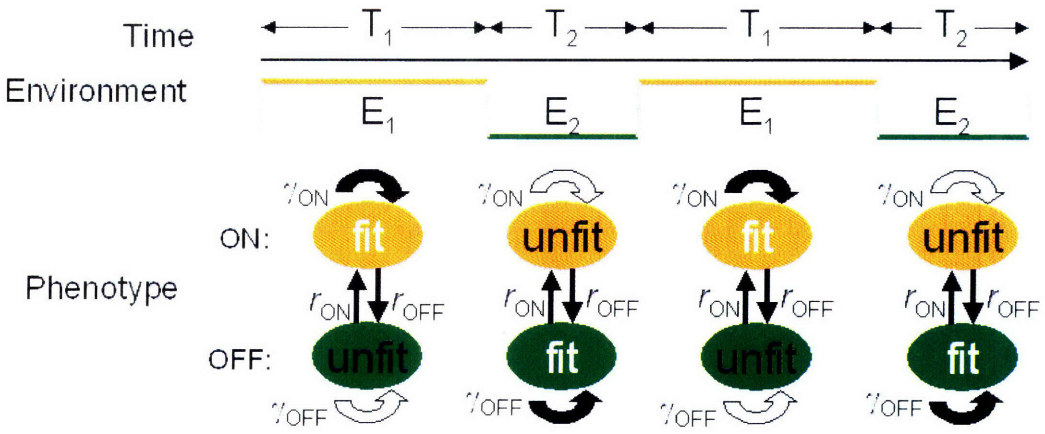


Figure 5.1.1 Two environment/two state hedge betting schematic. Two states (phenotypes) exist for each cell, ON (orange) and OFF (green). Cells randomly switch between the two states with frequencies r_{ON} and r_{OFF}. The first environment (E_1) has no uracil, while the second (E_2) has both 5-FOA and uracil. A cell is either fit or unfit to its environment depending on the specific phenotype it displays. For example in E_1, on-state cells are fit with a growth rate, γ_{ON}, but the unfit off-state cells proliferate with a smaller growth rate, γ_{OFF}.

If the stochastic switching between the two phenotypic states is much faster than the switching between the two environmental states, a high level of phenotypic diversity is expected (Figure 5.1.2, left panel). However, if the phenotypic switching rate is much slower than the environmental switching rate, the population is more homogenous at any given point in time (Figure 5.1.2, right panel).

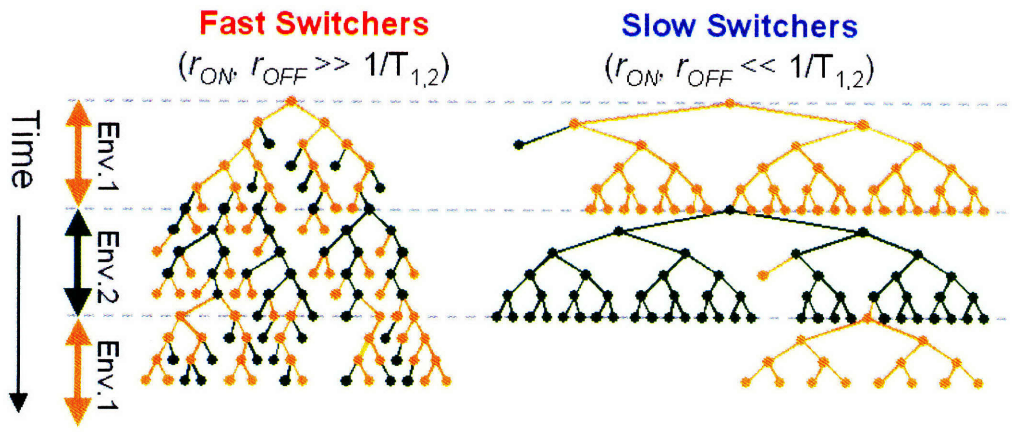


Figure 5.1.2 Cellular lineages for fast and slow switchers. Single cells with different switching frequencies (fast and slow) grow in alternating environments. Orange and green colors represent the first and second environments, respectively. Color change in the cellular lineage diagram corresponds to the change in phenotypic expression for a particular cell. If a cell finds itself in the unfit state after a switching event, it ceases to proliferate. In the case of slow switchers, reduced cell-to-cell variability in each environment is depicted by a more dominant use of a single color.

In a constant environment, the slow switchers maintain larger growth rates than fast switchers due to a greater fraction of cells in the fit state. However, slow switchers suffer greater losses and take more time to recover whenever an environmental transition occurs. In other words, a population must balance the rate at which cells transition between the phenotypic states with the frequency of environmental changes to maintain a reasonable number of cells in the unfit phenotype. This suggests that in order to optimally benefit from the diversity, cells should tune their switching rates to match the frequency of environmental changes (Lachmann et al. 1996; Thattai et al. 2004; Kussell et al. 2005; Kussell et al. 2005; Wolf et al. 2005).

5.1.3 Rewired Gal Network

To experimentally test this hypothesis we implemented this discrete switching system (Figure 5.1.1) *in vivo*. Bistable gene networks (Hasty et al. 2002; Angeli et al. 2004; Dubnau et al. 2006) provide promising experimental systems to implement discrete phenotypic states. The stochastic nature of gene expression drives rare transitions between these states (Hasty et al. 2000; Kepler et al. 2001; Kaern et al. 2005; Raser et al. 2005; Kaufmann et al. 2007). We utilize the bistable galactose utilization pathway of the budding yeast *Saccharomyces cerevisiae* (Acar et al. 2005; Hawkins et al. 2006), because the switching rates between the two phenotypic states can be tuned experimentally giving us precise control over the system (Acar et al. 2005). The wild type network was re-engineered as shown in Figure 5.1.3.

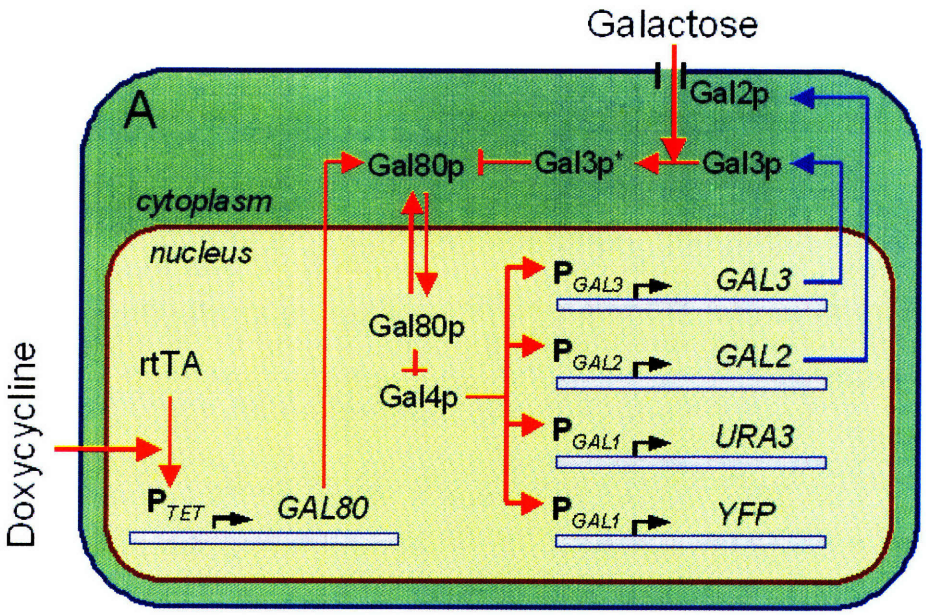


Figure 5.1.3 The galactose-signaling pathway. The activity of the galactose pathway is read out by using YFP driven by the *GAL1* promoter. Similarly, the endogenous *URA3* expression is also under the control of the *GAL1* promoter, coupling the synthesis of the Ura3 proteins to the activity of the *GAL* pathway. In environment E_1, ON cells will synthesize uracil and thrive, while in environment E_2 the *URA3* gene product is converted in a toxic intermediate in the presence of 5-FOA. By changing the extracellular galactose and doxycycline concentrations, the transition rates between the ON and OFF states can be altered, providing us with the fast and slow switchers.

The activity of the pathway was read out at the single cell level using yellow fluorescent protein (YFP) under the control of the *GAL1* promoter. For certain extracellular galactose concentrations, cells in an isogenic population display either a basal pathway activity (OFF) or an approximately 100-fold up-regulated activity (ON). Figure 5.1.4A and Figure 5.1.4B illustrate the stochastic transitions between these phenotypic states in single cells. Here a single progenitor cell initially in either the ON or the OFF state gives rise to a diverse population of both ON and OFF cells due to stochastic transitions between the two phenotypes.

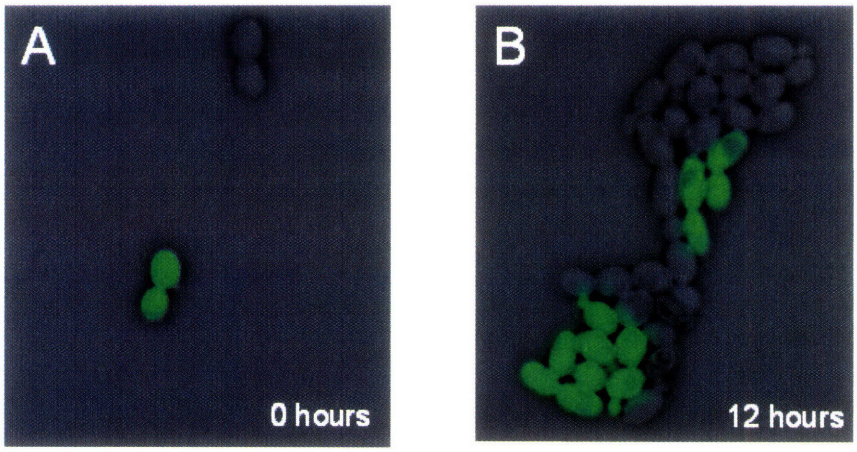


Figure 5.1.4 A single cell initially in the OFF state gives rise to both ON (green) and OFF (black) cells due to stochastic transitions between states. Similarly an ON cell (green) gives rise to a mixed population.

In our system, *GAL2* and *GAL3* remain under the control of their endogenous promoters, while the expression of *GAL80* is placed under the control of the *TET* promoter, which is induced using doxycycline. Because of this modification, this strain has externally tunable stochastic transition rates, r_{ON} and r_{OFF} (Figure 5.1.1), between the two phenotypic expression states. The rates are controlled by changing the extracellular concentration of galactose and doxycycline, with low (high) amounts of galactose and doxycyline leading to high (low) transition rates (Acar et al. 2005). This feature allows us to create populations of cells with both fast and slow switching rates.

To provide the two phenotypic states with two distinct growth rates we placed the endogenous *URA3* expression, necessary for uracil biosynthesis, under the sole control of the *GAL1* promoter. In this scenario, the two environments (E_1 and E_2) affect the growth rates of the two phenotypes in an antagonistic manner. Environment E_1 lacks

uracil and therefore favors the growth of ON cells to the disadvantage of the OFF cells. In contrast, E_2 contains both uracil and the small molecule 5-Fluoroorotic acid (5-FOA), which is converted into the toxic intermediate 5-Flurouracil in the presence of Ura3 protein, conferring a growth advantage to OFF cells (Boeke et al. 1984). Therefore, this experimental system provides us with a quantitative control over both the inter-phenotype switching rates and the growth rates of both phenotypes in the two different environments.

5.1.4 Switching Rate Measurement and Characterization

We first create two distinct populations of cells: one with fast and one with slow phenotypic transition rates, r_{ON} and r_{OFF}, by externally controlling the concentration of galactose and doxycycline present in the extracellular medium. Figure 5.1.5 experimentally demonstrates the behavior of these two populations in non-selective media.

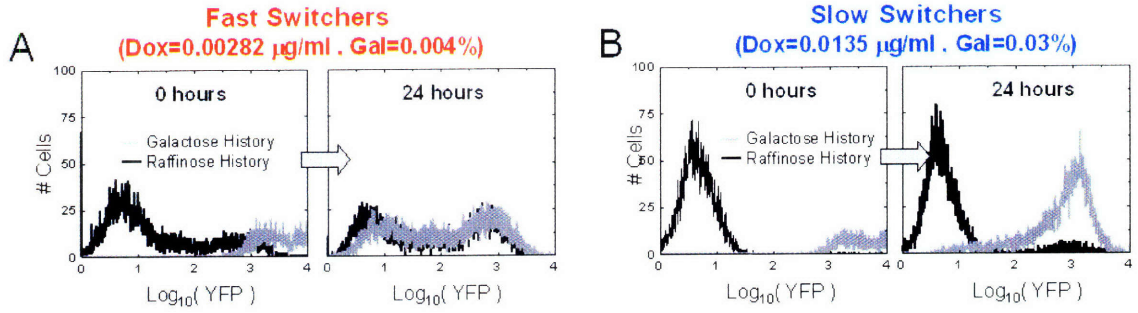


Figure 5.1.5 (A) YFP fluorescence distribution of fast switching cells that have been grown in non-selective media to obtain the t = 0 distributions (methods). After an additional 24 hours in non-selective media (with 0.004% galactose and 0.00282 μg/ml doxycycline corresponding to fast switchers), the distributions show very little history dependence indicating that cells are switching much faster than 1/24 hours^{-1}. **(B)** YFP fluorescence distribution of slow switching cells that have been grown in non-selective media to obtain the t = 0 distributions (methods). After an additional 24 hours in non-selective media (with 0.03% galactose and 0.0135 μg/ml doxycycline corresponding to slow switchers), the distributions show a large history dependence indicating that the switching rates are much slower.

Doxycycline increases the concentration of Gal80 biasing cells to the OFF state, while galactose sequesters Gal80 to the cytoplasm biasing cells to the ON state. Adding both compounds to the media causes transition rates between the two states to decrease accordingly. Due to these stochastic transitions between the two phenotypes, the t=0 hours expression distributions corresponding to fast (0.004% galactose, 0.00282 μg/ml doxycycline) or slow (0.03% galactose, 0.0135 μg/ml doxycycline) switching cells change at the end of the 24 hours growth period. By measuring the fraction of cells in each state at several points in time (Figure 5.1.6), we are able to estimate that the transition rates are roughly ten times greater for the fast switching cells, ($r_{ON} \approx 0.047, r_{OFF} \approx 0.035\,\text{hr}^{-1}$), than for the slow switching population, ($r_{ON} \approx 0.004, r_{OFF} \approx 0.007\,\text{hr}^{-1}$).

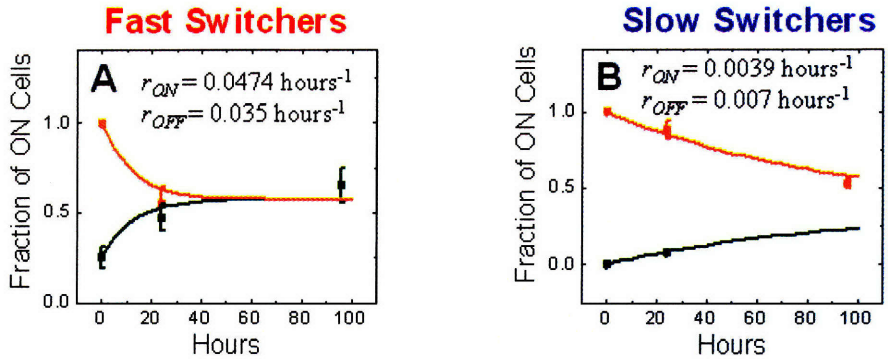


Figure 5.1.6 Determination of switching rates for fast and slow switchers. **(A)** The fraction of ON cells in a fast switching population as a function of time in non-selective media for the galactose (orange squares) and raffinose (green squares) history cells. **(B)** Fraction of ON cells for the slow switching population. To estimate the fraction at 96 hours the steady state distribution of cells (taken from Figure 5.1.7C-D) were determined. By fitting the data with the function:

$$f_{ON}(t) = \frac{r_{ON}}{r_{OFF} + r_{ON}} + \left(f_{ON}(t=0) - \frac{r_{ON}}{r_{OFF} + r_{ON}} \right) e^{-(r_{ON} + r_{OFF})t}$$

and minimizing the χ^2 cost function, we determine that the inter-phenotype switching rates for the fast switchers are r_{ON} = (0.0474 ± 0.026) hours^{-1} and r_{OFF} = (0.035 ± 0.020) hours^{-1}. The inter-phenotype switching rates for the slow switchers are: r_{ON} = (0.0039 ± 0.0002) hours^{-1} and r_{OFF} = (0.007 ± 0.0007) hours^{-1}. Solid lines represent the best fits to the data.

Figure 5.1.7 experimentally demonstrates the selection pressures imposed by each environment on the two phenotypes, ON and OFF. In the absence of any selection, both OFF and ON cells grow at very similar rates. In this case, for both the fast and slow switchers we observe steady-state bimodal distributions of network activity with roughly equal numbers of cells in either phenotypic state (Figure 5.1.7, top panels). However, when cells are grown for 4 days in either environment E_1 or E_2, the population significantly enriches in ON or OFF cells, respectively (Figure 5.1.7, bottom panels). The fast switchers display a more diverse distribution of expression values compared to the slow switchers, because the fast switchers transition more frequently to the unfit phenotype.

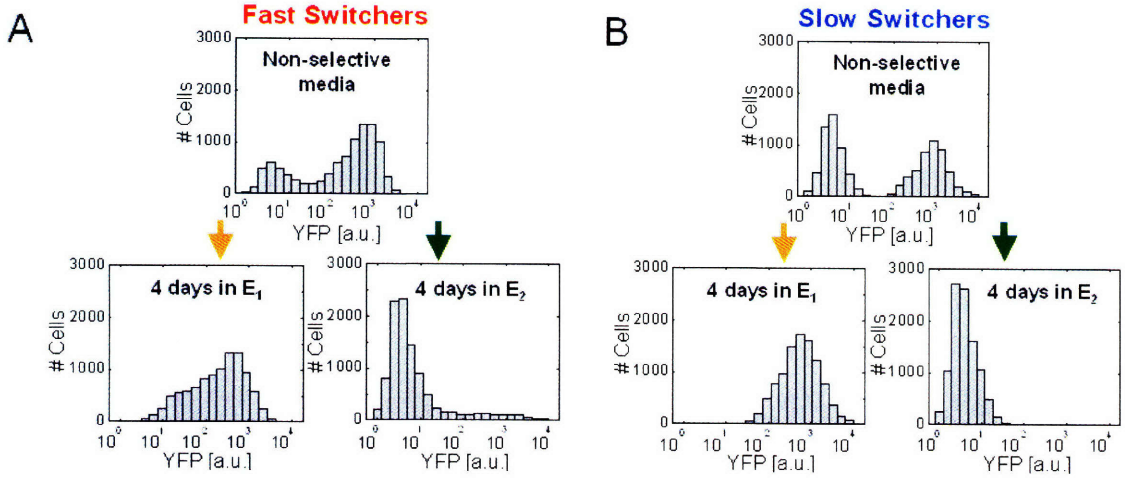


Figure 5.1.7 (A) Fast switching cells grown in non-selective media display a bimodal distribution. When cells are grown in E_1 (E_2), the interaction of the *URA*3 gene with the environment causes ON (OFF) cells to proliferate. **(B)** Similar selection is observed for slow switching cells, however fewer unfit cells are observed compared to the fast switchers.

We use a first order kinetic model to estimate the amount of time required for a cell to transition from one state to the other once it has decided to make the switch. Here we assume that ON cells produce YFP and Ura3 protein at a constant rate while OFF cells produce none. We assume that removal of these proteins is dominated by the dilution rate due to growth. Using this model we calculate that it requires approximately 1.3 hours for an average OFF cell to enter into the ON phenotype peak, while it requires approximately 6.5 hours for an ON cell to dilute enough YFP to be indistinguishable from an OFF cell. Both of these time-scales are smaller than the switching rates demonstrating that our approximation of the population as two distinct phenotypes is reasonable.

5.1.5 Turbidostat Setup

Next, we explored how this increased diversity impacts the growth rate of the population. To measure growth rates over long periods we utilized turbidostats (Bryson et al. 1952) enabling accurate and automatic on-line measurements of population growth rates (Figure 5.1.8). Cells are maintained at a constant optical density in liquid culture by continuous measurement of the culture's relative absorption coefficient (Figure 5.1.8B, upper panel – blue line). When the cellular density exceeds a pre-set threshold (Figure 5.1.8B, upper panel - red line) a pump is activated (Figure 5.1.8B, middle panel) which dilutes the culture back below the threshold. By measuring the amount of media needed to maintain the culture in this way, we can accurately calculate the population's growth rate (Figure 5.1.8B, lower panel) over time.

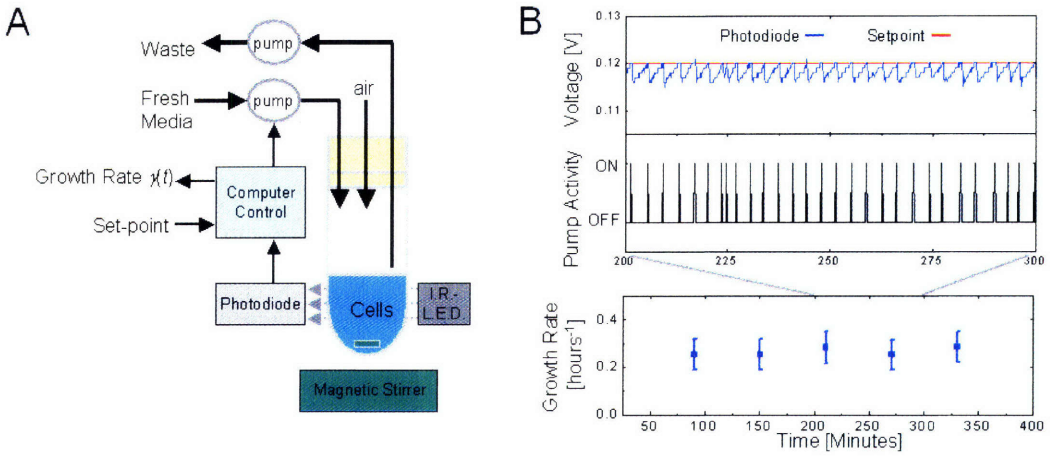


Figure 5.1.8 (A) Schematic of the custom-made turbidostat setup used for all growth rate measurements. An infrared LED (dark gray) and photodiode (light gray) pair were used to continuously measure the relative optical density of the culture. A peristaltic pump intermittently provided fresh media to dilute the population while second continuously operating peristaltic pump coupled to a pickup tube kept the culture volume fixed. **(B)** Whenever the photodiode voltage (blue line) went above a pre-set threshold (red line) the pump was activated (black line) to provide fresh media. The pumping rate was then used to calculate the dynamic population growth rate $\gamma(t)$ (blue circles).

Growth rates are calculated by recording the fraction of time the pump is actively providing fresh media during each hour-long interval and converting this into a pump flow rate by multiplying by the pump's maximal possible flow rate. This raw pump activity

is then converted into a growth rate by normalizing with the volume of the turbidostat culture using the formula, $\gamma(t) = \frac{p(t)}{V_{culture}}$, where $p(t)$ represents the measured pumping rate in ml/hr and $V_{culture}$ represents the culture volume in ml.

5.1.6 Growth Rate Dynamics

To quantify the relationship between population diversity and growth rate, we measured the growth rate dynamics as the population transitions from one environment to the other. We first prepared separate cultures of fast and slow switchers by allowing them to reach a steady-state YFP expression distribution in environments E_1 and E_2 (Figure 5.1.7, bottom panels). Subsequently, cells were transferred to the other environment (at time t = 0 as defined in Figure 5.1.9A-B) during which the population growth rate γ was monitored (Figure 5.1.9A-B). Figure 5.1.9A demonstrates the growth rate dynamics when the environment switches from E_2 to E_1. The growth rates of both the fast and slow switchers initially decreased rapidly within the first few hours since most cells were initially in the unfit (OFF) state and were unable to produce their own uracil. After a lag period of 3-5 hours, both fast and slow switching populations began to increase their growth rates, and reached a steady-state growth rate after about 18 hours (Figure 5.1.9A).

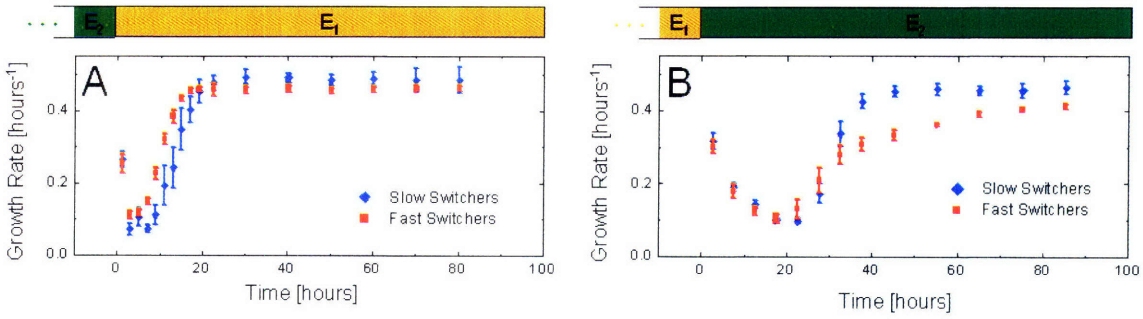


Figure 5.1.9 (A), Growth rates corresponding to cells prepared in an E_2 history and transferred to E_1 at t = 0 show a transition period and a steady state region. Fast switching cells (red line) recover from the effect of environment change faster than show switching cells (blue line) but have a lower steady-state growth rate. **(B),** Growth rates for cells prepared in an E_1 history and transferred to E_2 at t = 0. The red line corresponds to fast switchers and the blue line to slow switchers.

Although the growth rates for fast and slow switchers in Figure 5.1.9A-B show qualitatively similar dynamics, the quantitative differences between them illustrate the effects of diversity on the transient and steady-state growth rates. First, the fast switchers reached their steady-state about 4 hours sooner than the slow switchers. This is because they enter the new environment with a larger population of fit (ON) cells and

also because more members of the initially large fraction of unfit (OFF) cells can switch to the fit (ON) state during the first few hours. During this initial phase the fast switching strategy is more competitive than the slow switching strategy. However, once the growth rates have reached steady-state, the slow switching strategy becomes more competitive because the slow switchers less frequently transition to the unfit (OFF) phenotype. This results in a systematically larger steady-state growth rate compared to the fast switching strategy. We observe similar growth rate dynamics when cells transition from E_1 to E_2 (Figure 5.1.9B). The transient dynamics is overall slower in E_2 possibly due to the slow synthesis of the toxic intermediate 5-Flurouracil. After the adaptation period in E_2,both fast and slow switchers show similar growth rates to what was observed in environment E_1 (Figure 5.1.9A). Taken together, this shows that after transitioning to a new environment, fast switchers have higher growth rates during the transition to steady-state but lower growth rates than the slow switchers during the steady-state.

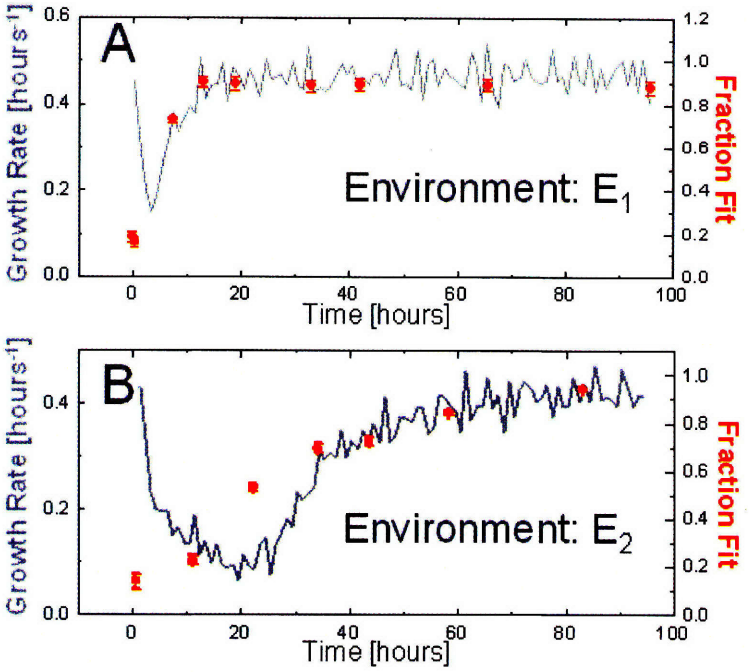


Figure 5.1.10 Dynamic measurement of cellular expression levels in the two environments. **(A)** Fast switching cells growing in E_2 were transferred to E_1 at t = 0 and the fraction of ON and OFF cells were measured at several points in time (red symbols). The increase in the dynamic growth rate (gray line) coincides with the increase in the fraction of fit cells (ON). **(B)** Fast switching cells growing in E_1 were transferred to E_2 at t = 0. The increase in growth rate coincides with a significant increase in the fraction of fit cells.

5.1.7 Simple Model

To further demonstrate that the observed behavior is due to the differences in switching rates between fast and slow switchers, we built a quantitative model composed of ordinary differential equations. The model quantifies the number of cells in each of the two discrete phenotypes, ON and OFF, which both grow at different rates, $\gamma_{ON}(t)$ and $\gamma_{OFF}(t)$ respectively (Figure 5.1.1). These growth rates reflect the fitness advantage or disadvantage conferred by the surrounding medium by quantifying both the steady state growth rate of each phenotype as well as the time required for each phenotype to reach steady state. In addition, cells also transition from the OFF to ON (ON to OFF) phenotypes at a constant rate r_{ON} (r_{OFF}) shown in Figure 5.1.6.

The model consists of two differential equations which characterize the dynamics of the number of cells in the ON and OFF states, N_{ON} and N_{OFF} respectively:

$$\frac{d}{dt}N_{ON} = \gamma_{ON}N_{ON} - r_{OFF}N_{ON} + r_{ON}N_{OFF}$$
$$\frac{d}{dt}N_{OFF} = \gamma_{OFF}N_{OFF} + r_{OFF}N_{ON} - r_{ON}N_{OFF}$$

The population growth rate is given by:

$$\gamma(t) = \frac{\frac{d}{dt}N_{ON} + \frac{d}{dt}N_{OFF}}{N_{ON} + N_{OFF}}$$

The parameter r_{ON} (r_{OFF}) characterizes the rate of transitions to the ON (OFF) state. The values γ_{ON} and γ_{OFF} are the instantaneous growth rates of the ON and OFF phenotypes and depend on which environment the cells are in. In E_1 the rates are given by $\gamma_{ON}(t) = \gamma_1(1-e^{-r_1 t})$ and $\gamma_{OFF}(t) = \gamma_1(e^{-d_1 t})$, while in E_2 the rates are given by $\gamma_{OFF}(t) = \gamma_2(1-e^{-r_2 t})$ and $\gamma_{ON}(t) = \gamma_2(e^{-d_2 t})$ (Figure 5.1.11). In these equations t represents the amount of time since the last environmental transition, and the constants γ_1 and γ_2 represent the steady-state growth rates achieved by the fit phenotype in each environment. The constants r_1 and r_2 represent the amount of time it takes for cells to recover from a transition from their unfit to their fit environment. Similarly d_1 and d_2 represent the time it takes for unfit cells to cease growing after a transition from their fit environment to their unfit environment.

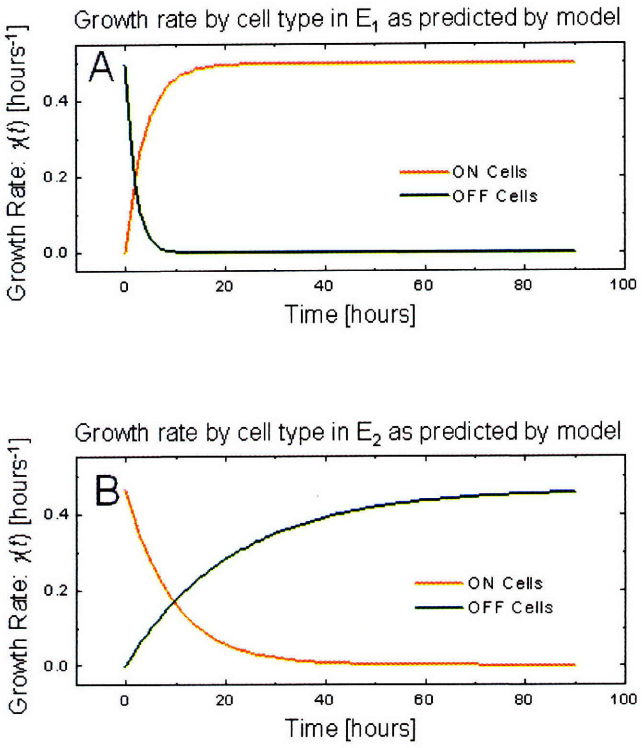


Figure 5.1.11 Modeling cellular growth in fluctuating environments. **(A)** In environment E_1, OFF cells (green) exponentially cease their growth rate due to the lack of intracellular uracil. Similarly, ON cells (orange) are assumed to need some time to recover from the previous environment (E_2 with 5-FOA) and begin growing at their maximal rate. **(B)** In E_2, the opposite scenario occurs where ON cells (orange) decrease their growth rate with time as the effects of 5-FOA build up in the cell, and OFF cells (green) recover their maximal growth rate on a slightly longer timescale.

The parameters r_1, r_2, d_1, and d_2 (Table 5) are separately fit to the data in Figure 5.1.9A-B by minimizing the χ^2 cost between the prediction and the measured growth rate values. The model fit is depicted by the red and blue lines in Figure 5.1.12A-B. The only difference between fast and slow switching cells in the model is the difference between transition rates (r_{ON}, r_{OFF}), indicating that the observed growth rate differences between fast and slow switching cells can be accounted for by switching rates alone.

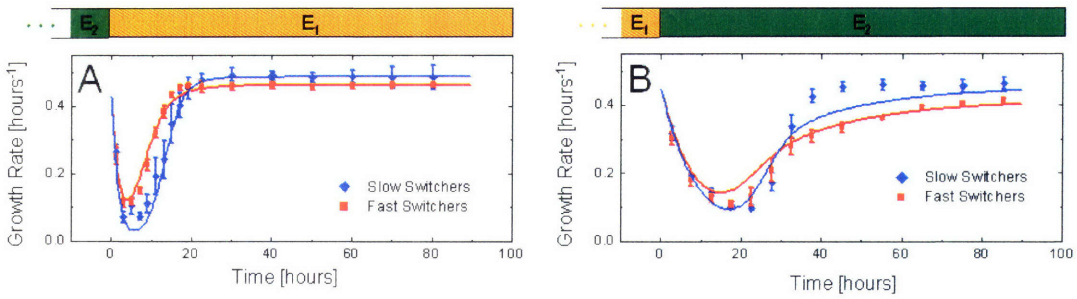


Figure 5.1.12 Model fits to the single environment swap data. Solid lines are generated by the model. **(A)** Growth rates corresponding to cells prepared in an E_2 history and transferred to E_1 at $t = 0$ show a transition period and a steady state region. Fast switching cells (red line) recover from the effect of environment change faster than show switching cells (blue line) but have a lower steady-state growth rate. **(B)** Growth rates for cells prepared in an E_1 history and transferred to E_2 at $t = 0$. The red line corresponds to fast switchers and the blue line to slow switchers.

Table 5. Model parameters

Parameter	Environment	Value	Source
r_{ON} (slow)	1 & 2	0.0039 ± 0.0002 hours^{-1}	Figure 5.1.6
r_{OFF} (slow)	1 & 2	0.007 ± 0.0007 hours^{-1}	Figure 5.1.6
r_{ON} (fast)	1 & 2	0.0474 ± 0.026 hours^{-1}	Figure 5.1.6
r_{OFF} (fast)	1 & 2	0.035 ± 0.02 hours^{-1}	Figure 5.1.6
γ_1	1	0.50 ± 0.01 hours^{-1}	Figure 5.1.9
γ_2	2	0.46 ± 0.01 hours^{-1}	Figure 5.1.9
d_1	1	0.62 hours^{-1}	Fit to Figure 5.1.9
d_2	2	0.17 hours^{-1}	Fit to Figure 5.1.9
r_1	1	0.12 hours^{-1}	Fit to Figure 5.1.9
r_2	2	0.042 hours^{-1}	Fit to Figure 5.1.9

5.1.8 Prediction of Growth in Fluctuating Environments

Using this population dynamics model, we can now predict which of the two switching strategies is more beneficial in fluctuating environments. In the model, the environment periodically alternates between environment E_1 (with a duration T_1) and E_2 (with a duration T_2) as depicted in Figure 5.1.1A. The fitness Γ of the population is defined as the average population growth rate over one period. The resulting fitness is calculated as a function of T_1 and T_2 for the fast (Γ_{fast}) and slow (Γ_{slow}) switching strategies. For both strategies short environmental periods lead to decreased fitness levels, since more time is spent in the transient recovery stage of growth. However, the fitness difference between slow switching and fast switching populations can either be positive or negative depending on T_1 and T_2 (Figure 5.1.13A). This analysis shows that in rapidly changing environments fast switchers would out-compete the slow switchers, whereas for slowly varying environments the slow switchers would dominate. Figure 5.1.13B shows which of the two strategies is more beneficial given the times T_1 and T_2.

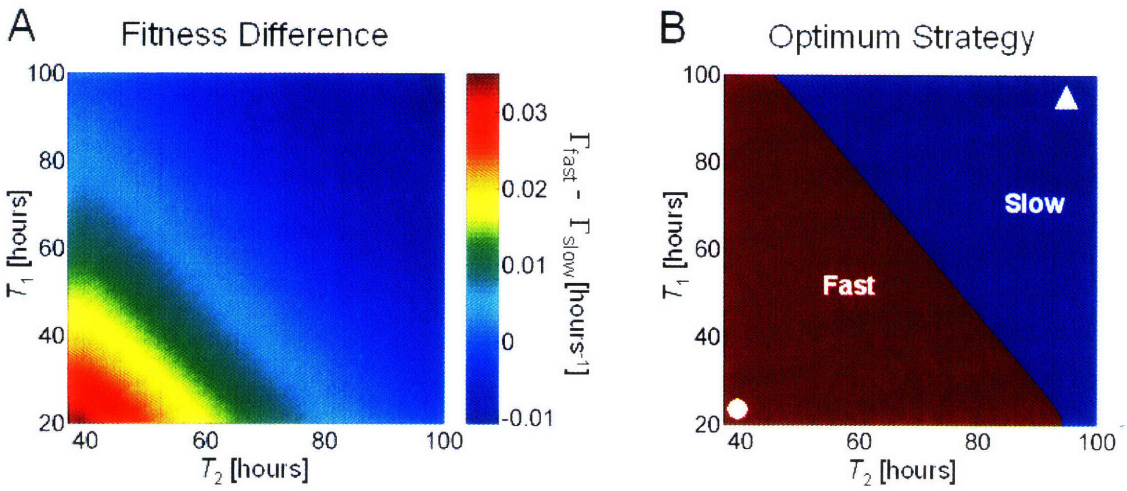


Figure 5.1.13 **(A)** Heat map showing the predicted fitness difference (mean growth rate) between fast and slow switching cells as a function of the environmental period (T_1, T_2). For short periods, the fast switchers display higher mean growth rates. On the other hand, for longer periods slow switchers show a fitness advantage over the fast switchers. **(B)** Phase diagram demonstrating regions where fast or slow switchers are predicted to be more fit. Two points were chosen in each part of the phase diagram to be further explored. The values of these points correspond to T_1 = 20 hours, T_2 = 37 hours (circle) and T_1 = 96 hours, T_2 = 96 hours (triangle).

To further demonstrate the effects of switching rate on the fitness of a population, we theoretically examine the opposite scenario, whereby the environmental periods are kept constant and the switching rates are modified. We examine two points in the phase

space shown in Figure 5.1.13 noted by a circle (short period) and triangle (long period). Both environmental periods show an optimal growth rate as noted by darkest shade of red in the heat map (Figure 5.1.14). As mentioned in section 5.1.2 this is because very large switching rates keep a large number of cells in an unfit state, while very slow switching rates are heavily impacted by environmental changes since very few cells are in the fit phenotype. The major difference between these calculations is that in short periods, the optimal switching rates (r_{ON} and r_{OFF}) are much faster than they are in the long period environment. This suggests that cells need to tune their switching rates to the rate of environmental changes with long periods requiring slow switching rates, while short periods require fast switching rates.

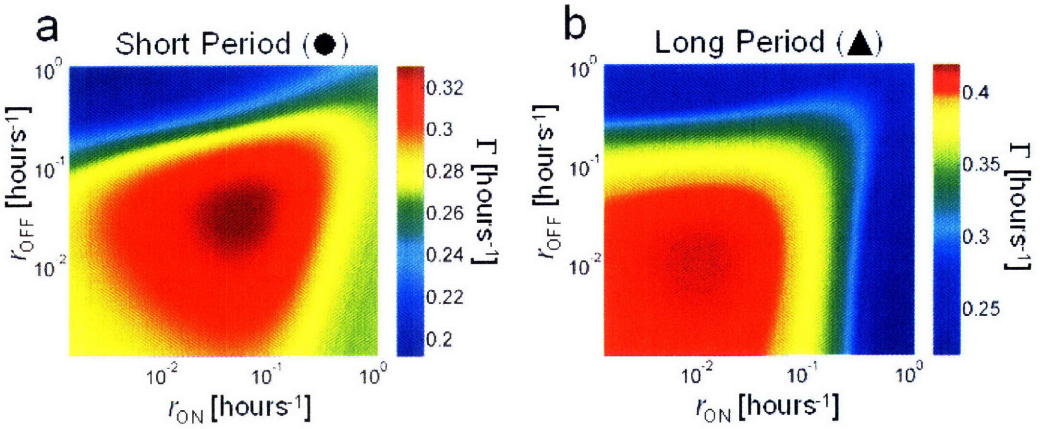


Figure 5.1.14 Growth rates calculated as a function of switching rate in **(A)** rapidly and **(B)** slowly varying environments. The optimal switching rate is faster in short period environments and slower in long period environments.

5.1.9 Growth in Fluctuating Environments

To experimentally test these predictions, we choose two points in the (T_1, T_2) space to measure fitness: T_1 = 20 hours, T_2 = 37 hours (circle, Figure 5.1.13B) and T_1 = 96 hours, T_2 = 96 hours (triangle, Figure 5.1.13B). As with the previous experiments, we grow the cells to steady state (4 days) in E_2 then transition them to E_1 at t = 0. After T_1 hours, the extracellular environment was changed again to E_2 for an additional T_2 hours. The dynamically measured growth rates for these two fluctuating environments are shown for both fast and slow switching populations in Figure 5.1.15A-B along with the population growth rates predicted by the model (Figure 5.1.15A-B, solid lines). As the phase diagram suggests, the fast switching cells out-compete the slow switchers in the rapidly changing environment (Figure 5.1.15A), while the situation is reversed in the slowly changing environment (Figure 5.1.15B). This is shown as a fitness difference between the fast and slow switchers in Figure 5.1.15E-F as quantified by the average growth rate. Here, in the rapidly changing environment the difference between growth rates is 20 percent, while in the slowly changing environment it is 6 percent. Due to exponential growth these differences greatly magnify the relative competitiveness over even moderate time-scales. This out-competition is made dramatically obvious when comparing the number of cells that would have been produced in a non-limiting environment at the end of each run (Figure 5.1.15C-D).

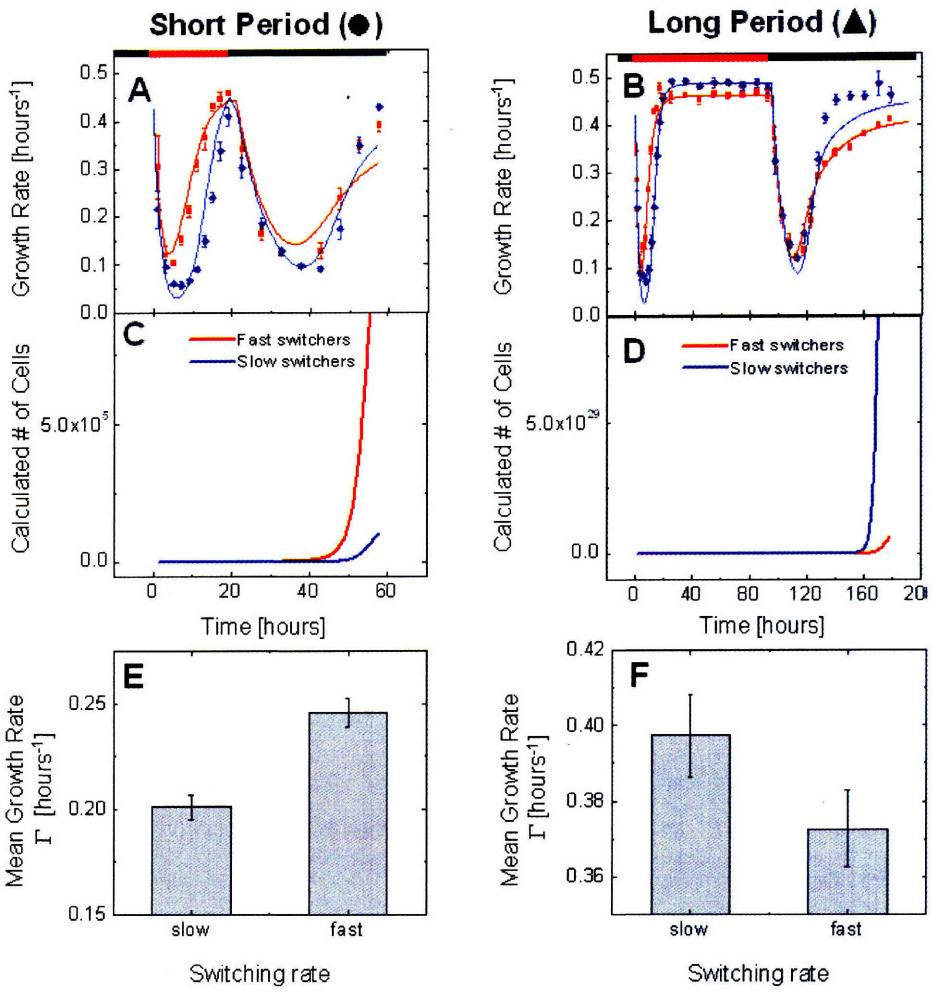


Figure 5.1.15 Testing the model predictions: growth dynamics in fluctuating environments with short and long periods. Growth rates for fast (red) and slow (blue) switchers grown for **(A)**, short environmental periods (20 hours in E_1, 37 hours in E_2) and for **(B)**, long environmental periods (96 hours in E_1, 96 hours in E_2) are compared to the growth rates predicted by the model (solid lines). The calculated number of cells is shown for **(C)**, the short period and **(D)**, the long period environment, highlighting the relative fitness advantage of fast and slow switchers respectively. The experimentally measured fitness (mean growth rate) **(E)**, in the short period and **(F)**, the long period environment. Error bars represent a 2.8% error estimated from the standard deviation of the growth rate measurements.

5.1.10 Discussion

Our data suggest that bet-hedging constitutes a simple, yet effective, survival strategy to cope with fluctuating environments. Following this strategy, an isogenic population improves its fitness by promoting phenotypic diversity so that, at any given time, some of its members are prepared for an unforeseen environmental fluctuation. The diversity is introduced naturally through the stochastic nature of biochemical processes allowing isogenic populations to mitigate risk by not putting "all of their eggs in one basket". A familiar analogue of this concept also occurs in finance, where diversification allows one to minimize the risk on a portfolio of investments by spreading the risk among many diverse assets.

Recent work suggests that cell-to-cell variability can have a large impact on the fitness of a population during times of adversity (Blake et al. 2006; Bishop et al. 2007). Here we show that it is the frequency of the environmental fluctuations that constrains the inter-phenotype transition rates. In particular, a population can enhance its chances of survival in fluctuating environments by tuning the phenotypic switching rates with respect to the durations of environmental exposures. In this 'resonant' condition bet-hedging provides an effective survival strategy by blindly anticipating environmental changes. This strategy could be utilized by cellular populations that are lacking dedicated signal transduction machinery for particular extracellular signals or when it is crucial for a population to act at a faster timescale than is possible by signal transduction.

5.1.11 Methods

Measurement of Inter-phenotypic Switching Rates

Cells were grown using non-selective synthetic dropout media (including 0.02 mg/ml uracil but lacking 5-FOA) in 2% raffinose. After overnight growth, cells were grown in 5 ml cultures for 30 hours in 2% raffinose ("raffinose history") and 2% raffinose and 2% galactose ("galactose history"). Subsequently, the raffinose and galactose history cells were separately grown for another 24 hours in media having 0.004% galactose and 0.00282 μg/ml doxycycline (fast switchers) and 0.03% galactose and 0.0135 μg/ml doxycycline (slow switchers). Culture volume during this period was 10 ml and the OD did not exceed 0.315. The expression distributions displayed were determined by flow cytometer (FACScan; Becton Dickinson).

Strain Construction and Growth Conditions

In order to couple the expression of the *URA3* gene to the activity of the galactose pathway, the diploid strain MA0188 (Acar et al. 2005) was modified. The promoter of the *URA3* gene was replaced by the *GAL1* promoter using homologous recombination of transformed PCR product. The *Candida albicans HIS5* gene was used as a marker. Integrations were verified by PCR.

Cultures were grown in synthetic dropout media with the appropriate amino-acid supplement and 2% raffinose as a carbon source. Prior to their turbidostat environments, cells were grown for 4 days in liquid culture in a shaker at 30°C. The culture volume was 5 ml for the initial overnight growth and 10-15 ml afterwards. These 'pre-turbidostat' media contained uracil (at a final concentration of 0.02 mg/ml) and 5-FOA (at a final concentration of 0.19 mg/ml). During this period, cells were prevented from reaching stationary growth phase by serial dilution. Next, cells were washed with their prospective 'in-turbidostat' media (lacking uracil and 5-FOA) and were transferred to the turbidostat. The turbidostat maintains constant optical density levels (OD) corresponding to exponential growth ($0.05 < OD_{600} < 0.2$). When switching to environment E_2, uracil (0.02 mg/ml) and 5-FOA (0.19 mg/ml) were added to the turbidostat media. The turbidostat temperature was maintained at 30 °C. Fast and slow switching phenotypes were obtained by supplementing the synthetic dropout media with galactose and doxycycline.

For the fast switching phenotype, 0.004% galactose and 0.00282 μg/ml doxycycline were used. For the slow switching phenotype, 0.03% galactose and 0.0135 μg/ml doxycycline were used. The expression distributions were determined by flow cytometer (FACScan; Becton Dickinson).

Turbidostat Measurements

All growth rate measurements were made using a custom made turbidostat (Figure 5.1.8). Cells were maintained at 10-15 ml volumes in test tubes while magnetic stir bars kept the cultures well mixed. An infrared LED/Photodiode pair (940 nm) was used to measure the relative OD through a D/A converter and custom Labview software. A peristaltic pump provided fresh media to dilute the population whenever the OD went above a critical threshold value. A second continuously operating peristaltic pump coupled to a pickup tube kept the culture volume fixed.

6 Perspectives

As demonstrated in Chapter 3.1, we were able to obtain a wealth of information about a simple biological system by studying how the network responded to dynamic inputs. Using periodic stimuli, we learned how the system responds to signals over a wide range of timescales. These frequency response data suggests that the system responds as a band-pass filter, filtering out both very rapid and very slow fluctuations. Using some simple analysis, a model was recovered from this data that is predictive in nature. Although many network structures are consistent with the observed dynamics, a simple observation of the underlying biology was enough to guide the analysis and produce a single network topology consistent with the data and known biology.

In general, a network under study may have much more complicated dynamics than those of the osmotic-response pathway, and it would be important for future work to address scaling these techniques. It is important to note that in this thesis, a single readout of the activity of the signal transduction cascade was measured. While this was sufficient information to deduce the network topology because there were only two dynamic variables and enough is known about the network to rule out alternative topologies. In general, however, these dynamic transfer functions could come from any one of an infinite number of suitable network topologies. Without further information, the underlying topology could not be as easily determined from the dynamic transfer function. For this reason, it would be important to further develop these techniques for multiple inputs and multiple output systems. As the number of dynamic variables needed to describe the system increases, the complexity of mechanistic network topologies that are able to give rise to the observed dynamics increases dramatically. This makes it unlikely that a few simplifying assumptions alone will be enough to inform a network topology. Rather than using a single input single output approach as applied here, it may be possible to simply expand these techniques using a multiple input, multiple output approach to gain the necessary data and network topologies needed. This type of analysis might be facilitated by multichannel microfluidics, which would allow application of various time dependent input signals independently, while simultaneously monitoring multiple output channels through fluorescent proteins or other techniques.

It may be a worthwhile endeavor to attempt to combine these types of dynamic systems analysis with other network inference techniques. Many network inference techniques have been developed to infer which nodes in a system influence other nodes

(Margolin et al. 2007; Price et al. 2007). Typically, however, these techniques rely heavily or entirely on steady state experimental data. While this may provide a set of interactions with varying degrees of confidence, the magnitudes and dynamics of these interactions may easily be lost due to the static nature of the experimental data. On the other hand, signals and systems analysis excels at developing dynamic phenomenological dynamic relationships between elements. Future work that addresses the connections between the static and dynamic system identification procedures might allow for more accurate inference of dynamic networks.

A few techniques to infer network connectivity have already been developed that rely on observing propagation information traveling within a network (Vance et al. 2002), usually in response to step inputs applied systematically to various nodes. Here, information about the time of arrival of various signals indicates causal relationships between connected components. In principle, this is similar to the dynamic systems methodology; however, it would be interesting to extend these analytical techniques to include period perturbations. One advantage that periodic perturbations have over step function perturbations lies in the frequency spectrum. In principle, a step function is a combination of low and high frequencies, and therefore contains all the different perturbation frequencies in a single input function. Unlike a delta function input, which has a uniform (flat) frequency spectrum, the step function input has higher amplitude in the low frequencies and lower amplitude in the high frequencies. This means that information about the fast responses and interactions has a tendency to be washed out against information coming from the slower timescale reactions. By applying sine (or square) wave pulses the amplitude of each frequency can be controlled individually, and the high frequency dynamics can be given equal weight to the low frequency dynamics.

Biological systems are inherently non-linear; for example, in a cell there will always be a maximum and minimum response regardless of how high or low the input signal magnitude is. In this work, we dealt with this fact by adding in a rectification element, but in general, non-linearities may complicate analysis of more complex networks. More work will likely be necessary before this type of analysis has the same amount of general applicability that it does in mechanical and electrical systems.

It is also important to note that these technique used in Chapter 3.1 will only identify the core network topology; that is, the simplest set of interactions necessary to accurately describe the dynamics involved. For example, this model has two internal state variables, while a full description of the network would involve many more

independent variables characterizing each individual intermediate in the signaling and response pathways. This core topology does not always make itself apparent from a very detailed set of interactions, and is often useful in that it may identify the key interactions or pathways in the system that are central in determining the observed dynamic behavior. Therefore, the signals and systems approach may often be the most useful when a system is very well characterized in terms of network structure but with relatively little known about the quantitative and temporal aspects of its function.

Further, these techniques might be useful for studying and understanding the molecular basis for disease. In many diseases, the healthy function of signaling pathways is modified in such a way that normal extracellular signals are misinterpreted due to errors in the signal processing cascades. These signals are often time varying in nature due to the desire of cells to respond to changes in the body. A standard method to treating these diseases involves looking for misregulated proteins in the pathway, and then applying theraputic agents that target these proteins attempting to return them to normally functioning levels. While this approach is attractive, it is not always clear that the protein under study is a *cause of* and not a *response to* the diseased state. Therefore blindly changing the function of a protein or pathway that is reacting to an error in another pathway may actually cause worsening of the disease state (Araujo et al. 2007). In these cases, it may be interesting use dynamic signal propagation as a metric to look for perturbations in signaling pathways that return a diseased pathway to normal dynamic signaling behavior.

Another interesting direction for future work may lie in exploring the connections between signal processing and signal noise discussed in this work. For example, while the directed intracellular signals studied in Chapter 4.3 seem to "confuse cells" and obscure the correct directionality of the extracellular signal, a recent study which builds on the work presented in this chapter suggests that it may actually do the opposite. (Andrews et al. 2007) has constructed an information theoretic approach to mapping perceived extracellular cues with intracellular decision-making processes. Here they find that in the presence of extracellular noise, it is beneficial for the cells to maintain an internal signal that points in the direction that they have previously determined the signal. This provides a sort of sensory inertia that buffers rapid noisy fluctuations in the extracellular signal. Similarly, pioneering work with *e. coli* chemotaxis has shown that the filtering properties of the network may have been optimized to provide the greatest amount of directional sensitivity given an inherently noisy extracellular signal (Andrews

et al. 2006). It might therefore be worthwhile to continue these studies into gene regulatory and signal transduction network topology, as this might reveal previously unknown connections between network structure, intracellular noise, signal noise and signal fidelity.

7 Works Cited

Acar, M., A. Becskei and A. van Oudenaarden (2005). "Enhancement of cellular memory by reducing stochastic transitions." Nature **435**(7039): 228-32.

Albertyn, J., S. Hohmann, J. M. Thevelein and B. A. Prior (1994). "GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway." Mol Cell Biol **14**(6): 4135-44.

Alepuz, P. M., E. de Nadal, M. Zapater, G. Ammerer and F. Posas (2003). "Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II." Embo J **22**(10): 2433-42.

Alon, U. (2007). "Network motifs: theory and experimental approaches." Nat Rev Genet **8**(6): 450-61.

Andrews, B. W. and P. A. Iglesias (2007). "An information-theoretic characterization of the optimal gradient sensing response of cells." PLoS Comput Biol **3**(8): e153.

Andrews, B. W., T. M. Yi and P. A. Iglesias (2006). "Optimal noise filtering in the chemotactic response of Escherichia coli." PLoS Comput Biol **2**(11): e154.

Angeli, D., J. E. Ferrell, Jr. and E. D. Sontag (2004). "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems." Proc Natl Acad Sci U S A **101**(7): 1822-7.

Araujo, R. P., L. A. Liotta and E. F. Petricoin (2007). "Proteins, drug targets and the mechanisms they control: the simple truth about complex networks." Nat Rev Drug Discov **6**(11): 871-80.

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**(4): 1633-48.

Arrieumerlou, C. and T. Meyer (2005). "A local coupling model and compass parameter for eukaryotic chemotaxis." Dev Cell **8**(2): 215-27.

Axelrod, D. E., Y. Gusev and T. Kuczek (1993). "Persistence of cell cycle times over many generations as determined by heritability of colony sizes of ras oncogene-transformed and non-transformed cells." Cell Prolif **26**(3): 235-49.

Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik and S. Leibler (2004). "Bacterial persistence as a phenotypic switch." Science **305**(5690): 1622-5.

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." Nat Genet **38**(6): 636-43.

Becskei, A., B. B. Kaufmann and A. van Oudenaarden (2005). "Contributions of low molecule number and chromosomal positioning to stochastic gene expression." Nat Genet **37**(9): 937-44.

Becskei, A., B. Seraphin and L. Serrano (2001). "Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion." Embo J **20**(10): 2528-35.

Benzer, S. (1953). "Induced synthesis of enzymes in bacteria analyzed at the cellular level." Biochim Biophys Acta **11**(3): 383-95.

Berg, H. C. and E. M. Purcell (1977). "Physics of chemoreception." Biophys J **20**(2): 193-219.

Bhat, P. J. and J. E. Hopper (1992). "Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast GAL/MEL regulon." Mol Cell Biol **12**(6): 2701-7.

Bialek, W. and S. Setayeshgar (2005). "Physical limits to biochemical signaling." Proc Natl Acad Sci U S A **102**(29): 10040-5.
Bishop, A. L., F. A. Rab, E. R. Sumner and S. V. Avery (2007). "Phenotypic heterogeneity can enhance rare-cell survival in 'stress-sensitive' yeast populations." Mol Microbiol **63**(2): 507-20.
Blake, W. J., G. Balazsi, 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." Mol Cell **24**(6): 853-65.
Blake, W. J., K. A. M, C. R. Cantor and J. J. Collins (2003). "Noise in eukaryotic gene expression." Nature **422**(6932): 633-7.
Block, S. M., J. E. Segall and H. C. Berg (1982). "Impulse responses in bacterial chemotaxis." Cell **31**(1): 215-26.
Boeke, J. D., F. LaCroute and G. R. Fink (1984). "A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance." Mol Gen Genet **197**(2): 345-6.
Bryson, V. and W. Szybalski (1952). "Microbial selection." Science **115**(3003): 45-51.
Burdon, R. H. (1986). "Heat shock and the heat shock proteins." Biochem J **240**(2): 313-24.
Cai, L., N. Friedman and X. S. Xie (2006). "Stochastic protein expression in individual cells at the single molecule level." Nature **440**(7082): 358-62.
Chung, J. D. and G. Stephanopoulos (1996). "On physiological multiplicity and population heterogeneity of biological systems." Chemical Engineering Science **51**(9): 1509-1521.
Cohen, D. (1966). "Optimizing reproduction in a randomly varying environment." J Theor Biol **12**(1): 119-29.
Colaianne, J. J. and A. E. Bell (1970). "Sonless, a sex-ratio anomaly in Drosophila melanogaster resulting from a gene-cytoplasm interaction." Genetics **65**(4): 619-25.
Colman-Lerner, A., A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce and R. Brent (2005). "Regulated cell-to-cell variation in a cell-fate decision system." Nature **437**(7059): 699-706.
Corrie, J. E. T. and D. R. Trentham (1993). Caged nucleotides and neurotransmitters. Biological Applications of photochemical switches. H. Morrison. New York, Wiley: 243-305.
Dallon, J. C. and H. G. Othmer (1997). "A discrete cell model with adaptive signalling for aggregation of Dictyostelium discoideum." Philos Trans R Soc Lond B Biol Sci **352**(1351): 391-417.
Danese, P. N., L. A. Pratt, S. L. Dove and R. Kolter (2000). "The outer membrane protein, antigen 43, mediates cell-to-cell interactions within Escherichia coli biofilms." Mol Microbiol **37**(2): 424-32.
Delbruck, M. (1945). "The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages)." J Bacteriol **50**(2): 131-5.
Dhar, N. and J. D. McKinney (2007). "Microbial phenotypic heterogeneity and antibiotic tolerance." Curr Opin Microbiol **10**(1): 30-8.
Dubnau, D. and R. Losick (2006). "Bistability in bacteria." Mol Microbiol **61**(3): 564-72.
Dworkin, M. and K. H. Keller (1977). "Solubility and diffusion coefficient of adenosine 3':5'-monophosphate." J Biol Chem **252**(3): 864-5.
Elowitz, M. B. and S. Leibler (2000). "A synthetic oscillatory network of transcriptional regulators." Nature **403**(6767): 335-8.
Elowitz, M. B., A. J. Levine, E. D. Siggia and P. S. Swain (2002). "Stochastic gene expression in a single cell." Science **297**(5584): 1183-6.

Ferrigno, P., F. Posas, D. Koepp, H. Saito and P. A. Silver (1998). "Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1." Embo J **17**(19): 5606-14.
Funamoto, S., K. Milan, R. Meili and R. A. Firtel (2001). "Role of phosphatidylinositol 3' kinase and a downstream pleckstrin homology domain-containing protein in controlling chemotaxis in dictyostelium." J Cell Biol **153**(4): 795-810.
Gamba, A., A. de Candia, S. Di Talia, A. Coniglio, F. Bussolino and G. Serini (2005). "Diffusion-limited phase separation in eukaryotic chemotaxis." Proc Natl Acad Sci U S A **102**(47): 16927-32.
Gardner, T. S., C. R. Cantor and J. J. Collins (2000). "Construction of a genetic toggle switch in Escherichia coli." Nature **403**(6767): 339-42.
Gennemark, P., B. Nordlander, S. Hohmann and D. Wedelin (2006). "A simple mathematical model of adaptation to high osmolarity in yeast." In Silico Biol **6**(3): 0018.
Geva-Zatorsky, N., N. Rosenfeld, S. Itzkovitz, et al. (2006). "Oscillations and variability in the p53 system." Mol Syst Biol **2**: 2006 0033.
Gillespie, D. T. (1977). "Exact Stochastic Simulation of Coupled Chemical Reactions." The Journal of Physical Chemistry **81**(25): 2340-2361.
Golding, I., J. Paulsson, S. M. Zawilski and E. C. Cox (2005). "Real-time kinetics of gene activity in individual bacteria." Cell **123**(6): 1025-36.
Gottschling, D. E., O. M. Aparicio, B. L. Billington and V. A. Zakian (1990). "Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription." Cell **63**(4): 751-62.
Grossman, A. D. (1995). "Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis." Annu Rev Genet **29**: 477-508.
Halme, A., S. Bumgarner, C. Styles and G. R. Fink (2004). "Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast." Cell **116**(3): 405-15.
Hasty, J., D. McMillen and J. J. Collins (2002). "Engineered gene circuits." Nature **420**(6912): 224-30.
Hasty, J., J. Pradines, M. Dolnik and J. J. Collins (2000). "Noise-based switches and amplifiers for gene expression." Proc Natl Acad Sci U S A **97**(5): 2075-80.
Hawkins, K. M. and C. D. Smolke (2006). "The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae." J Biol Chem **281**(19): 13485-92.
Hernday, A., M. Krabbe, B. Braaten and D. Low (2002). "Self-perpetuating epigenetic pili switches in bacteria." Proc Natl Acad Sci U S A **99 Suppl 4**: 16470-6.
Hohmann, S. (2002). "Osmotic stress signaling and osmoadaptation in yeasts." Microbiol Mol Biol Rev **66**(2): 300-72.
Hooshangi, S., S. Thiberge and R. Weiss (2005). "Ultrasensitivity and noise propagation in a synthetic transcriptional cascade." Proc Natl Acad Sci U S A **102**(10): 3581-6.
Ideker, T., T. Galitski and L. Hood (2001). "A new approach to decoding life: systems biology." Annu Rev Genomics Hum Genet **2**: 343-72.
Ingalls, B. P. (2004). "A frequency domain approach to sensitivity analysis of biochemical networks." Journal of Physical Chemistry B **108**(3): 1143-1152.
Insall, R., A. Kuspa, P. J. Lilly, G. Shaulsky, L. R. Levin, W. F. Loomis and P. Devreotes (1994). "CRAC, a cytosolic protein containing a pleckstrin homology domain, is required for receptor and G protein-mediated activation of adenylyl cyclase in Dictyostelium." J Cell Biol **126**(6): 1537-45.

Isaacs, F. J., J. Hasty, C. R. Cantor and J. J. Collins (2003). "Prediction and measurement of an autoregulatory genetic module." Proc Natl Acad Sci U S A **100**(13): 7714-9.
Janetopoulos, C., L. Ma, P. N. Devreotes and P. A. Iglesias (2004). "Chemoattractant-induced phosphatidylinositol 3,4,5-trisphosphate accumulation is spatially amplified and adapts, independent of the actin cytoskeleton." Proc Natl Acad Sci U S A **101**(24): 8951-6.
Johnston, M., J. S. Flick and T. Pexton (1994). "Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae." Mol Cell Biol **14**(6): 3834-41.
Kaern, M., T. C. Elston, W. J. Blake and J. J. Collins (2005). "Stochasticity in gene expression: from theories to phenotypes." Nat Rev Genet **6**(6): 451-64.
Kaufmann, B. B. and A. van Oudenaarden (2007). "Stochastic gene expression: from single molecules to the proteome." Curr Opin Genet Dev **17**(2): 107-12.
Kaufmann, B. B., Q. Yang, J. T. Mettetal and A. van Oudenaarden (2007). "Heritable stochastic switching revealed by single-cell genealogy." PLoS Biol **5**(9): e239.
Keegan, L., G. Gill and M. Ptashne (1986). "Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein." Science **231**(4739): 699-704.
Kepler, T. B. and T. C. Elston (2001). "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations." Biophys J **81**(6): 3116-36.
Kimmel, A. R. and C. A. Parent (2003). "The signal to move: D-discoideum go orienteering." Science **300**(5625): 1525-1527.
Kitano, H. (2002). "Systems biology: a brief overview." Science **295**(5560): 1662-4.
Kitano, H., A. Funahashi, Y. Matsuoka and K. Oda (2005). "Using process diagrams for the graphical representation of biological networks." Nat Biotechnol **23**(8): 961-6.
Klipp, E., B. Nordlander, R. Kruger, P. Gennemark and S. Hohmann (2005). "Integrative model of the response of yeast to osmotic shock." Nat Biotechnol **23**(8): 975-82.
Korobkova, E., T. Emonet, J. M. Vilar, T. S. Shimizu and P. Cluzel (2004). "From molecular noise to behavioural variability in a single bacterium." Nature **428**(6982): 574-8.
Kramers, H. A. (1940). "Brownian motion in a field of force and the diffusion model of chemical reactions." Physica **7**(4): 284-304.
Kussell, E., R. Kishony, N. Q. Balaban and S. Leibler (2005). "Bacterial persistence: a model of survival in changing environments." Genetics **169**(4): 1807-14.
Kussell, E. and S. Leibler (2005). "Phenotypic diversity, population growth, and information in fluctuating environments." Science **309**(5743): 2075-8.
Lachmann, M. and E. Jablonka (1996). "The inheritance of phenotypes: an adaptation to fluctuating environments." J Theor Biol **181**(1): 1-9.
Le, T. T., S. Harlepp, C. C. Guet, K. Dittmar, T. Emonet, T. Pan and P. Cluzel (2005). "Real-time RNA profiling within a single bacterium." Proc Natl Acad Sci U S A **102**(26): 9160-4.
Lederberg, J. and T. Iino (1956). "Phase Variation in Salmonella." Genetics **41**(5): 743-57.
Levchenko, A. and P. A. Iglesias (2002). "Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils." Biophys J **82**(1 Pt 1): 50-63.
Levin, M. D. (2003). "Noise in gene expression as the source of non-genetic individuality in the chemotactic response of Escherichia coli." Febs Letters **550**(1-3): 135-138.
Levin, M. D., C. J. Morton-Firth, W. N. Abouhamad, R. B. Bourret and D. Bray (1998). "Origins of individual swimming behavior in bacteria." Biophys J **74**(1): 175-81.

Levins, R. (1968). Evolution in changing environments: some theoretical explorations. Princeton, NJ, Princeton University Press.
Lilly, P. J. and P. N. Devreotes (1995). "Chemoattractant and GTP gamma S-mediated stimulation of adenylyl cyclase in Dictyostelium requires translocation of CRAC to membranes." J Cell Biol **129**(6): 1659-65.
Lipan, O. and W. H. Wong (2005). "The use of oscillatory signals in the study of genetic networks." Proc Natl Acad Sci U S A **102**(20): 7063-8.
Liu, J. Y., P. F. Miller, J. Willard and E. R. Olson (1999). "Functional and biochemical characterization of Escherichia coli sugar efflux transporters." J Biol Chem **274**(33): 22977-84.
Ljung, L. (1999). System Identification Theory for the User, Prentice Hall.
Luyten, K., J. Albertyn, W. F. Skibbe, B. A. Prior, J. Ramos, J. M. Thevelein and S. Hohmann (1995). "Fps1, a yeast member of the MIP family of channel proteins, is a facilitator for glycerol uptake and efflux and is inactive under osmotic stress." Embo J **14**(7): 1360-71.
Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (Mus musculus L.)." Nature **190**: 372-3.
Maeda, T., M. Takekawa and H. Saito (1995). "Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor." Science **269**(5223): 554-8.
Maheshri, N. and K. O'Shea E (2007). "Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression." Annu Rev Biophys Biomol Struct **36**: 413-434.
Maloney, P. C. and B. Rotman (1973). "Distribution of suboptimally induces -D-galactosidase in Escherichia coli. The enzyme content of individual cells." J Mol Biol **73**(1): 77-91.
Manahan, C. L., P. A. Iglesias, Y. Long and P. N. Devreotes (2004). "Chemoattractant signaling in dictyostelium discoideum." Annu Rev Cell Dev Biol **20**: 223-53.
Margolin, A. A. and A. Califano (2007). "Theory and limitations of genetic network inference from microarray data." Ann N Y Acad Sci **1115**: 51-72.
Meili, R., C. Ellsworth, S. Lee, T. B. K. Reddy, H. Ma and R. A. Firtel (1999). "Chemoattractant-mediated transient activation and membrane localization of Akt/PKB is required for efficient chemotaxis to cAMP in Dictyostelium." Embo Journal **18**(8): 2092-2105.
Meili, R. and R. A. Firtel (2003). "Follow the leader." Dev Cell **4**(3): 291-3.
Meinhardt, H. (1999). "Orientation of chemotactic cells and growth cones: models and mechanisms." J Cell Sci **112 (Pt 17)**: 2867-74.
Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak and A. van Oudenaarden (2006). "Predicting stochastic gene expression dynamics in single cells." Proc Natl Acad Sci U S A **103**(19): 7304-9.
Muller, H. (1930). "Types of visible variations induced by x-rays in Drosophila." Journal of Genetics **22**: 299-334.
Narang, A., K. K. Subramanian and D. A. Lauffenburger (2001). "A mathematical model for chemoattractant gradient sensing based on receptor-regulated membrane phospholipid signaling dynamics." Ann Biomed Eng **29**(8): 677-91.
Nerbonne, J. M., S. Richard, J. Nargeot and H. A. Lester (1984). "New photoactivatable cyclic nucleotides produce intracellular jumps in cyclic AMP and cyclic GMP concentrations." Nature **310**(5972): 74-6.
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**(7095): 840-6.

Nobelmann, B. and J. W. Lengeler (1996). "Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism." J Bacteriol **178**(23): 6790-5.
Novick, A. and M. Weiner (1957). "Enzyme Induction as an All-or-none Phenomenon." Proc Natl Acad Sci U S A **43**(7): 553-566.
Oppenheim, A. V., Willsky, A. S., Young, I. T. (1983). Signals and systems, Prentice-Hall.
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." Nat Genet **31**(1): 69-73.
Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman and A. Van Oudenaarden (2004). "Multistability in the lactose utilization network of Escherichia coli." Nature **427**(6976): 737-40.
Parent, C. A., B. J. Blacklock, W. M. Froehlich, D. B. Murphy and P. N. Devreotes (1998). "G protein signaling events are activated at the leading edge of chemotactic cells." Cell **95**(1): 81-91.
Patino, M. M., J. J. Liu, J. R. Glover and S. Lindquist (1996). "Support for the prion hypothesis for inheritance of a phenotypic trait in yeast." Science **273**(5275): 622-6.
Paulsson, J. (2004). "Summing up the noise in gene networks." Nature **427**(6973): 415-8.
Pedram, M., C. N. Sprung, Q. Gao, A. W. Lo, G. E. Reynolds and J. P. Murnane (2006). "Telomere position effect and silencing of transgenes near telomeres in the mouse." Mol Cell Biol **26**(5): 1865-78.
Pedraza, J. M. and A. van Oudenaarden (2005). "Noise propagation in gene networks." Science **307**(5717): 1965-9.
Peng, G. and J. E. Hopper (2002). "Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein." Proc Natl Acad Sci U S A **99**(13): 8548-53.
Postma, M. and P. J. M. Van Haastert (2001). "A diffusion-translocation model for gradient sensing by chemotactic cells." Biophysical Journal **81**(3): 1314-1323.
Powell, E. O. (1958). "An outline of the pattern of bacterial generation times." J Gen Microbiol **18**(2): 382-417.
Price, N. D. and I. Shmulevich (2007). "Biochemical and statistical network models for systems biology." Curr Opin Biotechnol **18**(4): 365-70.
Proft, M. and K. Struhl (2004). "MAP kinase-mediated stress relief that precedes and regulates the timing of transcriptional induction." Cell **118**(3): 351-61.
Ramsey, S. A., J. J. Smith, D. Orrell, M. Marelli, T. W. Petersen, P. de Atauri, H. Bolouri and J. D. Aitchison (2006). "Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast." Nat Genet **38**(9): 1082-7.
Rando, O. J. and K. J. Verstrepen (2007). "Timescales of genetic and epigenetic inheritance." Cell **128**(4): 655-68.
Rao, C. V., D. M. Wolf and A. P. Arkin (2002). "Control, exploitation and tolerance of intracellular noise." Nature **420**(6912): 231-7.
Rappel, W. J., P. J. Thomas, H. Levine and W. F. Loomis (2002). "Establishing direction during chemotaxis in eukaryotic cells." Biophysical Journal **83**(3): 1361-1367.
Raser, J. M. and E. K. O'Shea (2004). "Control of stochasticity in eukaryotic gene expression." Science **304**(5678): 1811-4.
Raser, J. M. and E. K. O'Shea (2005). "Noise in gene expression: origins, consequences, and control." Science **309**(5743): 2010-3.
Ray, D. T. and J. D. McCreight (1996). "Yellow-tip: a cytoplasmic inherited trait in melon (Cucumis melo L.)." The Journal of Heredity **87**(3): 245-247.

Reiser, V., H. Ruis and G. Ammerer (1999). "Kinase activity-dependent nuclear export opposes stress-induced nuclear accumulation and retention of Hog1 mitogen-activated protein kinase in the budding yeast Saccharomyces cerevisiae." Mol Biol Cell **10**(4): 1147-61.

Risken, H. (1996). The Fokker-Planck Equation: Methods of Solutions and Applications. Berlin, Springer-Verlag.

Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz (2005). "Gene regulation at the single-cell level." Science **307**(5717): 1962-5.

Samadani, A., J. Mettetal and A. van Oudenaarden (2006). "Cellular asymmetry and individuality in directional sensing." Proc Natl Acad Sci U S A **103**(31): 11549-54.

Samoilov, M., A. Arkin and J. Ross (2002). "Signal processing by simple chemical systems." Journal of Physical Chemistry A **106**(43): 10205-10221.

Samoilov, M. S., G. Price and A. P. Arkin (2006). "From fluctuations to phenotypes: the physiology of noise." Sci STKE **2006**(366): re17.

Savageau, M. A. (1976). Biochemical systems analysis. A study of function and design in molecular biology. Reading, MA, Addison-Wesley.

Schaffer, W. M. (1974). "Optimal efforts in fluctuating environments." Am. Nat. **108**: 783-790.

Seger, J. a. B., H. J. (1988). What is bet-hedging? Oxford, Oxford Univ. Press.

Sheff, M. A. and K. S. Thorn (2004). "Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae." Yeast **21**(8): 661-70.

Simpson, M. L., C. D. Cox and G. S. Sayler (2003). "Frequency domain analysis of noise in autoregulated gene circuits." Proc Natl Acad Sci U S A **100**(8): 4551-6.

Slatkin, M. (1974). "Hedging one's evolutionary bets." Nature **250**: 704-705.

Slepchenko, B. M., J. C. Schaff, I. Macara and L. M. Loew (2003). "Quantitative cell biology with the Virtual Cell." Trends Cell Biol **13**(11): 570-6.

Sontag, E. D. (2005). "Molecular systems biology and control." European Journal of Control **11**(4-5): 396-435.

Spudich, J. L. and D. E. Koshland, Jr. (1976). "Non-genetic individuality: chance in the single cell." Nature **262**(5568): 467-71.

Stearns, S. C. (1976). "Life-history tactics: a review of the ideas." Q Rev Biol **51**(1): 3-47.

Suel, G. M., J. Garcia-Ojalvo, L. M. Liberman and M. B. Elowitz (2006). "An excitable gene regulatory circuit induces transient cellular differentiation." Nature **440**(7083): 545-50.

Swain, P. S., M. B. Elowitz and E. D. Siggia (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proc Natl Acad Sci U S A **99**(20): 12795-800.

Tamas, M. J., K. Luyten, F. C. Sutherland, et al. (1999). "Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation." Mol Microbiol **31**(4): 1087-104.

Thattai, M. and A. van Oudenaarden (2001). "Intrinsic noise in gene regulatory networks." Proc Natl Acad Sci U S A **98**(15): 8614-9.

Thattai, M. and A. van Oudenaarden (2002). "Attenuation of noise in ultrasensitive signaling cascades." Biophys J **82**(6): 2943-50.

Thattai, M. and A. van Oudenaarden (2004). "Stochastic gene expression in fluctuating environments." Genetics **167**(1): 523-30.

Thorsen, M., Y. Di, C. Tangemo, et al. (2006). "The MAPK Hog1p modulates Fps1p-dependent arsenite uptake and tolerance in yeast." Mol Biol Cell **17**(10): 4400-10.

Tosh, D. and J. M. Slack (2002). "How cells change their phenotype." Nat Rev Mol Cell Biol **3**(3): 187-94.

Tranquillo, R. T., D. A. Lauffenburger and S. H. Zigmond (1988). "A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations." J Cell Biol **106**(2): 303-9.
Tyson, J. J., K. Chen and B. Novak (2001). "Network dynamics and cell physiology." Nat Rev Mol Cell Biol **2**(12): 908-16.
van der Woude, M. W. and A. J. Baumler (2004). "Phase and antigenic variation in bacteria." Clin Microbiol Rev **17**(3): 581-611, table of contents.
Van Haastert, P. J. and P. N. Devreotes (2004). "Chemotaxis: signalling the way forward." Nat Rev Mol Cell Biol **5**(8): 626-34.
Vance, W., A. Arkin and J. Ross (2002). "Determination of causal connectivities of species in reaction networks." Proc Natl Acad Sci U S A **99**(9): 5816-21.
Vilar, J. M., C. C. Guet and S. Leibler (2003). "Modeling network dynamics: the lac operon, a case study." J Cell Biol **161**(3): 471-6.
Volfson, D., J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring and J. Hasty (2006). "Origins of extrinsic variability in eukaryotic gene expression." Nature **439**(7078): 861-4.
Wakamoto, Y., J. Ramsden and K. Yasuda (2005). "Single-cell growth and division dynamics showing epigenetic correlations." Analyst **130**(3): 311-7.
Westfall, P. J. and J. Thorner (2006). "Analysis of mitogen-activated protein kinase signaling specificity in response to hyperosmotic stress: use of an analog-sensitive HOG1 allele." Eukaryot Cell **5**(8): 1215-28.
Westwick, D. T., Kearney, R. E. (2003). Identification of Nonlinear Physiological Systems, IEEE Press.
Wolf, D. M. and A. P. Arkin (2002). "Fifteen minutes of fim: control of type 1 pili expression in E. coli." Omics **6**(1): 91-114.
Wolf, D. M., V. V. Vazirani and A. P. Arkin (2005). "Diversity in times of adversity: probabilistic strategies in microbial survival games." J Theor Biol **234**(2): 227-53.
Wong, P., S. Gladney and J. D. Keasling (1997). "Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose." Biotechnol Prog **13**(2): 132-43.
Xiao, Z., N. Zhang, D. B. Murphy and P. N. Devreotes (1997). "Dynamic distribution of chemoattractant receptors in living cells during chemotaxis and persistent stimulation." J Cell Biol **139**(2): 365-74.
Xiong, W. and J. E. Ferrell, Jr. (2003). "A positive-feedback-based bistable 'memory module' that governs a cell fate decision." Nature **426**(6965): 460-5.
Xu, X., M. Meier-Schellersheim, X. Jiao, L. E. Nelson and T. Jin (2005). "Quantitative imaging of single live cells reveals spatiotemporal dynamics of multistep signaling events of chemoattractant gradient sensing in Dictyostelium." Mol Biol Cell **16**(2): 676-88.
Yi, T. M., Y. Huang, M. I. Simon and J. Doyle (2000). "Robust perfect adaptation in bacterial chemotaxis through integral feedback control." Proc Natl Acad Sci U S A **97**(9): 4649-53.
Yildirim, N. and M. C. Mackey (2003). "Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data." Biophys J **84**(5): 2841-51.