The Biological Role of Stochasticity

From Molecules to Communities

APPROVED BY SUPERVISORY COMMITTEE

Gürol Mehmet Süel, Ph.D., Supervisor

Rama Ranganathan, M.D., Ph.D., Chair

Neal Alto, Ph.D.

Jose Rizo-Rey, Ph.D.

DEDICATION

Dedicated to my family including my wife, Crystal, my parents, Srinarong and Vannee, and all those who have supported me throughout my life.

Acknowledgments

I would like to thank the members of my Dissertation Committee including former members of my committee: Gürol Süel, Ph.D.; Rama Ranganathan, M.D., Ph.D.; Neal Alto, Ph.D.; and Jose Rizo-Rey, Ph.D.; Christian Forst, Ph.D.; and Leon Avery, Ph.D. I would also like to thank members of the Süel Lab including Tolga Çağatay and Anna Kuchina.

THE BIOLOGICAL ROLE OF STOCHASTICITY FROM MOLECULES TO COMMUNITIES

by

MARK ANDREW KITTISOPIKUL, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

June 2013

The Biological Role of Stochasticity

From Molecules to Communities

Publication No. _____

Mark Andrew Kittisopikul, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2013

Supervisor: Gürol Mehmet Süel, Ph.D.

The presence of stochasticity in biology engenders the question of the role of such randomness. While stochasticity may be an artifact of biochemical interactions, it could also be an actively regulated aspect of life. Here I investigate how noise may be encoded within the molecular interactions of a biochemical network and then examine how those consequences propagate to higher levels of organization including isogenic populations of cells and host-pathogen interactions. I particularly look into how this variability impacts the ability of populations to respond to their environments. I conclude that stochasticity is selectable property of life that adds robustness to biological processes during periods of uncertainty.

Table of Contents

List of Tables

List of Figures

Prior Publications

1. **Kittisopikul, M.** & Süel, G.M. Biological role of noise encoded in a genetic network motif. Proceedings of the National Academy of Sciences, National Acad Sciences, 2010, 107, 13300-13305.

2. Orchard, R.; **Kittisopikul, M.**; Altschuler, S.; Wu, L.; Süel, G.M. & Alto, N. Identification of F-actin as the Dynamic Hub in a Microbial-Induced GTPase Polarity Circuit. Cell, Elsevier, 2012, 148, 803-815.

3. Asally, M.; **Kittisopikul, M.**; Ru, P.; Du, Y.; Hu, Z.; Çağatay, T.; Robinson, A.; Lu, H.; Garcia-Ojalvo, J. & Süel, G.M. Localized cell death focuses mechanical forces during 3D patterning in a biofilm. Proceedings of the National Academy of Sciences, National Acad Sciences, 2012.

Abbreviations

B. subtilis *Bacillus subtilis*

CDP Cell death pattern

FFL Feed-Forward Loop

GEF Guanine-nucelotide Exchange Factor

HGT Horizontal Gene Transfer

HGR Homologous Gene Recombination

IPTG Isopropyl β-D-1-thiogalactopyranoside

SSA Stochastic simulation algorithm

Chapter 1

Introduction

> "Life is like a box of chocolates. You never know what you're ganna to get." Forest Gump in *Forest Gump* (1994)

Given the immense knowledge we have gained about biology over the last century, are we any closer to being able to understand life? Despite being able to decode the basic blueprints of life, genomes, our ability to predict how biology functions is severely limited. While we understand why an apple falls to the ground on earth or how a plane can take flight, we have little ability to predict the future activities of our own cells or even bacteria. The complexity of biology not only presents a difficult system to understand, but also prevents us from knowing the entire state of the system. This uncertainty may also be an aspect inherent in the principle of life. While we may not currently use such principles in our current engineering methods, it would be more unusual for biology not to take advantage of such an inherent part of life rather than not. Despite not having complete information about the system to make an absolute prediction, we can still predict the likelihood of various outcomes. I will explore this stochastic aspect of biology from the scale of molecules to interactions between species.

What does it mean to be able to predict life? Given the initial state of an organism and it's environment, a prediction would be able to able to give the future state of the organism. That is for a set of environmental variables, E, and internal state variables, O, we would like to predict a future

distinguishable state, a phenotype, Ψ, at a time, t. The organism can then be imagined as a function f of the environment, internal state, and time mapped to a phenotype:

$$f(E,O,t) = \Psi \tag{1.1}$$

If a single outcome Ψ can be written based on the variables, then the system could be called deterministic. There are two main problems with trying to address a biological system in this manner. First, it is difficult to identify all the possible environmental and state variables that may affect the future outcome Ψ. Additionally, even if all the variables were known it would be difficult to be able to measure all of them. A measurement of all the variables will likely destroy the environment and the organism making a prediction of the future state no longer relevant. This introduces some uncertainty into the study and predictability of life. Even without this uncertainty, however, it may still be difficult to predict the phenotype of an organism at a future time point.

1.1 Chaos

One way to generate diverse phenotypes within a mostly homogeneous environment is to take advantage of very small differences within that environment. If small changes can deterministically lead to vastly different outcomes, then the biological system is called chaotic. These changes may not be external to a biological system, but could be in internal variable such as the number of ribosomes present in the cell or even the present locations of a molecule of mRNA. That is for a small change in the environment ε or difference in the initial state of an organism or cell θ that a very different outcome Ψ_2 is achieved:

$$f(E+\varepsilon, O+\theta, t) = \Psi_2 \tag{1.2}$$

In this way, the phenotype after some time, t, is highly sensitive to variation in the environment. Variation in the environment could thus result in result in variability of a population.

1.2 Stochasticity

Stochasticity occurs when we can predict the outcome of events better by probability of one event occurring out of many events rather than a certain outcome as in the deterministic case. That is rather than one outcome being possible given a set of parameters, we can better describe the process as follows:

$$Pr(\Psi_n|E,O,t) = \rho_n(E,O,t) \tag{1.3}$$

$$\text{where} \quad \sum_n \rho_n(E,O,t) = 1 \tag{1.4}$$

The result is better described as a distribution over the set of possible outcomes, $\{\Psi_n\}$, where the probability of each outcome, $Pr(\Psi_n|E,O,t)$ is conditional on the external and internal parameters as well as time. For example, when flipping an evenly weighted coin with two sides, heads and tails, the predicted outcome is that we have an equal chance of the coin landing with heads or tails up. That is to say there is a 50% probability of heads being the outcome and 50% probability of tails being the outcome. Because of this uncertainty, this outcome is non-deterministic. It cannot be determined ahead of time what the outcome will be for sure. Perhaps, if we had information on exactly how the coin was flipped, which side started up, and all the atmospheric conditions then we could make an accurate and certain prediction, but obtaining this information or controlling the conditions is likely beyond the time or resources available for such an event. If a large number of coins were flipped, perhaps a million coins, then the outcome of the set of coins tends to become deterministic. The larger the set of coins, the number of coins with heads should be about half of the coins flipped. While we still have limited information and predictability on an individual coin, the outcome of the population of coins becomes more certain. This is referred to as the Law of Large Numbers in probability theory.

The act of flipping a coin can be related to biological systems as will be discussed later in this dissertation in detail. The main concept though is that within a living organism, cell, or a compartment thereof there are often small numbers of molecules whose interactions determine the observable phenotypes. As with flipping a few coins, the outcome of these interactions can be better described stochastically than deterministically. Other phenotypes are determined by larger numbers of proteins or cells, however, so stochasticity is not always a given.

1.3 Heterogeneity

Variation within life is a simple fact even appreciated by Forest Gump. While we appreciate this intrinsic variability in life as much as we enjoy different flavors in a box of chocolates, this variability also confounds our study of biology. Rather than merely being an impediment to research, this variability may also be an important part of how cells function. Variability is a function of scale. Viewed from a different scale, heterogeneous mixtures become homogeneous or vice versa. The study of biology now spans scales from the molecular to the ecological. Observing how variation at the molecular level might affect variation at the ecological level could be critical to our understanding of life.

We can describe variation over a certain region, Γ, as a distribution. This distribution is can be defined as a frequency or probability of finding a certain phenotype Ψ_n of the set all possible phenotypes within region Γ.

$$P(\Psi_n|\Gamma) = \rho_n(\Gamma) \tag{1.5}$$

$$\sum_n \rho_n(\Gamma) = 1 \tag{1.6}$$

We can then define homogeneity and heterogeneity more precisely. If for any two regions of a similar scale Γ and Γ' within a larger region, Ω, the frequencies of finding a certain state, Ψ_n, are

the same, then we can say that Ω is homogeneous for, Ψ_n in regions of the scale of Γ. This can be stated as follows:

Ω is homogeneous at the scale of Γ if and only if

$$P(\Psi_n|\Gamma) \quad = \quad P(\Psi_n|\Gamma_2) \quad \forall \quad \Gamma,\Gamma_2 \quad \text{of similar scale}$$

If the distribution of the two regions are significantly different, then Ω is heterogeneous at the scale of Γ. For example, one operational region of life is the cell. It provides a natural compartment that creates a physical separation between different copies of life. For a population of cells as Ω, we can then discuss cellular heterogeneity or homogeneity.

If there is apparent heterogeneity at a certain scale, is this due to a stochastic process or a deterministic one? A true stochastic process can only be clearly established through elimination of all possible deterministic causes through precise control or measurement of all conditions that may affect the outcome. As noted above, a deterministic but chaotic process could produce many results with just the slightest of variations. However, there is an effective limit on the amount that can be known for the system not just for the research biologist but also for the biochemical processes that are attempting to respond to an environment. While an experimentalist is limited by the available measurement techniques, the cell is limited by biochemical sensors and their reliability. Meanwhile, by employing the Law of Large Numbers, we may be able to turn stochastic processes into deterministic systems. However, this still does not offer conclusive proof of stochasticity. In a complex system, we may simply be drowning out other inputs by increasing the relative contribution a single input. I thus approach stochasticity on an empirical basis aware of the possibility of hidden variables that may fully determine the system.

1.4 Bet-Hedging

Stochasticity therefore is a formalism for dealing with the unknown. When the parameters are uncertain or the outcome is uncertain based on known parameters, then stochasticity may be mechanism to deal with an uncertain future by creating heterogeneity. This is called bet-hedging. Rather than betting on a single outcome, a community of organisms may do better betting on multiple outcomes since failure of part of a population is often desirable to the possible demise of the entire population. There is an illustrative analogy of this to gambling. For a game of chance such as roulette where a number (1 through 38) and color (red or black) are randomly chosen, placing all of one's wealth on a single number or color could end an evening of risk taking thrills. Spreading the bet between many or all possibilities would allow a greater chance for another round of betting to take place. Bet-hedging through stochasticity is a mechanism to ameliorate risk. The continuity of life in the face of risks is a basic tenet of our understanding of life.

1.5 Genetic Regulatory Networks and other Biochemical Networks

Biological organisms integrate information about their environment through signalling networks involving proteins and nucleic acids among other functional molecules. These biological molecules form complex webs of interactions with each other that can be organized into signalling cascades and networks. In particular, a genetic regulatory network (GRN) involves how genes encoded in deoxynucleic acids, DNA, affect how other genes are transcribed, activated, or repressed.

A genetic regulatory network is a set of interactions between genes in which the genes regulate the activity of each other. The activity of a gene describes how many gene products are produced such as RNA transcripts that are copied from the gene by RNA polymerase. The interaction is directional in that one gene, the regulator, may affect the activity of another gene, the target, but not

vice versa. These genetic interactions also have a sign. An interaction is positive, or activating, if the activity of the target gene increases when the interaction is present. In contrast, it is negative when the activity of a gene decreases in the presence of the interaction. These networks are the biological mechanism that an organism uses to interpret their environment and function. There exists few copies of genes within a particular biological cell, usually one in prokaryotes, two in eukaryotes, and perhaps more during cell replication. Given that the number of copies of these genes are countable, it seems likely that biological activity concerning these genes may be stochastic.

A biochemical network is generalization of a GRN where the interactions may not involve regulation at the level of transcription, but rather consist of direct interactions between gene products. An example is the degradation of a protein by another protein called a protease. Another example is the activation of a G-protein through the enzymatic exchange of GDP for GTP by another protein. As in GRNs, small numbers of molecules in these networks may also produce stochastic effects.

1.6 Natural Selection

Natural selection is the process in which organisms that have more advantageous traits are able to reproduce more than organisms with less beneficial traits. Natural selection allows for properties that positively affect the survivability, or fitness, of an organism in it's environment to be propagated to future generations. Selectable properties must be able to affect fitness of an organism, while also being encoded in such a way that is inheritable by the next generation. For example, the organism that competes the most successfully for food while avoiding death will continue to reproduce in an ecological system. Meanwhile those organisms that cannot sustain themselves will not be able to reproduce and will perish. The ability to sense nutrients is a critical property that

can encoded in the amino acid sequence and structure of a protein. This allows that property to be selectable.

Since heterogeneity generated by stochasticity allows organisms to mitigate risk and certain biochemical networks involving low number of molecules may be subject to stochasticity, perhaps stochasticity is a trait that can be selected for in natural environments.

1.7 Hypothesis

I propose that stochasticity is a biological property that can be selected for by the environment of an organism. Therefore, *I hypothesize that stochasticity is encodable within an inheritable network of biochemical interactions and that stochasticity allows biological processes to be robust in uncertain environments.*

1.8 Approach

I will first examine how the numbers of molecules within a cell serve as the origin of molecular stochasticity, or noise. Furthermore, how noise might be encoded and controlled via networks of interactions will be discussed in Chapter 2. Bacteria will be used as a model for noise due to the relative simplicity compared to more complex organisms and the ability to experimentally manipulate and observe noise in Chapters 2 and 3. Further examples of how stochasticity affects communities of bacteria and host-pathogen interactions will be shown in Chapters 4 and 5. In the penultimate chapter, Chapter 6, I propose a set of experiments to address directly whether stochasticity is a selectable property. In this way, I hope to address the origins and consequences of stochasticity from the molecular scale up to an ecological community of organisms.

Chapter 2

Molecular basis of stochasticity encoded in genetic regulatory network architecture

2.1 Preamble

If variability and the stochasticity that generates it is critical to life, then there must be a way for this variability to be encoded. While genetics encodes proteins and other functional molecules, I demonstrate that the interactions between those molecules is how stochasticity may be encoded. Encoding stochasticity is critical to for it to play a role in biology. Without a way for stochasticity to be purposefully inherited, then natural selection and thus evolution of stochasticity would not be possible. In this chapter, I examine how a simple but common network motif called a feed-forward loop might encode stochasticity.

2.2 Abstract

Genetic circuits that regulate distinct cellular processes can differ in their wiring pattern of interactions (architecture) and susceptibility to stochastic fluctuations (noise). Whether the link between circuit architecture and noise is of biological importance remains, however, poorly understood. To investigate this problem, we performed a computational study of gene expression noise for all possible circuit architectures of feed-forward loop (FFL) motifs. Results revealed that FFL architectures fall into two categories depending on whether their ON (stimulated) or OFF (unstimulated) steady states exhibit noise. To explore the biological importance of this difference in noise

behavior, we analyzed 858 documented FFLs in Escherichia coli that were divided into 39 functional categories. The majority of FFLs were found to regulate two subsets of functional categories. Interestingly, these two functional categories associated with FFLs of opposite noise behaviors. This opposite noise preference revealed two noise-based strategies to cope with environmental constraints where cellular responses are either initiated or terminated stochastically to allow probabilistic sampling of alternative states. FFLs may thus be selected for their architecture-dependent noise behavior, revealing a biological role for noise that is encoded in gene circuit architectures.

2.3 Introduction

Cellular processes are typically regulated by genetic circuits with particular architectures of interactions among genes and proteins. However, it is not well understood whether different architectures of genetic circuits generate distinct properties that can be subject to selective pressures. For example, selection of circuit architectures can be driven by the requirement to generate biologically important dynamic behaviors such as oscillations [1]. However, other selective pressures must also exist because natural genetic oscillators, such as circadian clocks and cell cycle circuits, can differ in architecture [2, 3, 4, 5, 6]. Furthermore, a recent study in Bacillus subtilis showed that the dynamics of a natural cellular differentiation circuit could be reconstituted by a synthetic circuit with an alternative architecture but with differences in variability (noise) and physiology [7]. These and other studies suggest that gene circuit architectures can encode distinct properties such as susceptibility to noise that could be critical to the physiological process that they implement [3, 8, 9, 10, 11, 12, 13]. Systematic comparisons of alternative architectures could therefore reveal different properties supported by distinct topologies and help uncover the biological importance of gene circuit architecture.

Feed-forward loops (FFLs) constitute an ideal gene circuit motif for studying the relationship between circuit architecture and biological function because of their simple architecture and well characterized functional roles in organisms such as *Escherichia coli* [2] and Saccharomyces cerevisia [14]. In a FFL circuit, a transcription factor A regulates a second transcription factor B and both can regulate expression of the output gene C (Fig.2.1 A and E). Therefore, expression of the FFL output gene C represents the integration of the activities of A and B transcription factors. There are a total of eight possible FFL architectures because the regulatory links among A, B, and C can either be positive (activation) or negative (repression). Examples of all possible FFL architectures have been identified and shown to regulate a multitude of cellular processes in a diverse range of organisms ranging from bacteria to human cells [15, 16]. The large body of knowledge on FFLs makes this motif an appropriate model system to study the link between circuit architecture and biological function [4].

Continuous simulations based on ordinary differential equations have suggested that distinct FFL architectures can differ in their dynamics. In particular, differences have been observed between two types of architectures classified as coherent and incoherent FFLs based on whether the net sign of direct and indirect (through B) regulatory links from A to C are the same or opposite, respectively [2]. For example, it has been shown that coherent FFL architectures can serve as delay elements where the expression of the output gene C is delayed with respect to the activation of the input transcription factor A [2]. Compared to coherent FFLs, incoherent circuits in turn have been shown to have an accelerated output response to input, where the maximum expression of the output gene C is lower and thus reached sooner upon activation of A [17]. Therefore, continuous simulations have revealed that differences in the architectures and logics of FFLs can give rise to divergent dynamics.

While continuous simulations can predict gene circuit dynamics, they do not account for the stochastic behavior that is inherent to the biochemical reactions comprising FFLs. Stochastic fluctuations can alter the dynamics of genetic circuits and even induce qualitatively distinct behaviors [6, 18, 19, 20]. For example, probabilistic interactions among small numbers of molecules can generate stochastic bursts of gene expression. Single-cell and single-molecule measurements have characterized these bursts and implicated transcription and translation processes as possible sources [21, 22, 23]. Perhaps more importantly, recent studies have shown that gene expression bursts can serve a biological function [13, 24, 20]. Stochastic fluctuations have also been shown to depend on the architecture of genetic circuits [7, 25, 26]. Additionally, recent studies have begun to show that distinct FFL architectures can differ in behavior at the stochastic regime [27]. For example, it has been suggested that coherent FFLs amplify circuit-extrinsic noise at the C output, while incoherent FFLs attenuate such noise [26]. Furthermore, an analytical study has shown that the most abundant coherent FFL exhibits the lowest noise amplitudes of all FFL architectures [27]. In contrast, the most abundant incoherent FFL architecture exhibits the highest noise amplitudes [27]. It is however unclear if all possible FFL architectures differ in stochastic behavior and, more importantly, if differences in noise behavior are of biological importance.

To systematically investigate the relationship between FFL architecture, noise, and function, we performed discrete stochastic simulations for all possible three-component circuit architectures and three logic gates (AND, OR, and XOR). We found that all FFL circuit architectures could be classified into two categories according to how susceptible their ON and OFF steady states were to noise, independent of their logic gates. This noise behavior of FFL architectures is coupled to circuit function. In particular, these data show that FFLs with high noise in their OFF state preferentially regulate rare stochastic processes in E. coli such as the generation of antibiotic-resistant persister

cells [28]. In contrast, cellular processes that are typically in high demand, such as anaerobic respiration, are found to be regulated by FFLs with high noise in their ON state. While FFLs with higher noise in their OFF state can stochastically initiate rare cellular responses, FFLs with higher noise in their ON state can stochastically terminate cellular processes that are in high demand. These results suggest that specific FFL architectures may be selected based on their distinct noise behaviors to allow sampling of alternative cellular states and cope with environmental constraints.

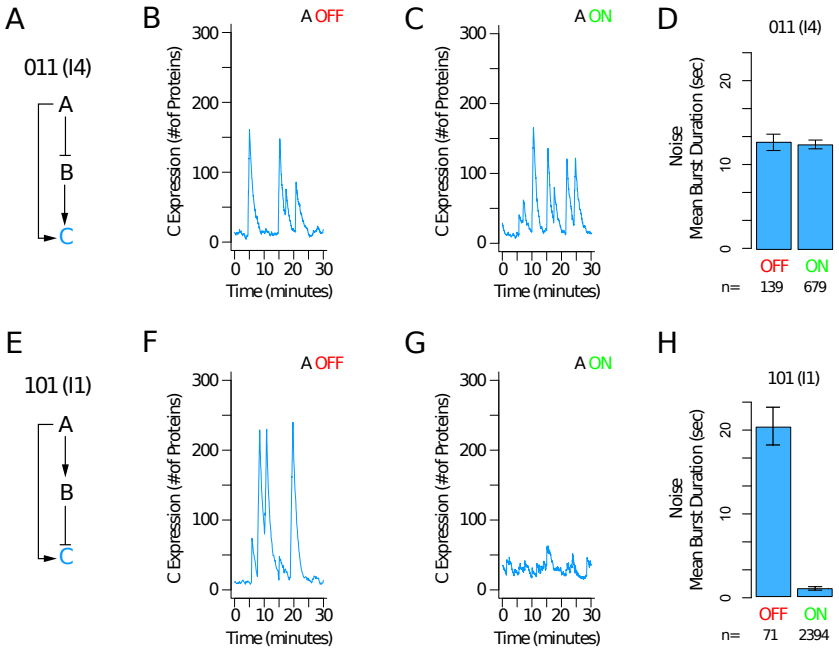
2.4 Results

2.4.1 Two Incoherent FFLs Differ in Their Susceptibility to Gene Expression Bursts

In order to begin analyzing if there were any noise differences, we selected two FFLs that seemed to perform a similar function: 011 (I4) and 101 (I1). Both of these incoherent feed-forward loops have the input gene directly activating the output gene, but indirectly they inhibit the output gene. The main difference is that the order of activation and repression differ in the indirect interactions between input and output genes. Alon et al. previously described differences in the dynamics of these two networks in response to ON and OFF steps.

We began by investigating a pair of FFL circuits with similar architecture (Fig.2.1A and 2.1E). In both circuits, the input node A transcriptionally activates output node C directly and also represses C indirectly through node B. Since the direct and indirect regulatory pathways have opposite signs (direct activating and indirect repressing), both circuits are traditionally classified as incoherent FFLs [2]. The only difference in architecture between these circuits is therefore the opposite order of sequential activation and repression reactions comprising their respective indirect pathways. We refer to circuits with such alternative architectures as isocircuits. For easier comparison, we adopt here a three digit binary nomenclature that classifies FFL circuits based on the

Figure 2.1: The 011 (I4) and 101 (I1) Incoherent FFLs Reveal Two Noise Behaviors. Stochastic simulations reveal two noise behaviors in incoherent FFLs. AD and EH pertain to the FFL circuits 011 (A) and 101 (E), respectively. The logic of integration for the regulation of the output node C is a Boolean AND gate. B, C, F, and G show time traces of C molecule numbers expressed from the C output promoter as obtained from stochastic simulations based on the Gillespie algorithm [29]. B and F show data for C expression obtained in the OFF steady state (unstimulated state of A). C and G depict data for the ON steady state (stimulated state of A). D and H show the mean durations of high gene expression bursts of C obtained from simulations and precisely defined by the time (seconds) the C promoter remains in the high expression state as determined by the binding state of transcription factors A and B.



signs of interactions between A-B, B-C and A-C nodes respectively, where 1=activating and 0=repressing (Fig. 2.1A and 2.1E). Despite similarities between the isocircuits, circuit 101 occurs more frequently in E. coli than its isocircuit partner 011 (165 versus 53 circuits, respectively). Therefore, a systematic comparison of isocircuits could reveal the biological importance of the architectural difference between them.

To investigate the differences between these isocircuits we constructed discrete stochastic models and simulated them using the Gillespie algorithm [29, 30]. Simulations described production and degradation of mRNA and proteins of A, B and C species as discrete reactions and also accounted intrinsically for the stochastic behavior of binding and unbinding events of A and B transcription factors to the C promoter. For simplicity, we first considered here an AND logic for the regulatory input of A and B into the C promoter. Using these simulations we studied the behavior of the incoherent FFL isocircuits 011 and 101. The dynamics of the isocircuits during transitions between ON and OFF steady states (as defined by whether input A was absent or present, respectively) are consistent with previous literature and ODE simulations [2, 14, 16, 17]. However, when the circuits remained at OFF (Fig. 2.1B and 2.1F) or ON (Fig.2.1C and 2.1G) steady states, stochastic simulations revealed that both circuits were subject to stochastic bursts in C promoter expression (Fig.2.1B, 2.1C, 2.1F and 2.1G). These bursts are particularly prominent with slow binding kinetics of transcription factors to C promoter, but also occur with fast binding kinetics (see Methods Section 2.10.5). Simulations allowed us to precisely determine the amount of time each circuit resides stochastically in the C promoter state(s) that gives rise to high expression bursts. Therefore, we define here burst noise as the amount of time spent in the high expression state, which in turn determines the duration and amplitude of the observed C promoter expression bursts. We note that multiple binding or unbinding events of A and B to the C promoter can lead to the high expression

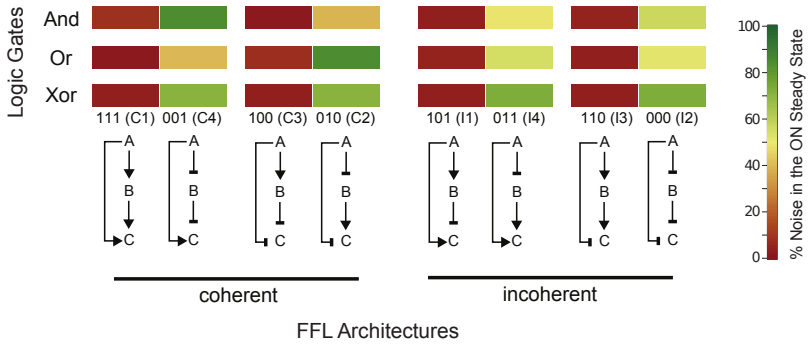
state from several equivalent low expression states. Therefore, C promoter expression bursts do not necessarily correspond to a single binding/unbinding event, but rather transitions between different expression levels. Even though each isocircuit displayed gene expression bursts, circuit 011 displayed burst noise in both the ON and OFF states, while circuit 101 exhibited higher noise, but only in the OFF state (Fig.1D and 1H).

What causes differences in C promoter expression bursts between the isocircuit steady states? In both circuits, bursts are generated by transient access to a high expression state of the C promoter due to stochastic binding and unbinding events of A and B transcription factors to the C promoter (Fig. 2.1). However, the amount of time spent in the high expression state of the C promoter is dictated by circuit architecture and is thus distinct between the two circuits. For example, in circuit 011, exit from the high expression state occurs by unbinding of either A or B. However, in circuit 101, exit from the high expression state occurs either by unbinding of A or binding of B. Binding reactions dependent on the concentration of the transcription factor and thus their rates can differ in ON and OFF steady states. Unbinding reactions on the other hand, are concentration independent. Therefore, differences in circuit architectures give rise to differences in noise behavior that can be defined by the mean durations of C promoter expression bursts in the ON and OFF steady states as described above (Fig.1D and 1H). Global parameter sensitivity analysis showed that these differences in noise behavior between isocircuits are consistently observed for a broad range of parameter values (2 fold change) as long as high concentrations of A and B transcription factors can effectively repress C promoter expression.

2.4.2 All Possible FFL Architectures Fall into Two Distinct Categories of Noise Behavior

Do architecture-dependent differences in noise profiles observed among isocircuit members 101 and 011 generally hold for all possible FFL architectures and even different logic inputs into the C promoter? To address this question we systematically performed discrete stochastic simulations for all possible eight FFL architectures and three binary input logics (AND, OR and XOR) for a total of 24 systems. Together, these simulations showed that the difference in noise profiles, taken as the proportion of mean burst duration in the ON state, observed between isocircuits 101 and 011 exists for all isocircuit pairs (Fig. 2.2). The percent noise of C promoter gene expression in the ON state appears to depend on circuit architecture, but surprisingly does not correlate with logic gates (Fig. 2.2). Regardless of logics, each isocircuit pair contains one circuit that has higher noise in its OFF steady state and one that does not. Global parameter sensitivity analysis once again showed that these results do not critically depend on parameter values. Together, these data suggest that architecture, but not logics at the C promoter, dictates the steady state noise profile of the output node C of FFLs. The steady state noise behavior of FFLs was observed to correlate with whether node A activates or represses node B. Specifically, simulations of all possible FFL circuits showed that when B is activated by A, noise in the OFF state is higher (Fig. 2.2). In contrast, for FFL circuits where B is repressed noise is similar in both the ON and OFF states. These results reveal a simple principle that appears to underlie the differences in steady state noise behavior of isocircuits and FFLs in general. When node A activates node B, the concentrations of A and B are correlated. Therefore, when the circuit is in the OFF steady state, both A and B are at low molecule numbers and thus subject to stochastic fluctuations. High noise in A and B in the OFF state in turn gives rise to higher noise in C promoter expression (Fig. 2.1F). Concurrently, when the circuit is in the ON state, A and B are at high concentrations and thus both are less noisy, which is again reflected in

Figure 2.2: FFL architecture determines noise behavior irrespective of logic gate. Shown are results from discrete stochastic simulations for all possible FFL circuit architectures and three Boolean logic gates (AND, OR, and XOR) for the regulation of the C output promoter. Each FFL is labeled according the three-digit binary nomenclature described in the main text as well as the standard classification. FFL architectures are paired as isocircuits and grouped according to the standard coherent and incoherent FFL classification [2]. Rectangles above each FFL circuit depict the proportion (%) of mean noise in the ON state divided by the total noise in the ON plus OFF states color coded according to the scale on the right. The amount of noise in ON and OFF states corresponds to durations of C promoter expression bursts as defined in Fig. 2.1. The noise pattern correlates with whether node B is activated or repressed by node A. Activation of B by A is associated with higher noise in the OFF state (left member of each pair), whereas repression of B by A is associated with higher noise in the ON versus OFF state (right member of each pair).



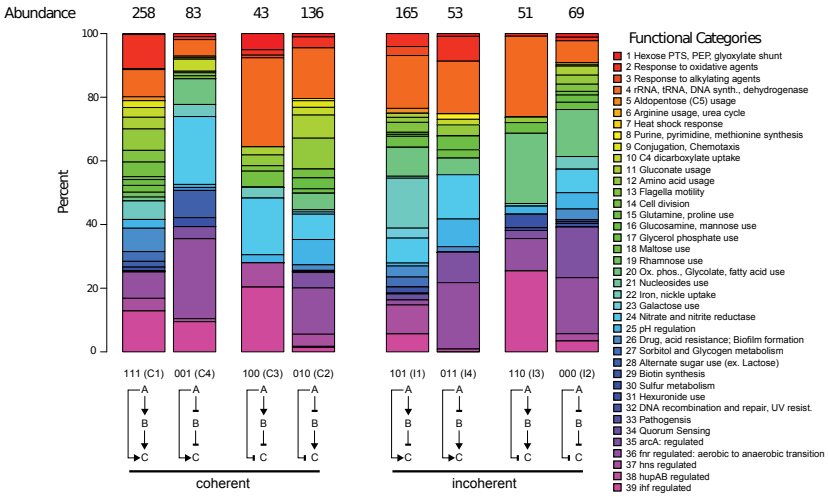
shorter durations of C promoter expression bursts (Fig. 2.1G). Therefore, in these circuits the OFF state will be noisy and the ON state will be quiet. However, in circuits where node A represses B, the concentrations of A and B vary oppositely. This inverse correlation in concentrations gives rise to an inverse correlation in noise of A and B. As a result, when the circuit is either in the ON or OFF state, only one of the two regulatory inputs into the C promoter is noisy while the other one is not. This inverse correlation distributes the noise across the ON and OFF states such that both states exhibit noise (Fig. 2.1B and 2.1C). Therefore, the mode of regulation of node B by A and not logics of C promoter regulation appears to dictate noise behavior at steady state, emphasizing the importance of FFL architecture.

2.4.3 Functional Profiles Discriminate Among FFL Architectures with Opposite Noise Behaviors

To determine if the distinct noise profiles of FFL isocircuits are of biological importance, we analyzed the well characterized and extensive data set of E. coli FFLs associated with various functional categories. This functional data set is comprised of 858 examples of FFLs grouped into 39 functional categories, obtained from the publicly available E. coli databases EcoCyc [31] and RegulonDB [32]. This data set contains examples of functional categories such as DNA synthesis that are regulated by all FFL architectures, as well examples such as biotin synthesis where only a specific FFL architecture (000) is assigned to it. Another difference among distinct FFL architectures and isocircuit members is that they vary in abundance. For example, while circuit 101 is more common (164 examples) and regulates a larger number of functional categories [20], its isocircuit counterpart 011 is less abundant (53 examples) and regulates fewer distinct functional categories [15] (Fig. 2.3). Therefore, each FFL architecture has a unique functional profile based on the number and categories of cellular functions assigned to it. Differences in these functional profiles suggest a relationship

between FFL architectures and their biological functions. Next we asked if the differences in functional profiles of FFL architectures are related to the differences in their steady state noise behavior. Specifically, we clustered all eight FFLs by functional categories (rows) and circuit architectures (columns), to determine which FFL architectures have similar functional profiles (Fig.2.3). Hierarchical clustering segregated FFL architectures into two groups with four circuits each (p ¡ 0.001 by Pearson G-test of Independence) using complete linkage and a Euclidean metric. Interestingly, FFLs within the same group exhibited similar noise behavior in stochastic simulations (Fig.2.3). In particular, FFLs with higher noise in the OFF state were grouped into one cluster and circuits with noise in their ON state formed the other. Therefore, clustering of functional profiles discriminated FFLs according to their architectures, since FFLs where node A activates B generate higher noise in the OFF state compared to those where A represses B (Fig.2.2). These results suggest that the functional profiles of FFLs contain information about the architecture-dependent noise behavior of these circuits at steady state. FFLs architectures are commonly classified based on whether the sign of direct and indirect pathways are the same (coherent FFLs) or opposite (incoherent FFLs). However, clustering of the functional data set did not discriminate circuits according to this classification. To investigate this issue further, we tested the robustness of clustering by analyzing how random perturbations to FFL functional profiles affect clustering (Fig.2.4A). Specifically, we repeatedly eliminated random subsets of functional categories and measured average clustering distances (linkage) among FFLs with opposite noise behavior and compared them to average distances among circuits with similar noise behavior (Fig.2.4A). A ratio greater than one was obtained 81% of the time, which indicated that distances among FFLs with similar noise profiles were consistently closer. Clustering of FFL circuits according to noise appears to be robust to random elimination of functional categories (Fig.2.4A). We then applied the same method to compare average clustering distances across

Figure 2.3: Functional profiles segregate FFLs according to architecture-dependent noise behavior. Matrix representation of the abundance (log color scale) of each FFL architecture for a given functional category in E. coli arranged according to two-dimensional hierarchical cluster analysis. Columns represent distinct FFL architectures, whereas functional categories are shown in rows. The dataset is comprised of a collection of 858 FFLs identified in E. coli and categorized into 39 distinct classes of biological functions according to the publicly available E. coli databases EcoCyc and RegulonDB. Dendrograms obtained from clustering were color coded to emphasize distinct clusters of FFL architectures (red and green) and three distinct groups of functional categories, orange (group 1), gray (group 2), and magenta (group 3). The FFL architecture clusters correspond to the predicted noise profiles shown in Fig. 2.2 and are accordingly colored green and red. For easier comparison, the noise results shown in Fig. 2.2 are depicted underneath respective circuit architectures.



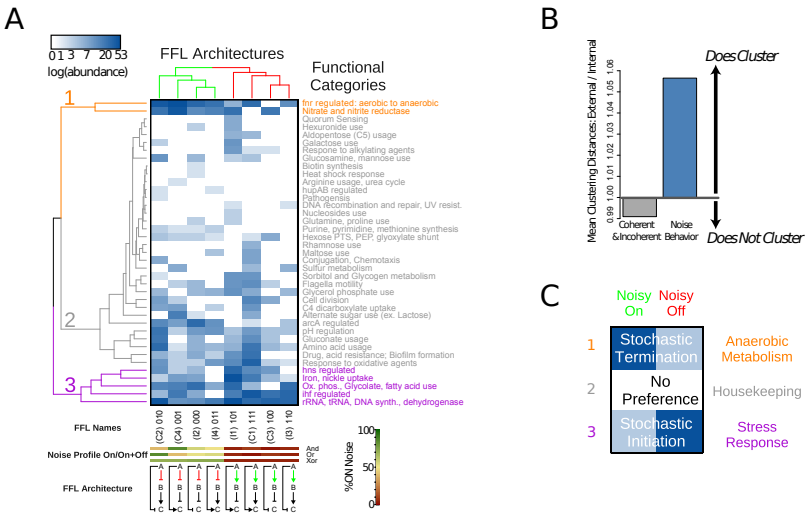
and within incoherent and coherent FFLs. Functional profiles systematically failed to cluster coherent/incoherent FFLs, as only 35% of iterations gave a ratio greater than one (Fig.2.4A). Together, these data show that functional profiles reliably segregate FFLs according to architecture-dependent noise, suggesting that this property is of biological importance.

2.4.4 Demands on Biological Processes Correlate with FFL Noise Behaviors

Cluster analysis more specifically revealed three groups of functional categories that diverged in their preference for FFL architectures with distinct noise profiles (p ¡ 0.001) (Fig.2.3 and 2.4B): Group#1 was enriched for FFL architectures with high noise in their ON states. Group#2 did not display a preference based on noise behavior. Finally, group#3 preferred FFL architectures that generate higher noise in their OFF state. Even though group #1 and #3 combined only constitute 18% of all functional categories, they are associated with 70% of the categorized FFLs. Therefore, we find that the number of FFLs is not evenly distributed across functional categories. Together, these results show that most of the FFLs identified in E. coli are involved in the regulation of a few functional categories that in turn appear to select for circuits based on their architecture-dependent noise properties.

What accounts for the enrichment of FFL architectures with specific noise profiles in functional groups #1 and #3? Many FFLs associated with these functional groups contain an interacting pair of global and functionally specific transcription factors, consistent with the hierarchical organization of gene regulatory circuits [4]. Specifically, many FFLs in group#1 contain the global regulator fnr (fumarate nitrate reductase) as their A node and a more specific transcription factor narL (nitrate reductase) as their B node, that together regulate E. coli metabolism under anaerobic conditions [33, 34]. An example of such a FFL circuit is shown in Fig.2.4C, where fnr regulator is

Figure 2.4: Robust clustering of FFL functional profiles reveals two noise-based cellular strategies. (A) Barchart of the percent of bootstrap samples (total = 100,000) that cluster according to noise (blue) or the traditional coherent/incoherent classification (gray) (SI Appendix Section 2.2.7). (B) Reduced matrix representation of the cluster analysis result and color coding depicted in Fig. 2.3. (A). Shown are the mean abundances of the two clusters of FFL circuit architectures and the three clusters of functional categories indicated in Fig. 2.3. Functional group 1 (orange) is comprised of categories associated with anaerobic metabolism and contains 30% (259/858) of FFLs. Functional group 2 (gray) contains, among others, housekeeping genes and the most number of functional categories (32/39), but contains only 54% (467/858) of FFLs. Functional group 3 (magenta) contains 5/39 functional categories and 62% (477/858) of FFLs that are in general related to stress responses. (C) Representative time traces from simulations depicting predicted stochastic termination and initiation expression patterns of genes dcuB (Top) and glpT (Bottom), respectively, according to the FFL architecture shown (Right).



active in the absence of oxygen and represses narL, and together these transcription factors regulate dcuB, the C4-dicarboxylate transporter necessary for anaerobic growth. Discrete stochastic simulations predict that the repression of narL (node B) by fnr (node A) gives rise to a noisy ON state expression of dcuB (node C) (Fig.2.4C). Noise in the ON state can result in the stochastic termination of dcuB expression during anaerobic growth (Fig.2.4C). This propensity for stochastic termination could permit occasional sampling of aerobic respiration that could be beneficial if environmental conditions unexpectedly change. Additionally, the noisy ON state of FFLs associated with group#1 could provide a mechanism to modulate the expression level of downstream targets of fnr and narL possibly through frequency modulation of stochastic bursts [35]. Therefore, enrichment of FFL architectures that give rise to noisy ON states may be a result of functional requirements associated with anaerobic respiration processes that make up functional group#1. Most FFLs in group#3 are associated with known stochastic processes in E. coli that respond to stress and give rise to heterogeneity. While group#1 and group#3 both exhibit a preference for FFL architectures based on noise, they do so in an opposite manner. In particular, the functional categories comprising group#3 are enriched for FFL architectures where node A activates node B and thus gives rise to high noise in the OFF state of node C. Concurrently, group#3 FFLs share a pattern where a global regulator such as IHF (integrative host factor) positively regulates a more functionally specific transcription factor such as fis (factor for inversion stimulation). Together, IHF (node A) and fis (node B) regulate genes such as glpT (glycerol-3-phosphate transporter) (node C) that have been implicated in the generation of resistance to antibiotics like fosfomycin (Fig.2.4C) [36]. IHF has also been identified in a screen for genes involved in the stochastic generation of antibiotic resistant persister cells [28]. In FFLs with this architecture, expression of glpT is predicted to be noisy in the OFF state, giving rise to stochastic bursts of glpT expression. Together, these data reveal that group#3 is comprised

of biological processes associated with stress responses that can be initiated in a stochastic manner even in the absence of stress stimuli. Consistent with this finding, the architecture of FFLs enriched in group#3 exhibit higher noise in their OFF state that can facilitate stochastic activation and thus sampling of alternative stress responses at the single-cell level. Such probabilistic behavior at the single-cell level has been shown to be a beneficial strategy under unpredictable environmental conditions and may thus explain the preference of group#3 functions for FFLs with high noise in the OFF state [7, 13, 36, 37, 38]. Functional categories comprising group#2 exhibit a lack of preference for FFLs architectures according to noise. Even though group#2 contains 82% of functional categories, only 54% of FFLs are associated with this group. FFLs within group#2 may of course have been selected for based on properties other than noise. However, the low number of FFLs contained within group#2 suggests that noise behavior is at least one of the important properties underlying FFL function. If noise behavior is not critical for the operation of biological processes contained within group#2, perhaps FFLs are a less preferred circuit motif. Interestingly, the functional categories comprising group#2 are among others, associated with housekeeping processes such as cell division and amino acid synthesis. These basic biological processes are not known to benefit from stochastic behavior. Together, the three groups of functional categories suggest that most FFLs are selected for based on their noise properties by cellular processes that may benefit from stochastic fluctuations. zo

2.5 Discussion

The comprehensive analysis presented here reveals a general trend regarding the preference of biological processes for FFL architectures based on their noise profiles. However, the following points have to be considered: 1) there may be additional unaccounted regulatory inputs into FFLs

other than the three nodes considered here. Such additional regulatory inputs could alter the noise behavior of any given FFL circuit. 2) it is important to note that FFLs do not exist in isolation. Genes can be shared among FFLs and there can be cross-regulation between individual circuits. 3) robust clustering of FFLs according to noise does not imply that other differences among FFL architectures cannot be of functional importance. It is therefore striking that functional profiles containing information on biologically relevant properties robustly cluster FFLs consistent with architecture-dependent noise behavior. Stochastic behavior thus appears to be at least one of the important biological properties of FFLs.

Many gene regulatory circuits contain pathways comprised of consecutive activation and repression reactions similar to those in FFLs. Therefore, the noise behavior of other gene regulatory circuits may also be determined by the order of regulatory links with opposite actions. Specifically, noise in target gene expression that is governed by a net negative linear cascade of transcription factors will depend on whether this regulation is mediated by the repression of an activator or the activation of a repressor. Regulation will therefore either be mediated by a high concentration of repressor, or a low and thus noisy concentration of activator. For example, the order of activation and repression reactions comprising a net negative feedback loop of a bacterial differentiation circuit has been shown to dictate stochastic fluctuations in circuit dynamics [7]. Therefore, in instances where gene expression is regulated by a net negative cascade of transcription factors with opposite regulatory modes, noise of gene expression may depend on the order of activation and repression reactions.

Susceptibility to stochastic bursts of gene expression may of course not be the only functionally relevant property of FFLs that gives rise to clustering according to architectures. We considered another possible explanation for the observed clustering pattern known as demand theory

[39, 40]. Demand theory predicts that depending on the organisms native environment, genes in high demand are regulated by activators whereas genes in low demand are regulated by repressors. Since functional groups#1 and #3 display opposite clustering preferences where node A either represses or activates B respectively, demand theory would predict that node B is in low demand in group#1 while it is in high demand in group#3. Group#1 FFLs are involved in anaerobic metabolism and thus expected to be in high demand in the native environment of E. coli such as the mammalian colon. However, transcriptional regulators corresponding to node B such as narL and nikR are more often repressed than activated in group#1 FFLs. FFLs in group#3 on the other hand are involved in stress responses and thus expected to be in low demand assuming that E. coli is well adapted to its environment. In particular, stress response regulators that correspond to node B such as baeR, HN-S, and fis should be in low demand and thus expected to be repressed according to demand theory. However, these transcriptional regulators are found to be activated in group#3 FFLs. Therefore, rather than demand theory, the noise behavior of circuit architectures is more consistent with clustering of FFLs.

Stochastic bursts of gene expression have been demonstrated to be physiologically important for the many systems such as lambda and Lac repressors as well as the differentiation of Bacillus subtilis cells into the state of competence [6, 20, 41, 42, 43, 44, 45]. Consistent with these findings, FFL architectures that generate stochastic bursts of gene expression appear to be favored by E. coli stress responses such as the stochastic generation of antibiotic resistant persister cells [28, 46]. In addition to stochastic initiation, our results also suggest that some cellular processes such as anaerobic metabolism of E. coli, may prefer the ability to stochastically terminate their response. These data indicate a possible new relationship between the default state of a cellular process and the noise behavior of the associated FFLs. Depending on whether the default state of the cellular process

is active (ON) or inactive (OFF), FFLs with higher noise in ON or OFF steady states can enable sampling of alternative states by stochastic termination or activation respectively. These findings suggest a possible a link between the demand on a cellular process, and the type of noise generated by the associated FFL architecture. Furthermore, these results suggest that particular architectures of FFLs may have been selected for their inherent noise properties to cope with distinct environmental constraints. It may thus be possible to decode functional properties and selection pressures from architectures of gene regulatory circuits.

2.6 Methods

2.6.1 Programming Language and Statistical Computing Environment R

Version 2.9.1 of R from The R Foundation for Statistical Computing ISBN 3-90051-07-0 was used as distributed in the Debian GNU/Linux package r-base version 2.9.1-2. R was used to run stochastic simulations and analyze data.

2.6.2 Stochastic Simulation Software: GillespieSSA

Version 0.53 of the GillespieSSA package for R by Mario Pineda-Krch was used for stochastic master equation simulations [47].

2.6.3 Feed Forward Loop Simulations

The simulation environment was constructed in R as a wrapper around GillespieSSA. The three components of the FFLs, A, B, and C, each were modeled as genes with corresponding transcriptional promoters and translated protein species. The protein products for A and B then served as transcription factors for downstream genes as shown in Fig. 2.1. Transcription and translation were modeled explicitly as a single step with propensity determined by the binding state of tran-

scription factors to the promoter as opposed to using a cis regulatory input function. Transcription factor binding was modeled with a Hill coefficient of 2.

2.6.4 Databases: RegulonDB and EcoCyc

Genetic transcriptional network information was downloaded from RegulonDB [32] from the Regulatory Network Interactions section of datasets as NetWorkSet.txt, Version 6.3, released on January 30, 2009. Genetic regulatory interactions were consistent with EcoCyc [31] due to data sharing between the two databases. We excluded ambiguous or unknown interactions from our analysis, but did not exclude microarray or electronically inferred interactions.

2.6.5 Gene Annotation

Genes in FFLs were identified using FANMOD [48]. Functional annotation and grouping was based on that of Ma et al. [49]. Further gene classification was done with the assistance of EcoCyc. All genes involved in a feed-forward loop were annotated, and analysis was done both by considering three annotations per feed-forward loop (Fig. 2.3).

2.6.6 Cluster Analysis

With R, hierarchical clustering was applied to both functional categories and FFL types using a Euclidean distance metric and complete linkage based on the abundance of the number of FFLs. Bootstrap resampling analysis was done using the boot package available from CRAN [50].

2.7 Impact on Literature

Biological feed-forward loops may not solely involve genes and protein transcription factors, but could also include other molecules such as a miRNA that can also modulate gene activity,

involve only proteins, or even be networks of neurons within the central nervous system. In this work, we specifically examined the effect of architecture, or topology, on stochastic gene expression by quantifying bursts of transcripts produced by transcription factor binding and unbinding. We found that the topology of the FFLs affected the copy number of regulatory transcription factor molecules available. This created differences in the noise profiles in the ON and OFF states of the networks corresponding to stimulated or unstimulated expression of the input transcription factor.

An article that also involved feed-forward loops also investigated another implementation that used microRNAs for post-transcriptional control along with transcriptional control as in a traditional GRN [51]. There they saw that the 101 (I1) FFL was significantly more noisy as measured by the coefficient of variation and attributed the difference to binding and unbinding of transcription factor, LacI, as was examined in this work.

Other research took note of the differences in the noise profiles in the ON and OFF states [52, 53] and noted the importance of correlation or anti-correlation in transcription factor copy numbers [54]. Correlation and anti-correlation of molecule copy numbers is quite general and can introduce state-dependent noise into other biochemical networks. If there are degeneracies in network implementations, the use of correlated or anti-correlated relationships may have significance with regard to the required noise properties for a particular function [55]. Furthermore, Koh et al. noted how different stochastic profiles related to architecture may affect evolution if they control a selectable trait such as persistence in *E. coli* [56].

The extent of influence on stochastic properties by network topology should not be overstated, however. Other parameters such as binding kinetics could have a significant impact upon the output of a network. Lok-Hang So et al showed in a survey of 20 different bacteria promoters that transcriptional burst size was gene-independent and rather had a strong dependence on mean expres-

sion level [57]. In particular, the gene expression level was modulated by use of the transcriptional off rate, k_{off}. Essentially this means that higher expression is achieved by having longer bursts of expression while lower expression occurs when bursts are kept short in duration.

There the authors argued that gene-independence also implied that transcriptional noise was also independent of topology since the coefficient of variation did not depend on what transcription factors affected a particular gene's promoter. The supplemental material should be noted in that 15 of the 20 promoters studied were mutational variants of each other which weakens the gene and topology independence argument. Furthermore, the effect of topology on noise in this work on FFLs was via the copy number of transcription factors affecting the final output gene. High molecular copy number could extend a positive transcriptional burst in the case of activation or shorten it in the case of repression by modulating the apparent k_{off} rate of a gene. Low copy number would effect positive transcriptional bursts in the opposite manner. In fact, the noise measurement used here in this FFL study was mean burst duration based upon transcription factor binding state. Burst duration from the view of gene activity may not be able to resolve two binding state fluctuations in quick succession. If I examine the 011 (I4) and 101 (I1) FFLs shown in Figures 2.1 and 2.9 I observe that the 011 FFL has many bursts in quick succession during the ON state that could be considered a single burst. Additionally, the extended time traces in Figure 2.9 show that the 011 FFL in the ON state has a higher mean expression level to correspond. The other states seem to have mean expression levels that correspond with their mean burst durations. While mean burst duration and mean expression may be linked, this does not necessarily exclude the effect that topology may have on both.

Several reviews noted the joint impact that this work and work in the lab by Çağatay et al.[7, 58, 59]. Çağatay demonstrated that an alternate circuit topology that switched the order of repression and activation in a negative feedback loop produced deterministic behavior in contrast to

the stochastic behavior observed in the original circuit. As with the feed-forward loops, the deterministic synthetic circuit involved high copy numbers of the molecules that suppressed stochastic behavior when stimulated. This work on FFls demonstrated that the relationship between architecture and copy number could be generalized to predict noise behavior. Research on the synthetic alternate circuits with different noise properties is continued in Chapter 3.

2.8 Materials and Methods

2.8.1 Computational Environment

2.8.1.1 R

Version 2.9.1 of The R from The R Foundation for Statistical Computing ISBN 3-90051-07-0 was used. The Debian GNU/Linux package *r-base* version 2.9.1-2 was used. R is also available directly from CRAN, *http://cran.r-project.org*. R was used to both run stochastic simulations and analyze data.

2.8.1.2 GillespieSSA

Version 0.5-3 of the GillespieSSA package for R by Mario Pineda-Krch was used for computational simulations and downloaded from CRAN as noted above.

2.8.1.3 Feed Forward Loop Simulations

The simulation environment was constructed in R as a wrapper around GillespieSSA. The three species A, B, and C each had corresponding transcriptional promoter and translated protein species. Promoter complexes with regulatory proteins bound were simulated as a distinct and discrete species and not as genetic regulatory functions. Additional versions of the simulation were conducted to include mRNA transcripts and post-translationally activated protein units. These additional simulations are included here, but are were not included in the simulation as published.

2.8.2 Informatics

2.8.2.1 RegulonDB

Genetic transcriptional network information was downloaded from RegulonDB from the *Regulatory Network Interactions* section of datasets as NetWorkSet.txt Version 6.3 released on January 30th, 2009.

2.8.2.2 EcoCyc

Genetic regulatory interactions were confirmed using EcoCyc.

2.8.2.3 Data Processing

Subsequent text processing was done using GNU Awk and GNU Sed. All genes involved were mapped to a unique positive integer for use in FANMOD. A three column tab delimited file was created listing the number of the gene that encodes the regulator, the number of the gene being regulated, and the *color* of the regulation modality (induction or repression).

2.8.2.4 FANMOD

FANMOD, a tool for fast network motif detection was downloaded from *http://theinf1.informatik.uni-jena.de/~wernicke/motifs/index.html* and was downloaded on January 30th, 2009. The version used was last updated on December 23rd, 2006.

2.8.2.5 Annotation

Functional annotation and grouping was based on that of Ma et al. in Ma2004. Additional genes were categorized by referencing EcoCyc. All genes involved a feed forward loop were annotated and analysis was done both by considering three annotations per feed forward loop and by examining the the annotations for the A, B, and C specie classes individually.

2.8.2.6 Cluster Analysis

Hierarchical clustering was done using a Euclidean distance metric and complete linkage based on the raw abundance.

2.8.2.7 Bootstrap Methods

Bootstrap analysis was used to both evaluate the robustness of clustering and to identify functional subsets that were able to achieve various clustering patterns. Bootstrapping involved taking randomly weighted subsets of functional categories and evaluating the ratio of the mean external distances between each cluster and the mean internal distances within each cluster. Bootstrap analysis was done using the *boot* package available from CRAN.

2.9 Feed Forward Loop Nomenclature

2.9.1 Coherent / Incoherent Nomenclature

The naming scheme introduced by Mangan et al. split feed forward loops into two groups: coherent and incoherent. Coherent indicates that the direct and indirect regulation of C, the output node, by A, the input node are the same, whereas incoherent indicates that the type of regulation is different. The naming scheme then numbers the four coherent and four incoherent loop in a mostly arbitrary manner: C1-C4, I1-I4.

2.9.2 Binary Nomenclature

Rather than continuing to use the coherent and incoherent nomenclature which places emphasis on presumed comprehensibility of the circuits, we internally chose to use a binary nomenclature that describes the actual structure of the feed forward loop. We opted to later use this nomenclature in writing because we found this naming system made it easier in our own discussions to see patterns.

The binary nomenclature indicates the type of relationship between the three species. The first digit indicates if A represses (0) or induces (1) B. The second digit indicates the regulation of

Coh / Inc	Binary	Binary	Coh / Inc
C1	111	000	I2
C2	010	001	C4
C3	100	010	C2
C4	001	011	I4
I1	101	100	C3
I2	000	101	I1
I3	110	110	I3
I4	011	111	C1

Table 2.1: Table to translate between coherent and incoherent nomenclatures introduced by Alon et al. and binary nomenclature introduced here

C by B. The final digit indicates the direct regulation of C by A.

2.9.3 Translation Table

To facilitate the use of the binary nomenclature with prior literature we offer the following table as a translation. Whether a loop is coherent or incoherent can be derived from the binary nomenclature by a boolean relationship.

From the binary nomenclature it is possible to determine whether a loop is coherent or incoherent. Considering 0 to be FALSE and 1 to be TRUE, a coherent loop is one where the equality of the first two numbers is reflected by the third number. For example, in the 001 loop, the first digit, 0, equals the second digit, 0, and the third digit, 1, indicates that this equality is TRUE.

From the coherent and incoherent nomenclature it is impossible to determine the binary nomenclature completely without a reference. Interestingly, an even numbered circuit (C2, C4, I2, I4) indicates that A represses B, whereas an odd numbered circuit (C1, C3, I1, I3) indicates that A activates B. This is convenient because our work shows that the regulation of B by A is important in determining the distribution between the state where A is ON versus where A is OFF.

2.10 Stochastic Modeling

2.10.1 Coupled Transcription-Translation Simulations

The simulations included in the paper assumes that transcription and translation are coupled as in prokaryotes or in some circumstances in eukaryotics (i.e. nuclear ribosomes). These simulations therefore do not address whether the bursts are necessarily transcriptional in nature (a single promoter state change generates many mRNA transcripts from which many proteins are translated) or translational in nature (a single promoter state changes produces one or few mRNAs which are translated into many proteins). Transcription and translation are simulated as a single step with a single rate constant dependent on the activity of the state such that the number of proteins created is proportional to the time spent in a particular promoter state.

In considering all of the sixteen possible boolean integration schemes for the effect of A and B on C, only the AND, OR, and XOR boolean gates were relevant. The remaining schemes were either trivial such as TRUE, C is always active regardless of the binding state of A and B to its promoter region, or already represented in a different feed forward loop, such as NAND (NOT AND) which would turn a 111 loop into a 100 loop.

2.10.1.1 Reactions

$$A \underset{\text{trans}}{\overset{\text{degrade}}{\rightleftharpoons}} \emptyset \qquad A + B_p \underset{\text{unbind}}{\overset{\text{bind}}{\rightleftharpoons}} AB_p \qquad A + C_p \underset{\text{unbind}}{\overset{\text{bind}}{\rightleftharpoons}} AC_p \qquad A + BC_p \underset{\text{unbind}}{\overset{\text{bind}}{\rightleftharpoons}} ABC_p$$

$$B \underset{\text{trans}}{\overset{\text{degrade}}{\rightleftharpoons}} \emptyset \qquad B + C_p \underset{\text{unbind}}{\overset{\text{bind}}{\rightleftharpoons}} BC_p \qquad B + AC_p \underset{\text{unbind}}{\overset{\text{bind}}{\rightleftharpoons}} ABC_p$$

$$C \underset{\text{trans}}{\overset{\text{degrade}}{\rightleftharpoons}} \emptyset$$

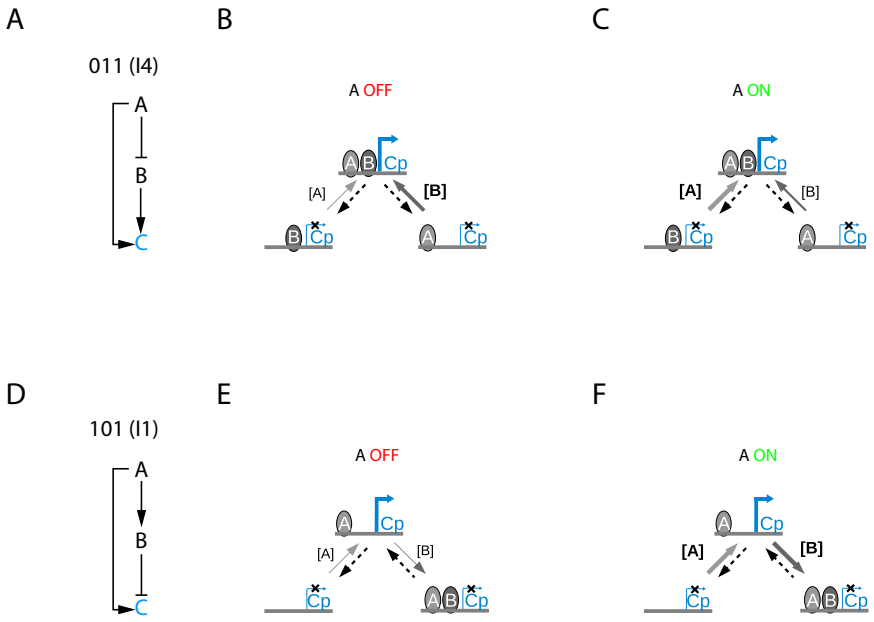


Figure 2.5: Mechanism of Burst Initiation for the and011 and and101 FFLs. The elevated promoter binding states indicates the transcriptionally active state. Bursts are initiated by an unbinding (dashed arrows) or binding (solid arrows) of transcription factors A or B depending on the circuit architecture as indicated by the arrows pointed to the active binding state. The burst duration is due to the amount of time it takes for one of two binding or unbinding events to take place as indicated by the arrows oriented downwards.

2.10.1.2 AND Boolean Logic Gate

The AND Logic Gate functions such that only one state is the transcriptionally-active promoter state. For the 111 loop, this means that both A and B must be bound for active transcription to occur. The main difference between the possible AND Logic gate schemes is whether the active state is achieved by transcription factor binding, unbinding, or a combination of both.

2.10.1.3 OR Boolean Logic Gate

The OR Logic Gate is similar to the AND logic gate in that only one of the four transcriptional states are different than the other three. In this case, however, there is only one transcriptionally-*inactive* state. For the 111 loop, this is the state where neither A nor B are bound to the promoter for C.

2.10.1.4 XOR Boolean Logic Gate

The XOR logic gate can be thought of as a subtraction of the AND logic gate from the OR logic gate. That is for the 111 loop, the promoter is only transcriptionally active if either A or B are bound but not both. This arrangement is different than the other two because any stochastic binding or unbinding event will result in a change of the transcriptional activity of the C promoter. Furthermore, with XOR integration logic, two states are transcriptionally-active and two states are transcriptionally-inactive. Also, XOR logic causes two of the FFLs to degenerate to be exactly the same loop. The 100 loop is transcriptionally active if either A or B are not bound but at least one must be bound. This is equivalent in every way to the 111 loop with XOR logic.

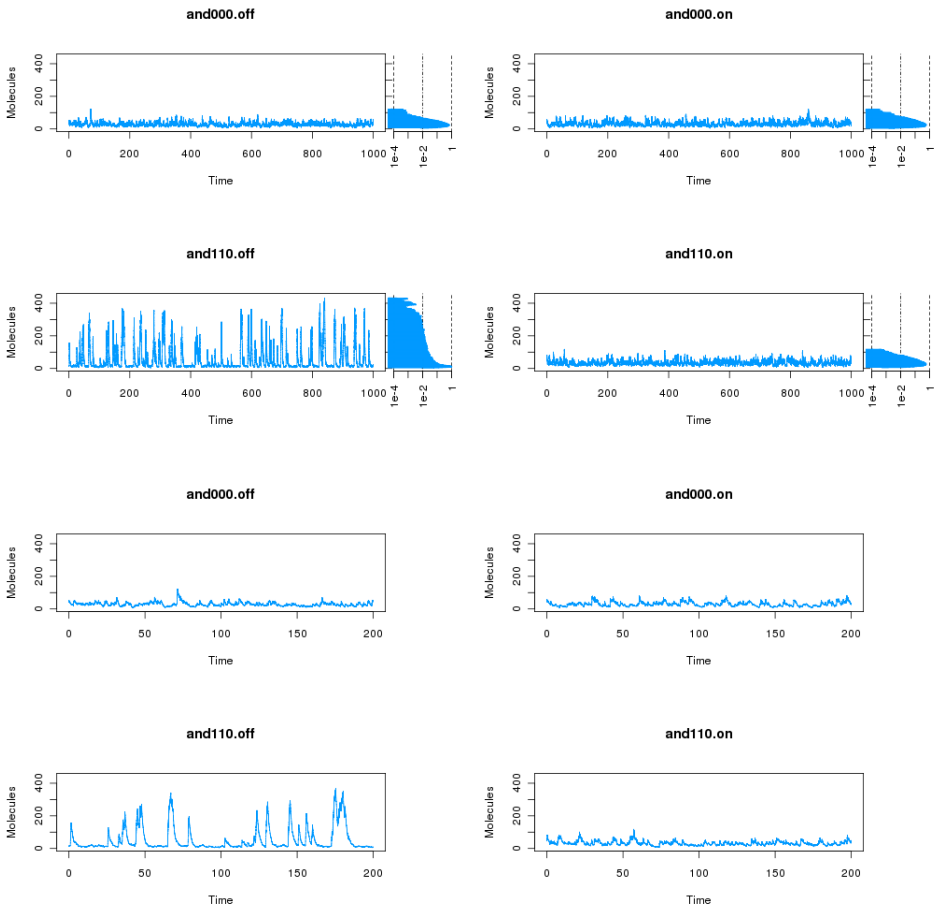


Figure 2.6: and000 vs and110

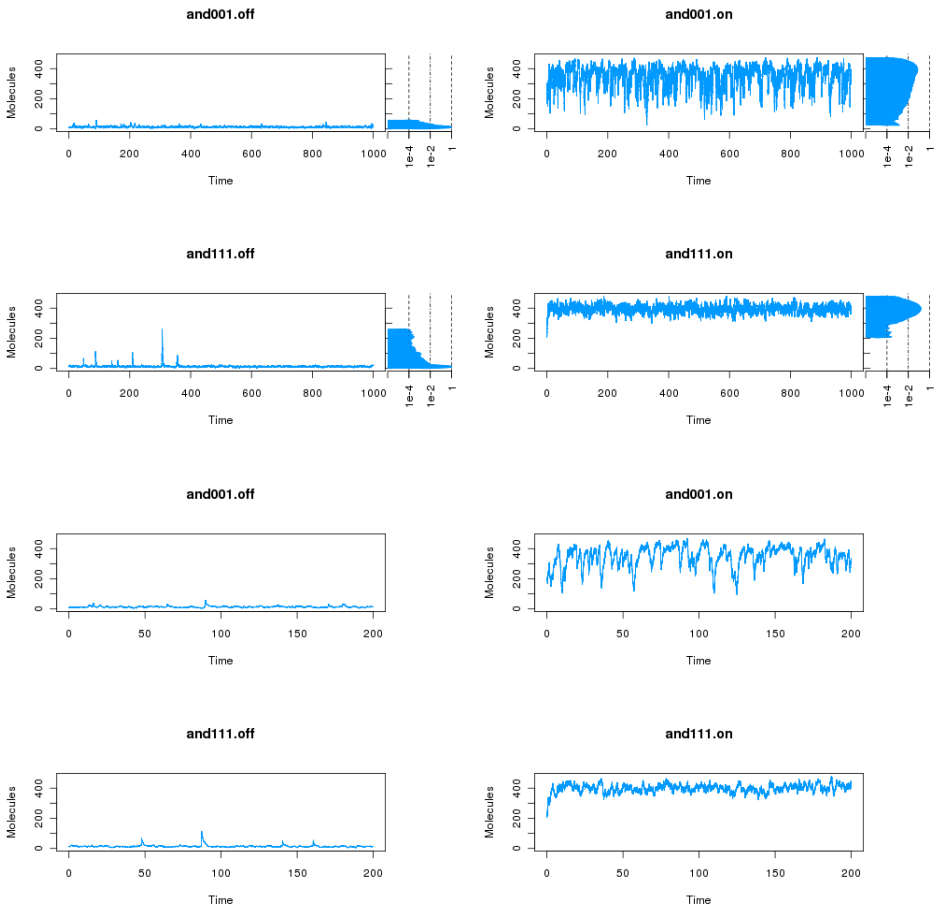


Figure 2.7: and001 vs and111

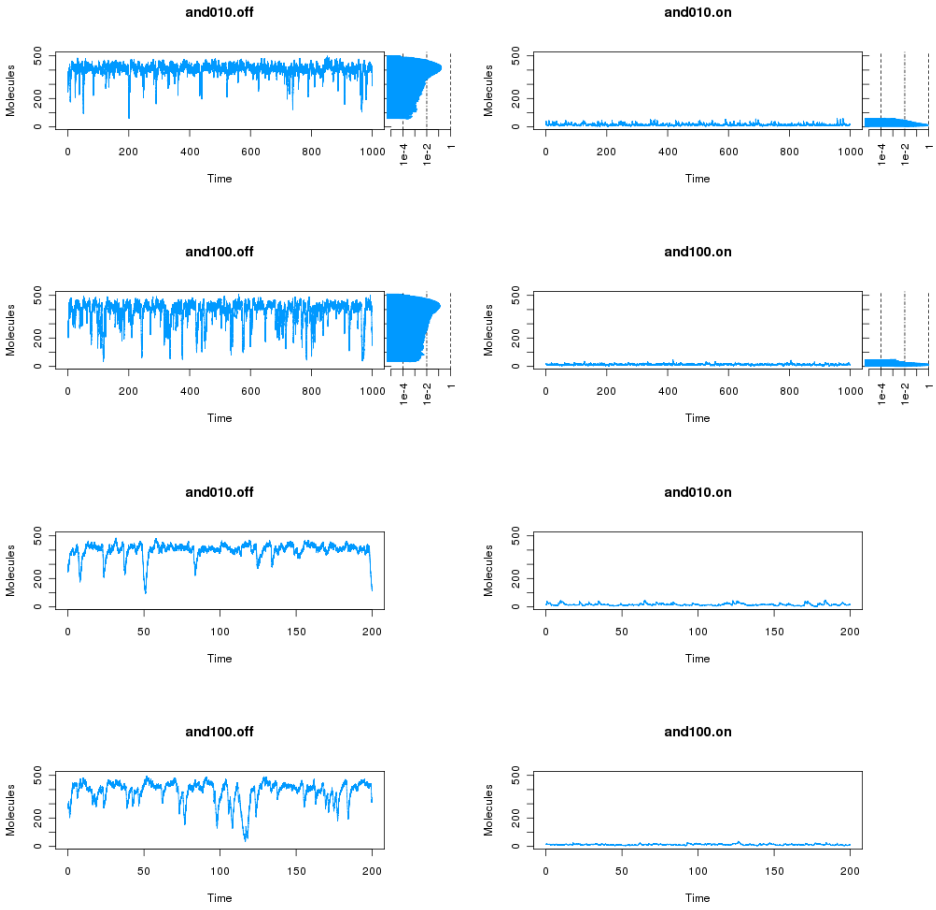


Figure 2.8: and010 vs and100

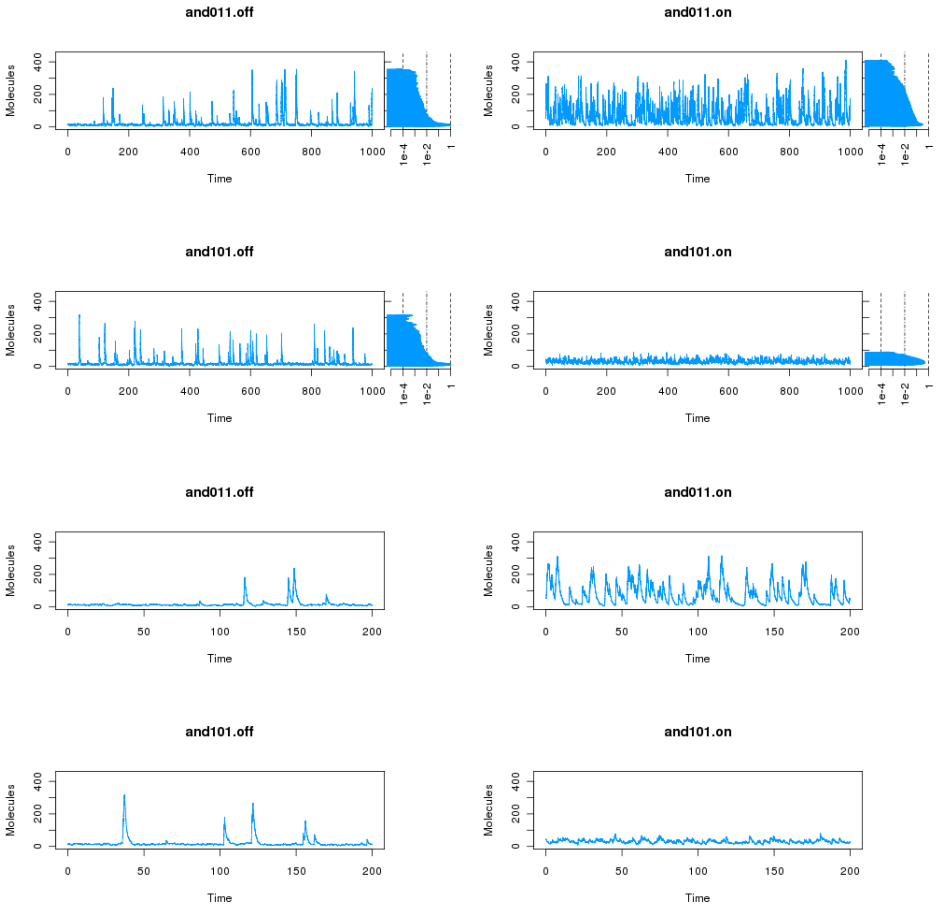


Figure 2.9: and011 vs and101

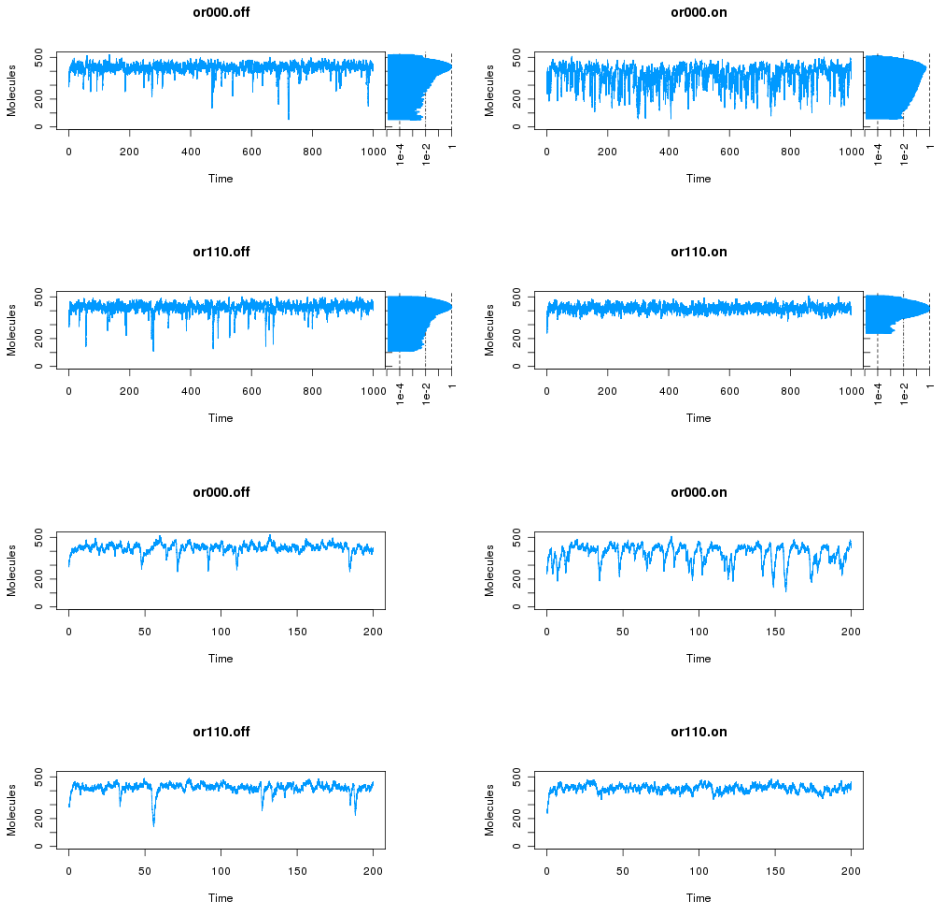


Figure 2.10: or000 vs or110

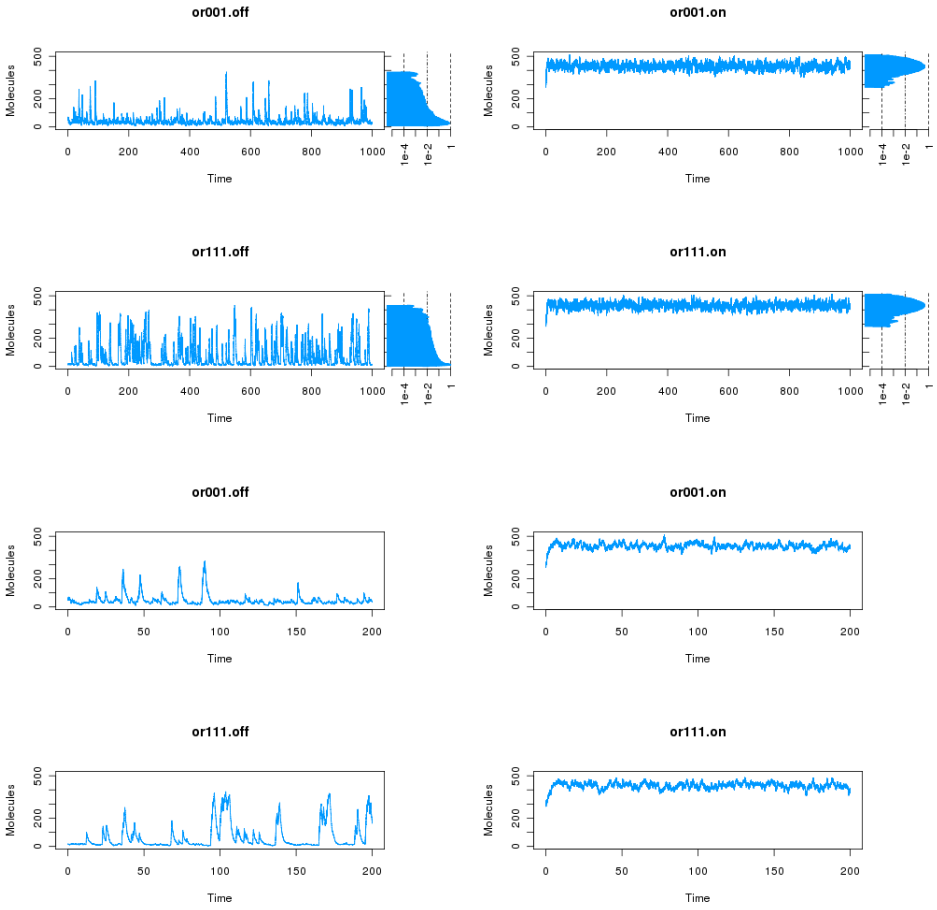


Figure 2.11: or001 vs or111

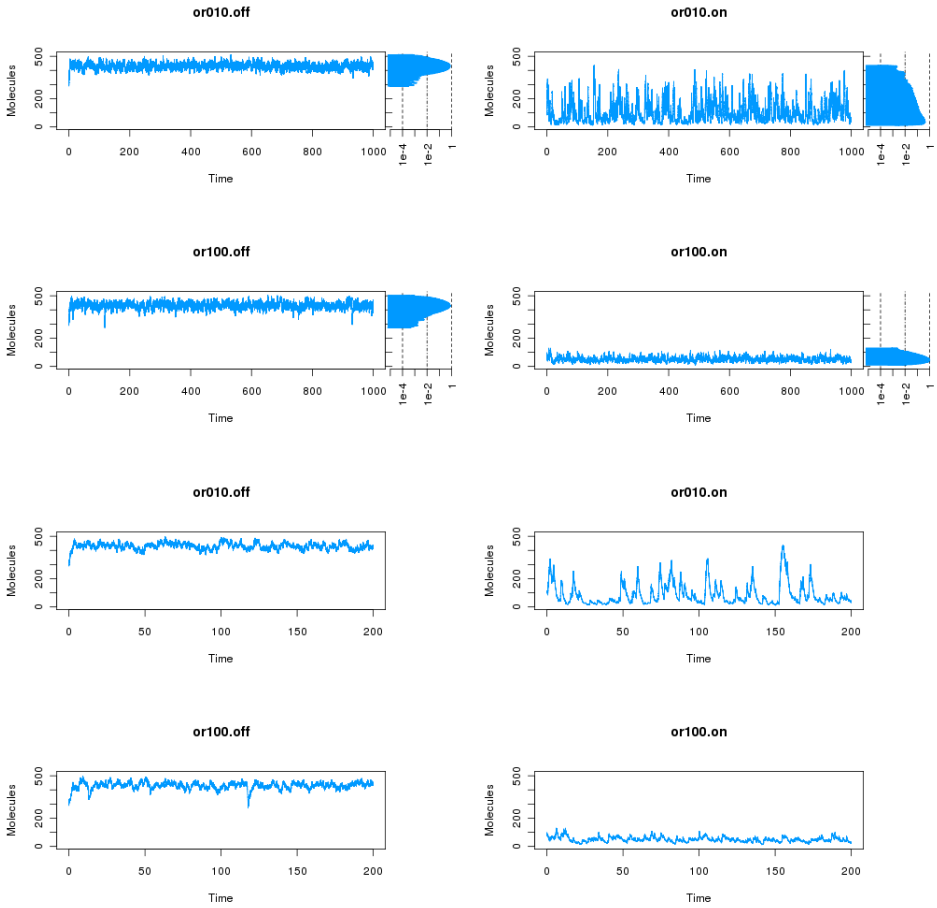


Figure 2.12: or010 vs or100

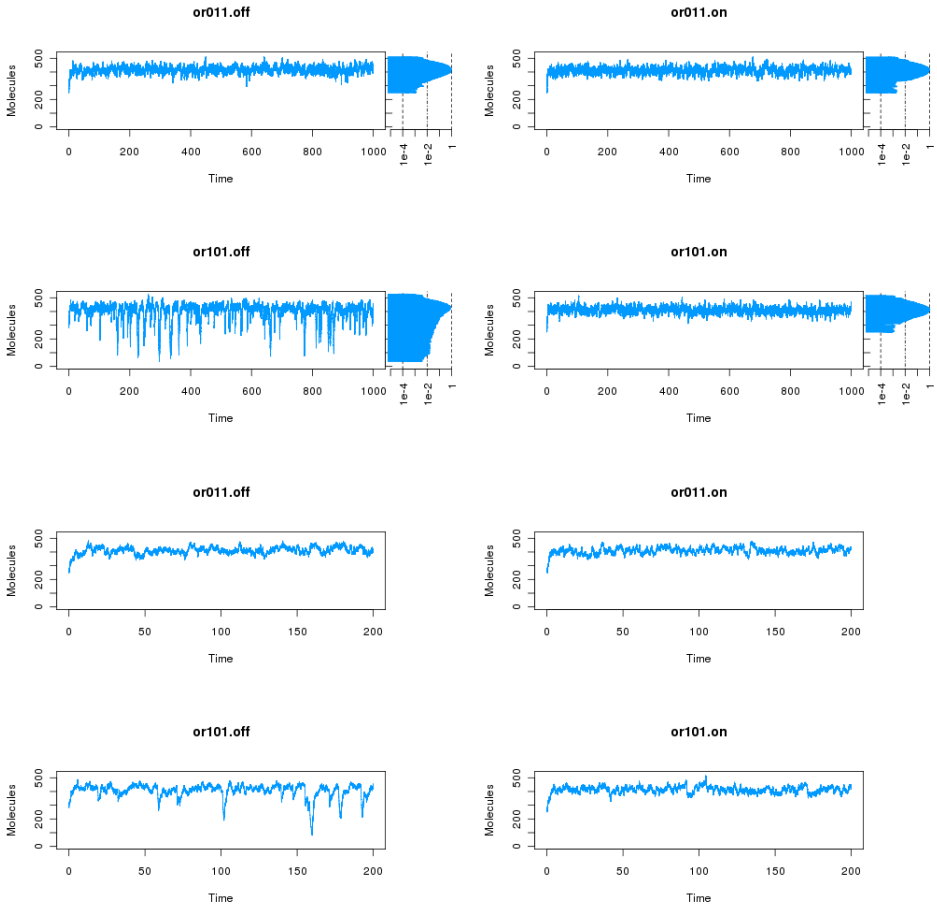


Figure 2.13: or011 vs or101

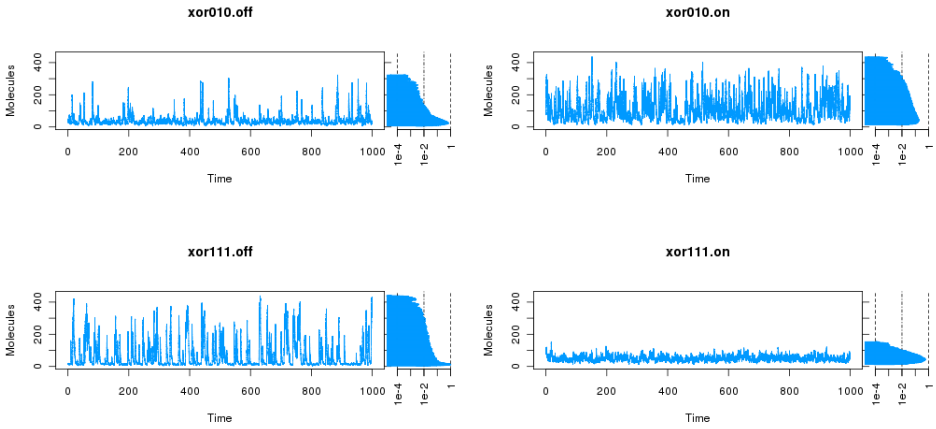


Figure 2.14: xor010 vs xor111

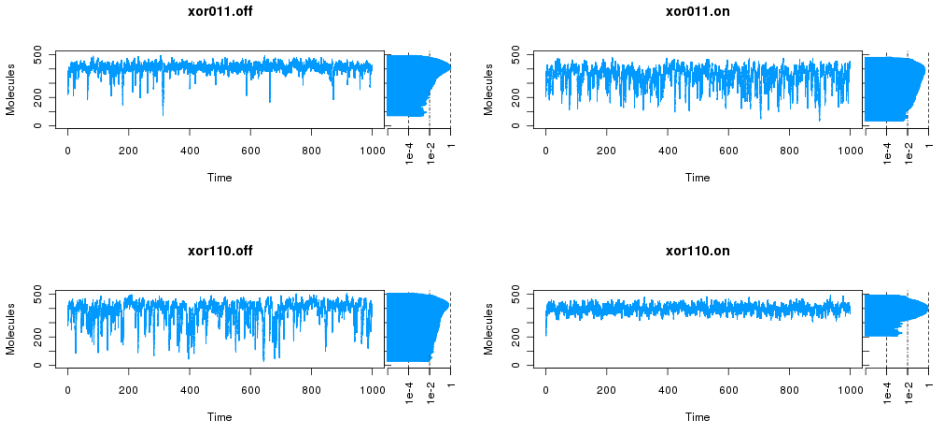


Figure 2.15: xor011 vs xor110

2.10.2 Noise Analysis

In order to analysis and quantify the noise, we focused on measuring noise relating to bursts as opposed to that coming from birth-death processes surrounding the mean. We initially investigated using the coefficient of variation, a standard measure of noise, as a metric but found the the C.V. was more apt for normally distributed noise. We subsequently turned to metrics more closely associated with properties of the burst. As others have found [23], the duration of bursts approximates a exponential distribution. More precisely, we found that each of the four distinct promoter states had dwell times that followed exponential distributions, and that the dwell times of burstable states were linear combinations of exponential distributions. The exponential distribution is described by the parameter which represents the mean dwell time in agreement with prior work [23].

Many binding states may have the same transcription activity. Therefore, the mean dwell time for a transcriptional state is not equivalent to the mean dwell time for an individual binding state.

In order to justify the use of the mean dwell time we recognized that dwell time and burst amplitude, the difference between the maximum and minimum protein levels during the transcriptional state change, were linearly related when the starting protein level was not near full saturation or depletion.

To compare different circuits we choose to use the proportion of the mean burst amplitude in the ON state versus the sum of the means in the ON and OFF state. This allowed us to develop a scale where a value near 0% indicates greater deviations in protein levels in the OFF state and a value near 100% indicates greater deviations from bursts in the ON state. To illustrate this scale we mapped this proportion to a Red-Yellow-Green color gradient as seen in the manuscript.

An intermediate visualization of this proportion can be seen in the following figure. The portion of the bar extending to the left from zero in the negative direction indicates the mean burst amplitude in the OFF state over the sum of the means in the ON and OFF state. The portion of the bar extending to the right right from zero in the positive direction indicates the analogous proportion for the ON state. The absolute width of the bar should equal unity. Bars extending further to the negative direction (left) are colored red whereas bars extending to the positive direction (right) are colored green.

2.10.3 Uncoupled Transcription-Translation Simulations

In the simulation time traces presented in the article, transcription and translation is coupled as would be expected in prokaryotic systems where transcription and translation are occurring simultaneously. While there is some evidence that this may be true in some cases for eukaryotic systems, eukaryotic systems present situations where coupling of transcription and translation may not be true. To explore this, we can explicitly simulate the transcription of mRNA and its translation as distinct processes. We find that the applicability of stochastic bursts due to transcription factor binding and unbinding is dependent on the relative speed of transcription versus translation.

2.10.4 Post Translational Regulation

While post translation regulation is out of the scope of this work, we can hypothesize on ways similar effects could occur. The key aspect of stochastic transcriptional regulation is the single discrete nature and subsequent amplification of the signal. Similar effects could be done by activation of other enzymatic cascades which only require one or few molecules to be active, but would have large effects.

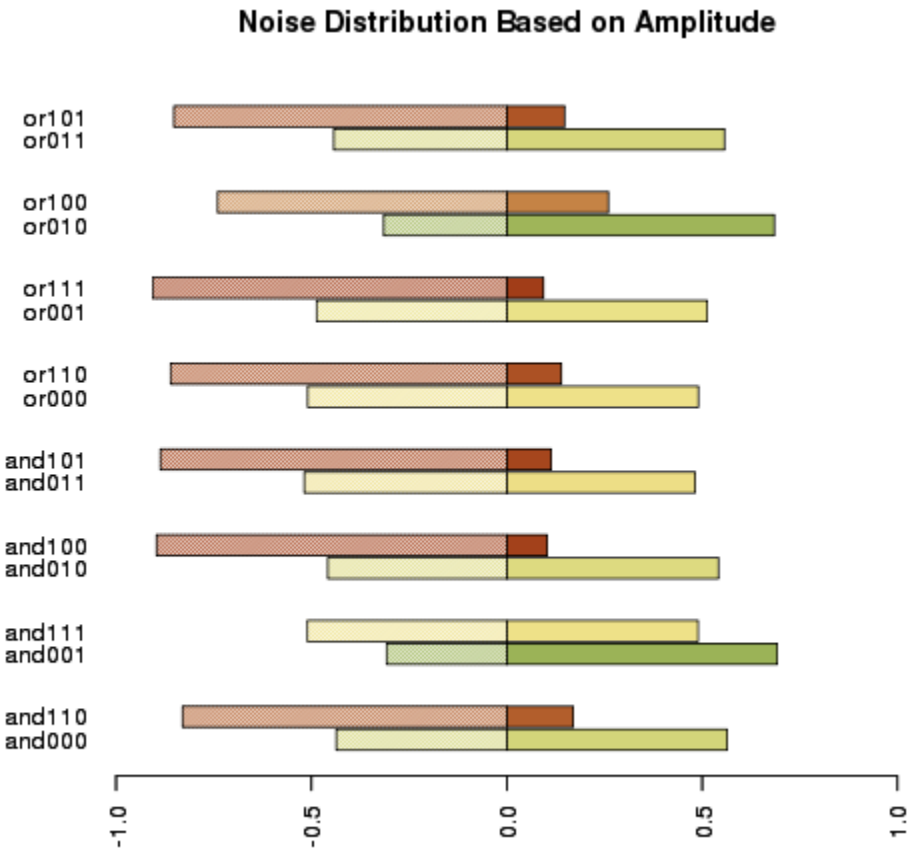


Figure 2.16: Noise is measured here by mean burst amplitude. The proportion of noise in the OFF state is indicated by negative deviations of the bar, whereas the proportion of noise in the ON state is indicate by positive deviations of the bar. The total width of the bar is equal to unity. The bars are colored according to a red-yellow-green color gradient as used in the article.

2.10.5 Parameter Choice and Variation

Parameters were set to arbitrary time units with the first order degradation rate constant being 0.8 protein / time unit. The time units are approximately on the scale of minutes. The equilibrium constants, ratios of kinetics parameters, were chosen such that the range of transcription factor concentration between ON and OFF states would result in changes of transcriptional activity while the kinetics of binding and unbinding were selected to be within an order of magnitude of transcription as described. The first order degradation and dilution rate constant of a protein was set to 0.8 per minute corresponding to a protein half-life of 0.87 minutes. Protein production (transcription and translation) was set such that an active promoter would produce approximately 200 to 300 proteins per minute and that an inactive promoter would produce 10 proteins per minute. Independent binding kinetics were simulated with a Hill coefficient of 2 to account for dimerization or non-specific binding. Kinetics were such that at high concentration of transcription factors (250), binding would occur about 10 times per minute, whereas at low concentrations binding would occur less than one time per minute (0.1) on average. Binding is psuedo first order due to the single copy nature of the promoter. Thus bursts due to binding of transcription factor are rare if the concentration is much less than the dissocciation constant. Hill binding kinetics were implemented using the following propensity rate law: $P(binding) = k_{bind}([A]/n)^n$. That is to say that a molecule A binding functions as n molecules binding cooperatively. The concentration of the DNA promoter is left out since this is modeled as unity.

The overall binding, unbinding, degradation, and transcription-translation constants were varied individually and in combination by increasing and decreasing parameters by 1.2-fold and 2-fold.

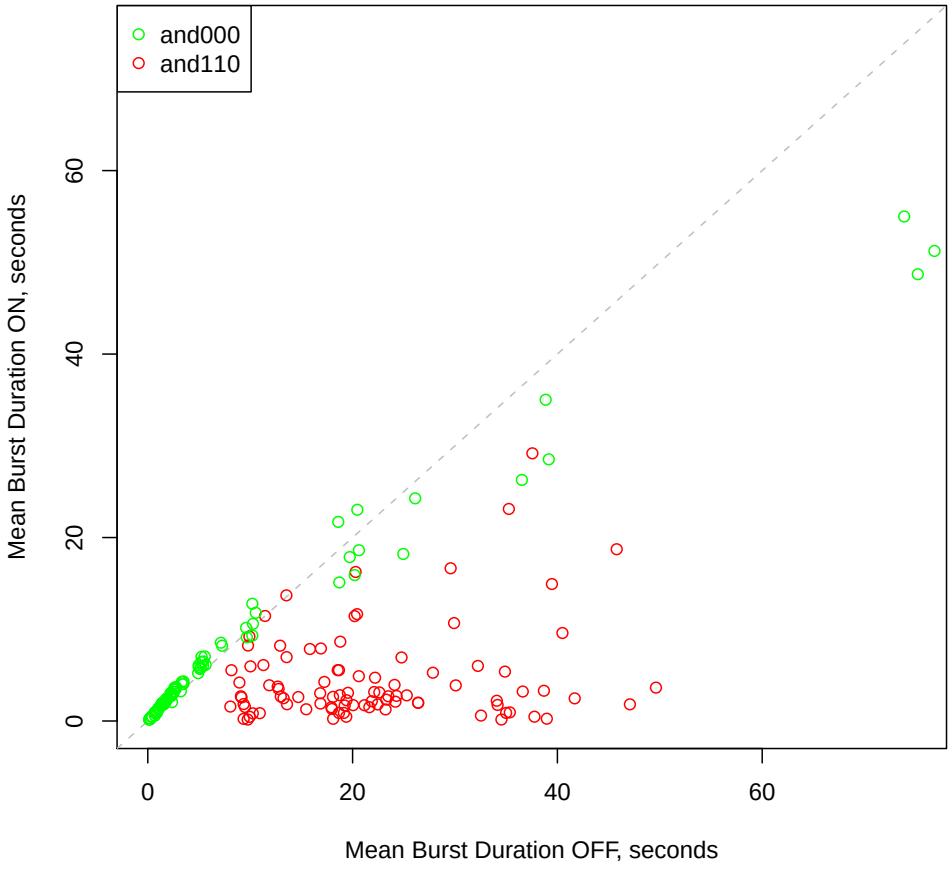


Figure 2.17: Parameter Variation of the and000 and and110 isocircuits

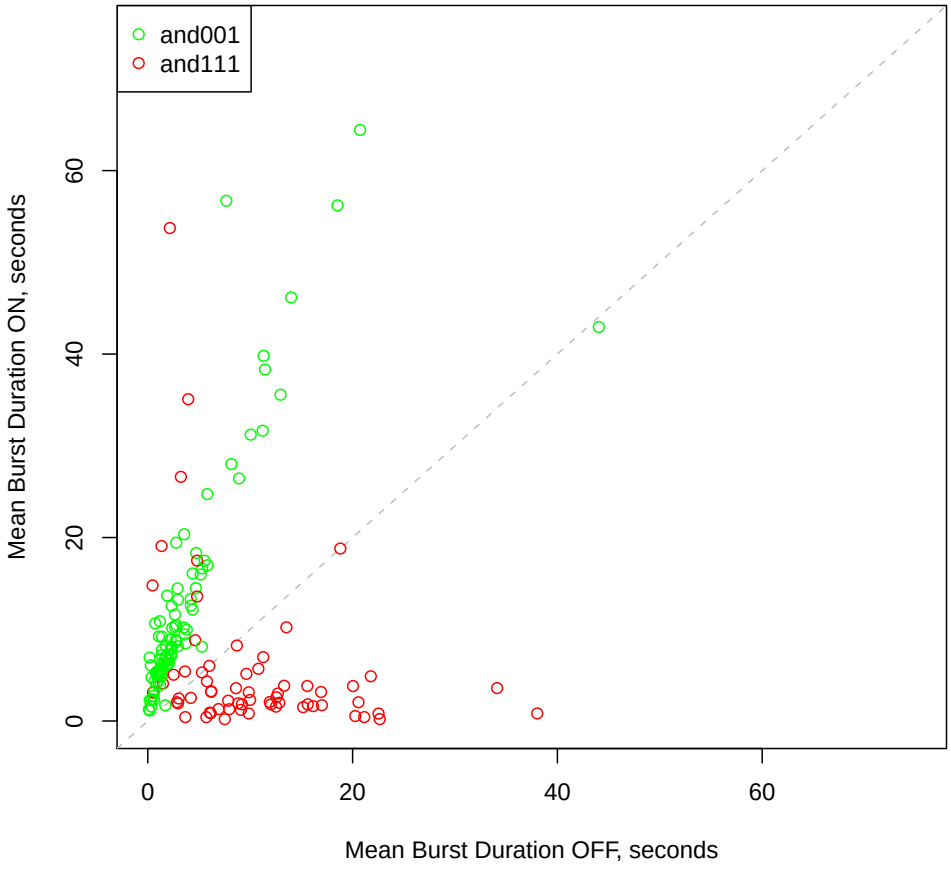


Figure 2.18: Parameter Variation of the and001 and and111 isocircuits

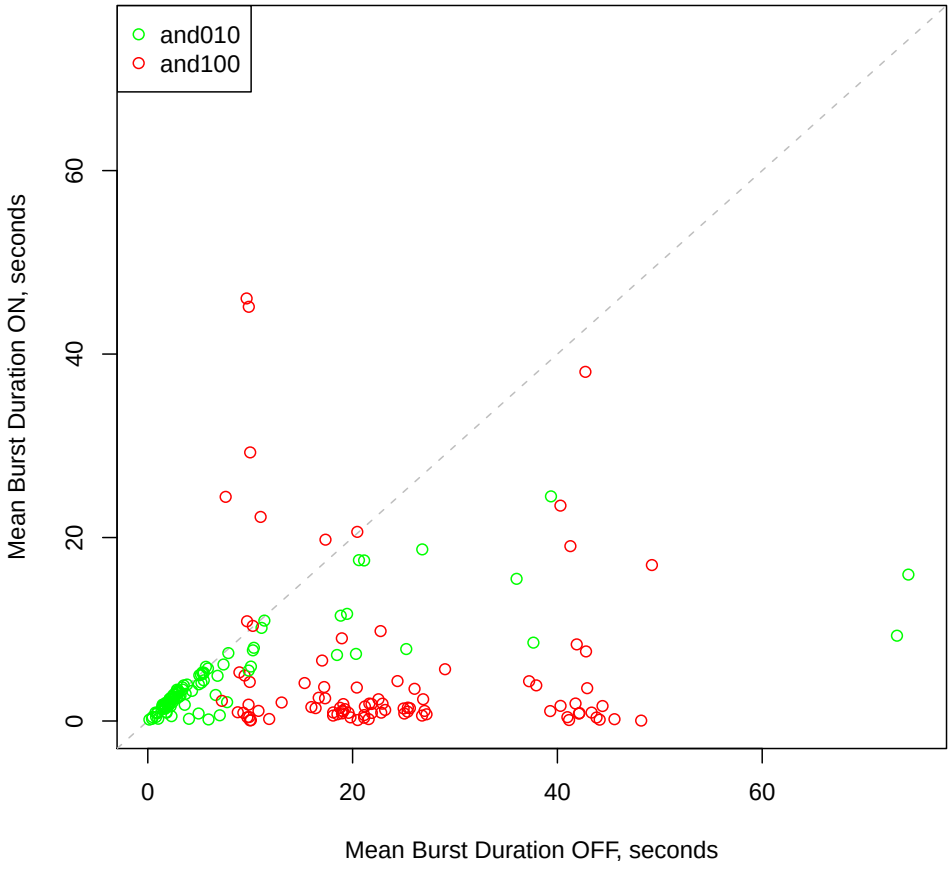


Figure 2.19: Parameter Variation of the and010 and and100 isocircuits

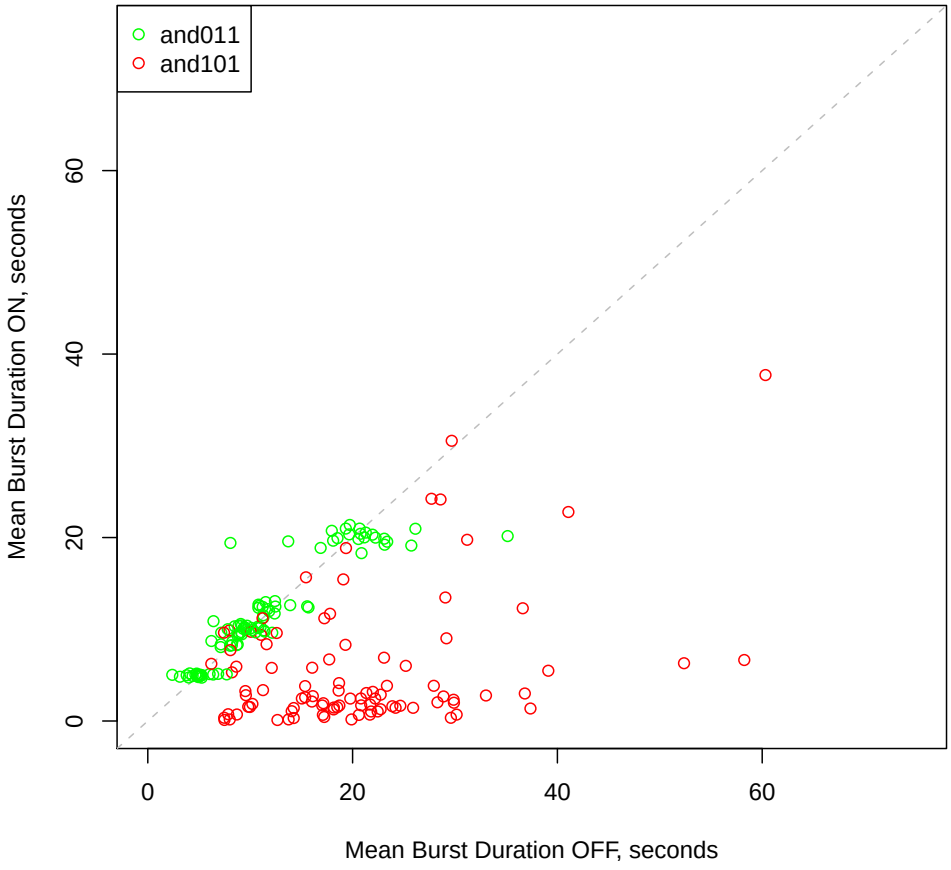


Figure 2.20: Parameter Variation of the and011 and and101 isocircuits

Name	Value	Name	Value	Comment
$Trans_A$	240	$Degrade_A$	0.8	Transcription & Translation, Degradation of A
$Trans_B$	284	$Degrade_B$	0.8	Transcription & Translation, Degradation of B
$Trans_C$	336	$Degrade_C$	0.8	Transcription & Translation, Degradation of C
$Trans_{Basal}$	10			Basal Transcription & Translation
$Bind$	0.0006	$Unbind$	2.4	Protein-Promoter Binding / Unbinding
$Hill$	2			Hill Coefficient for Binding

Table 2.2: Rate constants used in the simulation: Unit time is based upon the average time for one protein to degrade when only one is present. Transcription rate constants are used for unimolecular rate equations with the units as the average number of proteins produced per unit time per active promoter. Degradation rate constants are used with unimolecular rates with the units of average number of proteins degraded per unit time per total protein concentration. Basal transcription rate constants are in units of average number of proteins basally produced per unit time. Binding rate constants are bimolecular with units average number of complexes formed per unit time per promoter regions present per binding protein present. This is effectively unimolecular since the number of promoter regions is unity and constant. Unbinding rate constants are unimolecular with units average number of complexes dissociated per unit time per complexes present. This is effectively constant when a complex is present since the number of a DNA-protein complex is either unity of when present or zero when absent.

2.10.6 Priming of Transitions Between ON and OFF states

A significant implication of large stochastic bursts during steady state is that this may accelerate the transition of output between ON and OFF states. Dynamic behaviors when switching between ON and OFF states include pulse generation and accelerated response curves. Bursts of gene expression can stochastically accelerate the response by placing the output of the network closer to the new canonical output after the signaling change before a signal change has even occurred. For pulse generators this would create an extended plateau before the pulse dissipates.

2.11 Bioinformatics

2.11.1 Network Information

Network information was downloaded from RegulonDB. Awk and sed were used to modify the initial table for a format appropriate for FANMOD. This involved numbering the species and

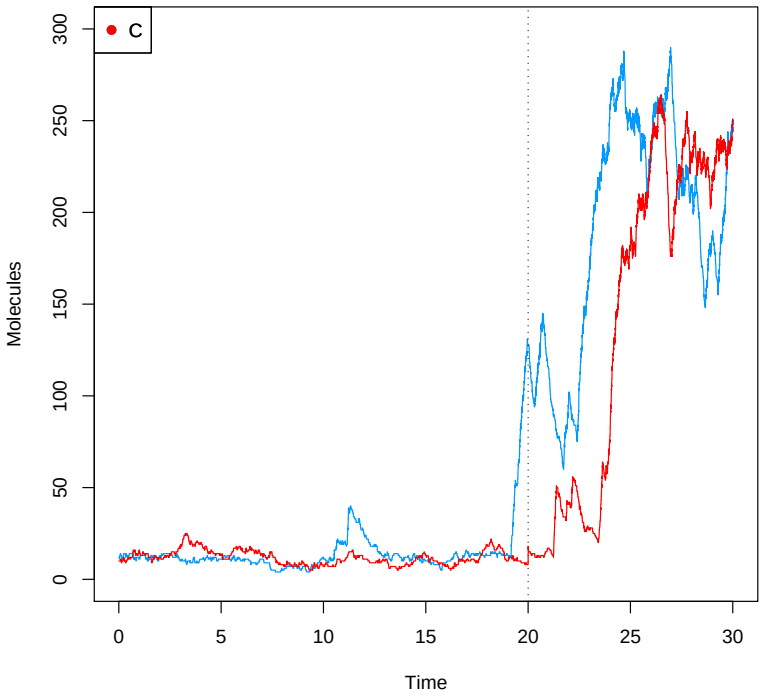


Figure 2.21: A stochastic burst of gene expression can accelerate the transition between OFF and ON states. In blue, the And001 FFL has a stochastic burst of gene expression that primes and accelerates the transition shortly before an OFF to ON signaling transition at 20 minutes. In red, the And001 FFL does not have a burst of gene expression and transitions to the ON signaling state normally without priming.

removing autoregulation. FANMOD was used to discover and type the FFLs by extracting information from the dump file. The assignment of the FFL nodes A, B, and C was not done automatically by FANMOD. Assignment was done by counting the number of receiving interactions since the number of interactions for A, B, and C would be 0, 1, and 2 respectively. Unlike prior analyzes of FFLs filtering for microarray or electronically inferred interactions was done conducted.

Categories were assigned based on a hierarchial analysis[49] of the *E. Coli* transcriptional regulatory network and summaries from EcoCyc.

2.11.2 Chi-Square Testing

Chi-square testing can be done by analyzing the the FFL loop classifications and functional categorizations as a contingency table. Expected values for a random distribution can be calculated by product of the two marginals for a loop-function cell of interest divided by the total observations.

2.11.3 G-Tests

For sparse contingency table groupings as is manifested without grouping loop classes or functions the chi square test are not valid. The G-test can be applied as a maximum likelihood statistical significance test in place of the chi-squared test.

2.11.4 Bootstrap Resampling Analysis

Bootstrap resampling was applied to determine the robustness of clustering by resampling the functional categories. This involved selecting a sample of 39 categories from the original 39 categories with replacement. A ratio of the average of the distances within a FFL cluster to the average distances between the two FFL clusters was determined the quality of clustering. A ratio of one would indicate that the distances between FFLs within a cluster and between clusters were

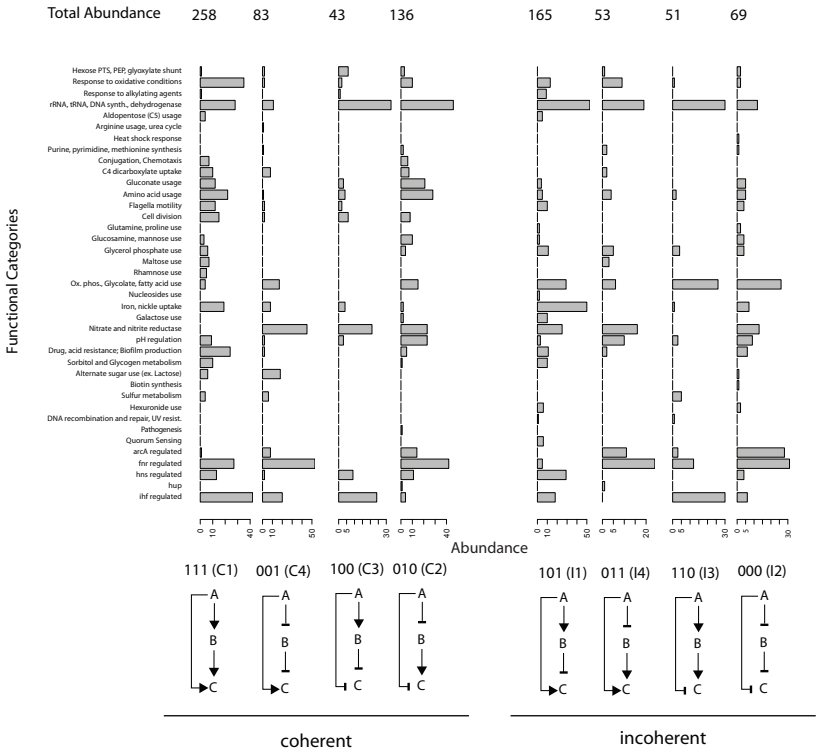


Figure 2.22: Bargraph representation of FFL abundances categorized by type and by functional category.

equal. Higher ratios greater than one would indicate that the FFLs were more closely related in terms of functional profile distribution, whereas ratios less than one would indicate the lack of clustering. The percent of samples that clustered were those with ratios greater than one.

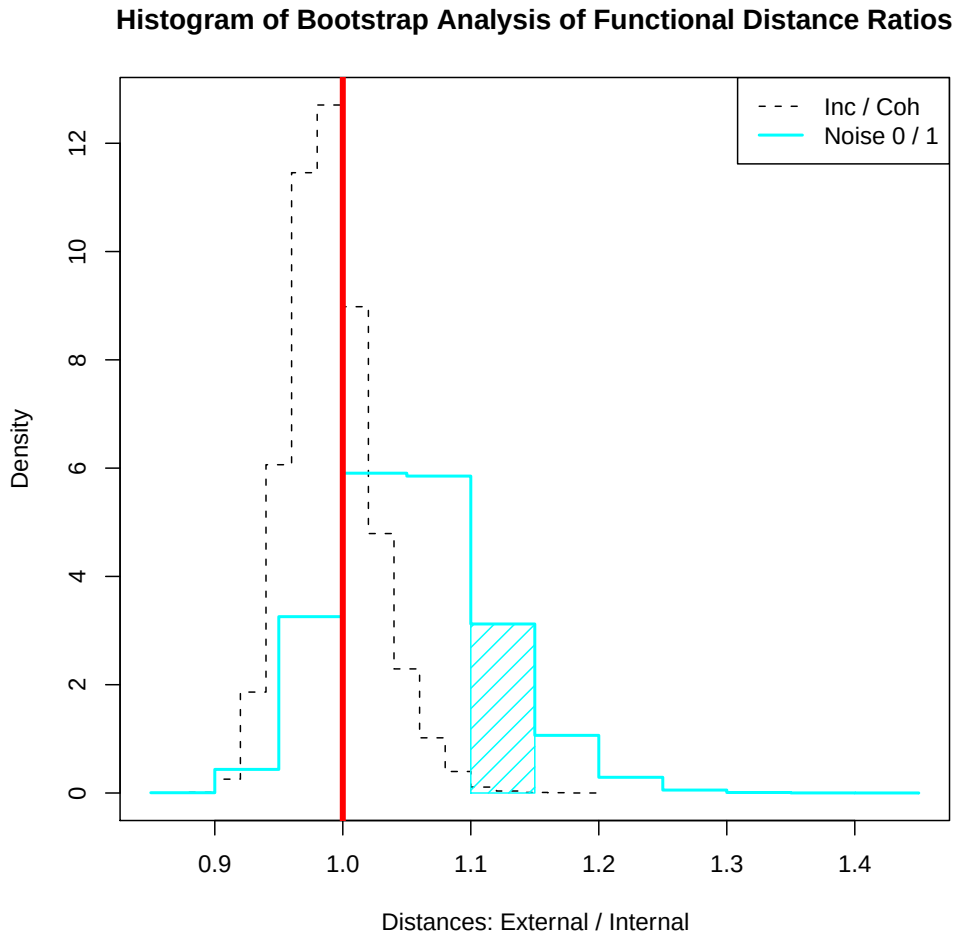


Figure 2.23: Bootstrap Analysis

Chapter 3

Stochasticity as an Arbiter of Robustness for Genetic Competence in Bacillus Subtilis

3.1 Preamble

Does noise have a physiological purpose? Previously, we saw that feed-forward loops with different stochastic properties segregated into different functions possibly based upon the demands put upon those functions by the environment. Subsequently, though it has been difficult to find feed-forward loops that fit into the needed kinetic requirements for substantial differences to occur and that can be isolated for study. The competence circuit in *Bacillus subtilis*, however, has known stochastic properties and can actually be seen as a feed-forward loop that controls itself. The input and the output node are the same gene. The direct regulation is an positive autoregulatory feedback loop, while the indirect relationship is a negative feedback loop. I will thus use this system to examine some of the concepts proposed in the previous chapter.

3.2 Introduction

In order to examine the effect of molecular copy number, I will examine a known stochastic system with a more deterministic analog. Competence in *Bacillus subtilis* allows the bacterium to integrate exogenous genes or alleles into it's genome through homologous recombination. This mechanism permits horizontal gene transfer to occur. A synthetic analog of the native competence machinery was also recently created by Çağatay et al [7]. Working along the same principles as

the feed-forward loops, the two functional molecules of the competence control circuit can either be correlated or anti-correlated. The native circuit is anti-correlated such that noise is more evenly spread between the vegetative and competence states. In the synthetic circuit, the two components are correlated making the active competence state much more deterministic.

This is most clearly seen in the duration of competence. The native circuit has a highly variable duration based on the rate of competence initiation. The synthetic circuit however has a very well defined duration due to a stable limit cycle. In a previous experiment, it was found that the native competence circuit was more successful at integrating exogenous DNA at different DNA concentrations than the synthetic circuit. In this work, I modify the native and synthetic circuit such that we could control induction of ComK directly. In particular, we used an IPTG inducible promoter derived from the Lac operon in E. coli. When IPTG is added, LacI unbinds from the promoter region and permits expression from the operon.

3.3 Results

3.3.1 Creation of a Tunable Synthetic Competence Circuit

In order to create a synthetic competence circuit that is tunable in a stress-like manner, we first need to reconfigure the original synthetic organism. The original SynEx created by Çağatay modified the PY79 strain of *Bacillus subtilis* with the following chromosomal alterations:

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}comS$	Spectinomycin
SacA	$P_{comGKbox1}mecA^{xp}$	Chloramphenicol
$\Delta srfA, comS$	$P_{comG}cfp$	Neomycin / Kanamycin

A SynExSlow variant was also created:

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}comS$	Spectinomycin
SacA	$P_{comGKbox1}mecA^{xp}, P_{comG}comS$	Chloramphenicol
$\Delta srfA, comS$	$P_{comG}cfp$	Neomycin / Kanamycin

In order to create *ComK* tunable versions of these strains we needed to replace the tunable $P_{hyperspank}comS$ with an internally controlled promoter for the ribosomal gene *rpsD*. Then we also needed to add in $P_{hyperspank}comK$. This created the SynExK variant:

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}comK$	Spectinomycin
SacA	$P_{comGKbox1}mecA^{xp}$	Chloramphenicol
$\Delta srfA, comS$	$P_{comG}cfp$	Neomycin / Kanamycin
GltA	$P_{rpsD}comS$	Phleomycin

and the SynExKSlow variant:

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}comK$	Spectinomycin
SacA	$P_{comGKbox1}mecA^{xp}$	Chloramphenicol
$\Delta srfA, comS$	$P_{comG}cfp$	Neomycin / Kanamycin
GltA	$P_{rpsD}comS$, $P_{comG}comS$	Phleomycin

These two new strains now are resistant to four antibiotics and have *comS* expressed from a ribosomal promoter providing for a basal level of expression. ComK in these strains can also be induced by IPTG in order to trigger competence. Since stress signals are usually integrated at the comK promoter, IPTG is in effect acting as a proxy for stress.

To study the native competence circuit, two strains were used. The first strain is a variant from a previous study [13].

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}$ comK	Spectinomycin
SacA	P_{comG} cfp, P_{comS} yfp	Chloramphenicol

The second strain was constructed so that it could be differentiated from the SynEx strains in a co-culture:

Locus	Construct	Antibiotic Resistance
AmyE	$P_{hyperspank}$ comK	Spectinomycin
SacA	P_{comG} yfp, P_{rpsD} mCherry	Chloramphenicol

3.3.2 Synthetic Competence has a Smaller Dynamic Response Range

In order to examine how the two strains responded to increasing amounts of artificial stress, I captured time-lapse micrographs of populations of both strains at single cell resolution. The native competence circuit had previously been characterized [13]. In that prior study, the native competence circuit was observed to transition from an excitable state, then an oscillatory state, and finally a monostable state stuck with high ComK expression when the IPTG level was increased. The transition from the excitable to the oscillatory state occurred around 3 uM IPTG while the high comK monostable state was achieved at 100 μM IPTG.

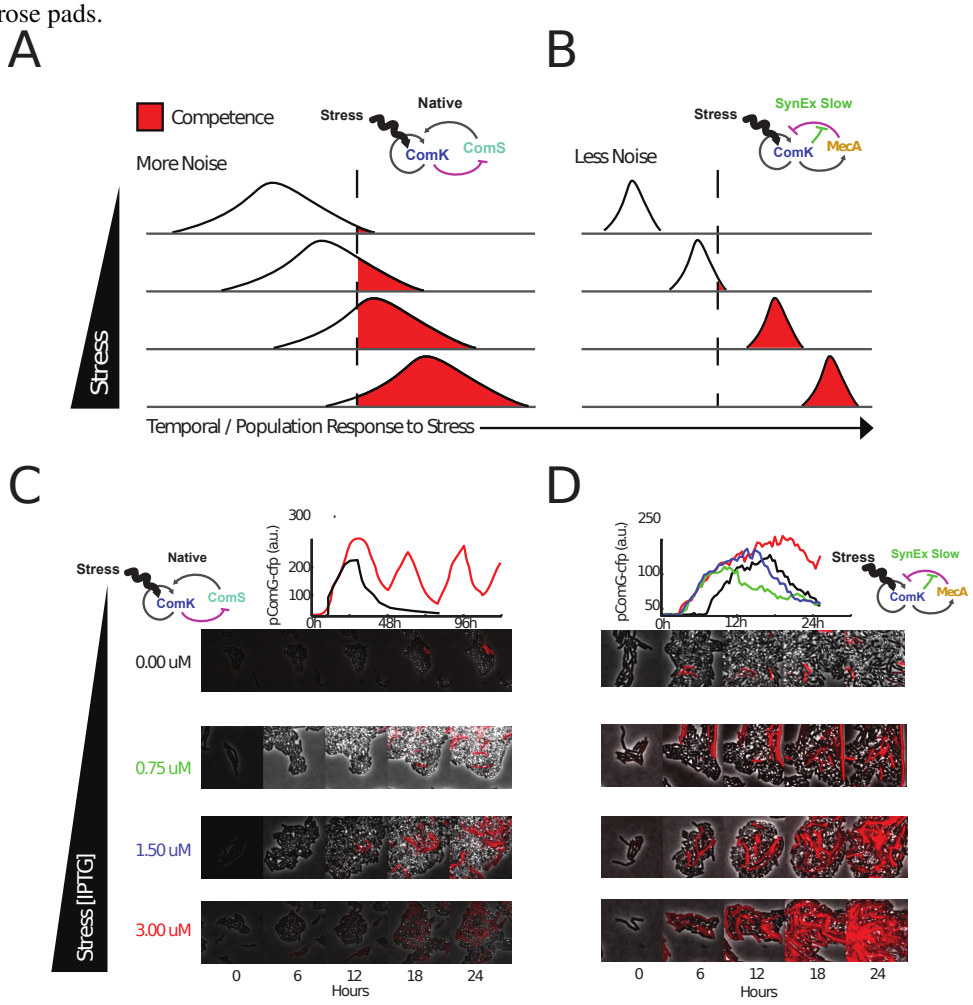
In comparing the two competence circuits at different levels of artificial induction we found the ability of the synthetic circuit to exit from competence was compromised at moderate IPTG concentrations of 3 μM IPTG. Rather than entering an oscillatory state as in the native circuit, it was found that the SynExKSlow circuit was already stuck in the high ComK competence state with induction by 3 μM IPTG (Fig 1). Furthermore, cells of the SynExKSlow strain were elongated due to inhibition of ftsZ ring formation necessary for division.

3.4 Materials and Methods

3.4.1 Sterlini-Mandelstrom Resuspension Media

Sterlini-Mandelstram Resuspension Medium was used during time-lapse microscopy and followed the protocol as in references [60, 61]. The actual protocol used consists of making two salt solutions: A and B. Solution A consists of 0.089 g of $FeCl_3 6H_2O$, 0.830 g of $MgCl_2 6H_2O$ and 1.979 g $MnCl_2 4H_2O$ in 100 mL of filtered water. Solution A is filter sterilized (not autoclaved) and stored at 4 °C. Solution B consists of 53.5 g NH_4Cl, 10.6 g Na_2SO_4, 6.8 g KH_2PO4, and 9.7 g NH_4NO_3. Solution B is then also filter sterilized and stored at 4 °C. Sporulation salts are made by combining

Figure 3.1: A and B) Illustration of the population based competence response to increasing concentrations of IPTG for the Native (A) and SynExKSlow (B) strains. Time-lapse micrographs of the Native (C) and SynExKSlow (D) strains with 0, 0.75, 1.5, and 3 μM IPTG in low melting point agarose pads.



adding 1 mL of Solution A and 10 mL of Solution B to filtered water for a total 1 L. This solution is then autoclaved. The final Resuspension media is created by combining 93 mL of sporulation salts, 2 mL of 10% v/v L-glutamate, 1 mL of 0.1M $CaCl_2$, and 4 mL of 1M $MgSO_4$ on the day of the experiment.

3.4.2 One-Step Transformation Media

One-step transformation media consists of 6.25 g of $K_2HPO_43H_2O$, 1.5 g of KH_2PO_4, 0.25 g of trisodium citrate, 50 mg of $MgSO_47H_2O$, 0.5 g of Na_2SO_4 at pH 7.0, 125 μL of 100 mM $FeCl_3$, 5 μL of 100 mM $MnSO_4$, 1 g of glucose, and 0.5 g of glutamate added into filtered water for a total of 250 mL. The media is filter sterilized using 0.2 micron Millipore filters.

3.4.3 2xYT Medium

2xYT recovery medium consists of 16.0 g of Tryptone, 10.0 g of Yeast Extract, 5.0 g of NaCl added to filtered water to a total volume of 1L. The media is then filter sterilized using 0.2 micron Millipore filters.

3.4.4 Time-Lapse Microscopy

Cells of *Bacillus subtilis* were prepared by streaking the cells from glycerol stocks onto LB agar plates containing the appropriate antibiotic for maintenance. Single colonies were then selected from the plates and grown in LB broth for three to four hours at 37 °C until an OD of 1.6 to 1.8 is reached.Agarose pads are made by pouring 6 mL of 0.8% w/v Low-Melting Point Agarose in Resuspension medium onto a glass coverslip. Another glass coverslip is placed on top of the agarose pad, and it is left to congeal while the culture is grown. Before the deposition of cells, the glass coverslip is removed. Cells are imaged by dropping 2 μL drops of culture in resuspension media on

pre-heated low melting point agarose pads. The pads were cut into squares with a 5mm edge. After the pads were dried for one additional hour, the pads flipped over and placed on a glass-bottom dish. The dish was then sealed with parafilm. Images of the culture were then obtained at 100X magnification on an Olympus IX-82 system ImagePro software from MediaCybernetics along with customized macros.

3.4.5 Plasmid construction

Template plasmids with homologous recombination arms for the *Bacillus subtilis* chromosomal loci were modified through restriction enzyme digest and ligation of DNA inserts. The inserts were created by polymerase chain reaction using primers from Integrated DNA Technologies while using genomic DNA or other plasmids as templates.

3.4.6 Strain construction

The PY79 strain of *Bacillus subtilis* was modified through homologous recombination through use a One-Step Transformation protocol by inducing competence. 50 ng of plasmid DNA was replicated in TOP10 *E. coli* cells (Invitrogen, Life Sciences, Inc) and purified using a MiniPrep spin column (Sigma-Aldrich). The DNA was then mixed with culture growing in minimal salts for thirty minutes and then subsequently were rescued using 2xYT rich medium. Positive colonies were then selected on LB agar plates containing selective concentrations of antibiotics.

3.5 Discussion

In this study, I constructed a strain of *Bacillus subtilis* that was tunable using an inducible promoter and contained a synthetic competence circuit that had exhibited competence durations with less variation. This more deterministic strain, SynExKSlow, was compared with a tunable

version of the Native circuit. While strains containing both circuits eventually reached a monostable high ComK state, the SynExKSlow strain appeared to reach this state at a lower concentration of IPTG than the strain containing the Native competence circuit. At the same concentration of IPTG, the Native circuit was known previously to enter and exit competence consecutively in a manner resembling oscillations.

One possibility is that the stochastic properties of the Native competence circuit allowed for this strain to exit competence despite sufficient artificial induction to re-enter the competence state. By making the competence circuit less stochastic in the SynExKSlow circuit due high copy numbers of both ComK and MecA, perhaps the ability to exit competence was no longer present. Perhaps then by making the competence circuit more deterministic, I also removed the ability for the circuit to respond gradually to increasing stimulus.

Another possibility is that the ability of MecA to assist in the degradation of ComK is limited at high concentrations. However, the ClpXP protease is known to be highly robust. Additionally, a concentration of 3 μM IPTG is not a very high induction level.

The differences in response to IPTG concentration between tunable strains containing either the Native competence circuit or the SynExKSlow competence circuit may be related to differences in their susceptibility to stochastic effects based upon copy number. These differences, however, may also have consequences for the ecological fitness of the two strains. In Chapter 6, I propose a set of experiments to examine the relative fitness of these two strains.

Chapter 4

Localized Cell Death Focuses Mechanical Forces During 3D patterning in a Biofilm

4.1 Preamble

The *Bacillus subtilis* strain NCBI 3610 forms complex structured populations of bacteria called biofilms. These communities of bacteria appear to be subject to stochastic effects manifested in a dynamic and spatially heterogeneous cell death pattern (CDP). While the initial phase of cell death appears to be mechanistically related to wrinkles by creating mechanical instability, the later stages of cell death also appear correlated with the rugose quality of these biofilms. This pattern as well as the localization phenomena in the next chapter show how stochastic systems can retain their spontaneity but also be cued by initial or external factors. In this work, I collaborated with a team of researchers to explore this complex system to draw biological and physical insights into how spatial heterogeneity can lead to the development of these macroscopic structures.

4.2 Abstract

From microbial biofilm communities to multicellular organisms, 3D macroscopic structures develop through poorly understood interplay between cellular processes and mechanical forces. Investigating wrinkled biofilms of Bacillus subtilis, we discovered a pattern of localized cell death that spatially focuses mechanical forces, and thereby initiates wrinkle formation. Deletion of genes implicated in biofilm development, together with mathematical modeling, revealed that ECM pro-

duction underlies the localization of cell death. Simultaneously with cell death, we quantitatively measured mechanical stiffness and movement in WT and mutant biofilms. Results suggest that localized cell death provides an outlet for lateral compressive forces, thereby promoting vertical mechanical buckling, which subsequently leads to wrinkle formation. Guided by these findings, we were able to generate artificial wrinkle patterns within biofilms. Formation of 3D structures facilitated by cell death may underlie self-organization in other developmental systems, and could enable engineering of macroscopic structures from cell populations.

4.3 Introduction

Populations of Bacillus subtilis can form communities called biofilms. These biofilms consists of the bacteria themselves in addition to extracellular material that they secrete including polysaccharides and amyloid.

Self-organization in space and time is a fundamental developmental process, defined by the autonomous formation of 3D macroscopic structures by replicating cell populations [62, 63, 64]. Such 3D pattern formation underlies the development of all multicellular organisms and cellular communities, and appears to be governed by two principal processes. First, genetic programs control cellular processes, such as growth, death, and differentiation. Second, 3D structure formation involves macroscopic movement of cell populations that are determined by mechanical properties and physical forces [65]. Recent studies have investigated each of these processes separately in different biological systems [66, 67, 68, 69]. However, insight into the direct interplay between cellular and mechanical processes that drives development requires simultaneous measurement of both processes, and thus constitutes a major challenge.

Compared with multicellular organisms, microbial biofilms are simpler systems for in-

vestigating the interaction between cellular and mechanical aspects of 3D self-organization during development. Interestingly, these microbial communities still exhibit diverse cellular behaviors and complex spatial organization [70, 71, 72, 73, 74]. For example, biofilms can develop from a single cell and give rise to complex 3D wrinkle structures that are visible to the naked eye, comprising billions of cells [70, 71, 75] (Fig. 4.1A). Aside from replication, bacterial cells can also exhibit other behaviors, such as genetically controlled cell death [70, 76, 77] and excretion of ECM components [70, 74, 78, 79, 80, 81]. In fact, one of the defining features of any biofilm is that cells are embedded within an ECM composed of diverse molecules, such as polysaccharides and amyloid fibers [79, 80, 81]. The ECM is required for wrinkle formation and appears to provide the biofilm with resilience against environmental extremes as well as mechanical support [70, 80, 81, 82]. It is also conceivable that replicating cells within the ECM can generate forces by pushing against each other, whereas cell death could provide an outlet for such forces. However, the interplay between mechanical forces and cellular processes, such as cell growth and death, and their role during biofilm development remains unclear.

Here, we investigated the spatiotemporal dynamics of cell death and mechanical processes during self-organization of Bacillus subtilis cells into wrinkled biofilms. Specifically, we combined quantitative measurements of cell death, movement, and mechanical properties to analyze development of WT and mutant biofilms. These measurements were obtained with fluorescence time-lapse microscopy, tracking of fluorescent beads embedded within biofilms, and atomic force microscopy (AFM) nanoindentation, respectively. Our findings suggest that lateral mechanical forces build up during biofilm growth and expansion, and are then focused in space by localized cell death, thus triggering vertical buckling and subsequent wrinkle formation. The formation of wrinkle structures facilitated by cell death may constitute a population-level stress response of biofilms. These insights

could allow the engineering of desired macroscopic structures from cellular populations in the future and provide exciting opportunities at the interface between synthetic biology and material sciences.

4.3.1 Contributions

Experimental analysis of biofilms was initiated by Munehiro Asally, PhD, in the supervision of Professor Gürol Süel. I contributed to the design, analysis, and writing of the manuscript including conception of the mechanical model, quantification of results, and interpretation of the data. While my initial interest was in the apparent stochasticity of the cell death pattern, my quantitative and physical expertise contributed to extracting information from the microscopy images and connecting those observations to a physical mechanism of wrinkle formation. Electron microscopy and other supporting research was performed by Andra Robinson which helped confirm the actual death of cells. Pau Rue and Professor Jordi Garcia-Ojalvo contributed and designed the mathematical model. Yingjie Du, Zhenxing Hu, and Professor Hongbing Lu measured load-displacement curves and analyzed stiffness measurements.

Particle tracking using the ImageJ Mosaic suite of plugins also required modification in order to scale with the large number of fluorescent beads tracked. An earlier prototype of deformation measurements was developed by a Summer Undergraduate Research Fellowship student, Helen Wu, who studied at Harvey Mudd College. Her worked laid the foundation for the convergence maps that colocated with the cell death pattern.

This work was originally published as "Localized cell focuses mechanical forces during 3D patterning in a biofilm" in the Proceedings of the Natural Academcy of Sciences in 2012. Please see Prior Publications for the full citation.

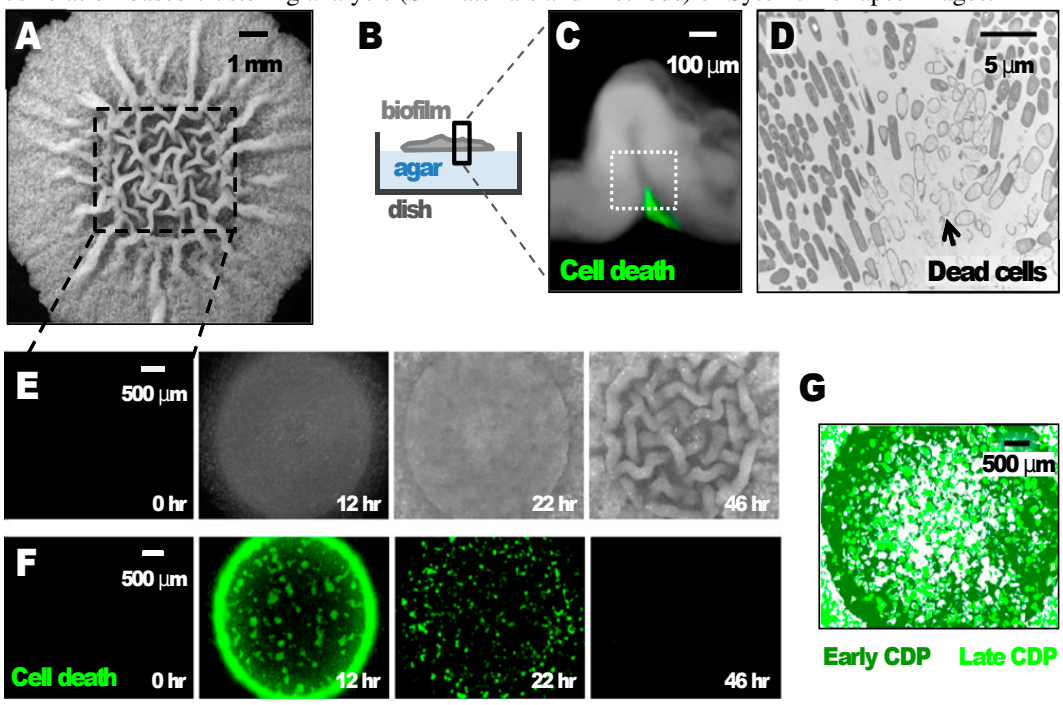
4.4 Results

4.4.1 Early Cell Death and Wrinkles Colocalize

Multicolor fluorescence and transmission EM imaging of biofilms indicated that cell death and wrinkles are spatially correlated. We imaged cross-sections of B. subtilis biofilms that contained fluorescent reporters for dividing cells and death (Fig. 4.1 B and C). In particular, we used expression of CFP from the promoter of the cell division operon ftsAZ as a reporter for cell density [83] (Fig. 4.1C, gray). At the same time, we measured cell death using Sytox Green, a commercially available fluorescent marker of cell death (Fig. 4.1C, green). This high-affinity nucleic acid stain only penetrates permeabilized cells and has been established as a specific reporter for cell death in various systems, including B. subtilis [84, 85] (Fig. S1 A and B). Importantly, this reporter does not affect biofilm formation (Fig. S1C). Cross-sections of biofilms revealed that cell death is localized at the bottom of biofilms, and more specifically at the center of folded wrinkle structures (Fig. 4.1C). Transmission EM confirmed the presence of dead cells at the bottom of wrinkle interfaces (Fig. 1D). These findings are consistent with at least two scenarios, namely, that wrinkle formation results in the death of cells or that cell death occurs first and perhaps facilitates wrinkle formation.

To discriminate between these possible scenarios, we began by measuring the spatiotemporal dynamics of cell death during B. subtilis biofilm development using fluorescence time-lapse microscopy. Unexpectedly, measurements uncovered a striking heterogeneous distribution of cell death that occurs before wrinkle formation (Fig. 4.1 EG). A similarly heterogeneous cell death pattern (CDP) is generated even when the biofilm is started from a single cell (Fig. S1E). Together, these data suggest that the replicating cell population during biofilm development gives rise to a heterogeneous CDP that may be genetically controlled.

Figure 4.1: Localized pattern of cell death correlates with the site of wrinkles in biofilms. (A) Three-day-old B. subtilis biofilm structure. The dashed square indicates the region of interest in this study. (B) Schematic of the microscope setting for fluorescent and bright-field imaging of a biofilm (depicted in gray). Dashed lines indicate the direction of observation. (C) Cross-section fluorescence image of a 30-h-old biofilm wrinkle, pseudocolored (PftsAZ-CFP in gray and a cell death reporter Sytox in green). (D) Transmission electron (TEM) micrograph of a sectioned biofilm wrinkle shows dead cells (black arrow) in the wrinkle interior (white box in C depicts observed location). (E) Film strip shows biofilm morphology observed from above the colony. (F) Film strip shows CDP during early (22 h) development of biofilm, imaged from below. Brightness and contrast are individually adjusted for each time point. (G) Early (dark green) and late (light green) CDPs detected by a correlation-based clustering analysis (SI Materials and Methods) of Sytox time-lapse images.



4.4.2 Deletion of ECM Genes Reduces Spatial Heterogeneity of Cell Death

To identify genetic control of the CDP, we deleted 32 genes involved in eight diverse aspects of biofilm development and analyzed the resulting CDPs quantitatively (Fig. 4.2 AF). Specifically, we analyzed at least three movies for each mutant strain and generated overlaid images representing a simplified history of when and where cell death occurred (Fig. 4.1G, Fig. 4.5, and SI Materials and Methods). We then quantified the spatial complexity (heterogeneity) of this CDP. In particular, we calculated the perimeter-to-area ratio (, where P is the perimeter and A is the area) for all CDPs so as to assign a numerical value to their complexity (Fig. 4.2A and SI Materials and Methods). This ratio is equal to unity for a perfect circular pattern and has a higher value for more complex patterns (WT CDP = 2,337 $\pm$ 623, n = 12). This analysis allowed us to identify the connection between gene activities and the spatiotemporal heterogeneity of the CDP (Fig. 4.2A).

The most pronounced change in the CDP phenotype is generated by deletion of genes implicated in ECM production. These genes appear to be required to generate localized cell death. Specifically, five gene deletion strains generated CDPs with statistically significant (P ¡ 0.01) decreased spatial heterogeneity. Three of those genes, srfA, spo0A, and sinI, are known to regulate, among others, matrix production through cell-to-cell signaling and transcriptional regulation. The other two genes, tasA and epsH, are directly involved in production of the ECM. In particular, the tasA gene expresses an extracellular protein that forms amyloid fibers within the matrix [79], whereas the epsH gene is necessary for synthesis of polysaccharides that comprise the biofilm ECM [71]. Deletion of the epsH gene resulted in a CDP with the lowest spatial heterogeneity, where cell death occurred globally and homogeneously (Fig. 4.2F). Consistent with our findings (Fig. 4.2 AF), colony wrinkle morphology has been shown to diminish with matrix deficiency [71, 74, 86, 79]. Here, we further show that these mutant biofilms also exhibit reduced CDP heterogeneity before

deficiency in wrinkle morphologies becomes evident. These results demonstrate that expression of matrix components contributes to the heterogeneity of the CDP in both space and time.

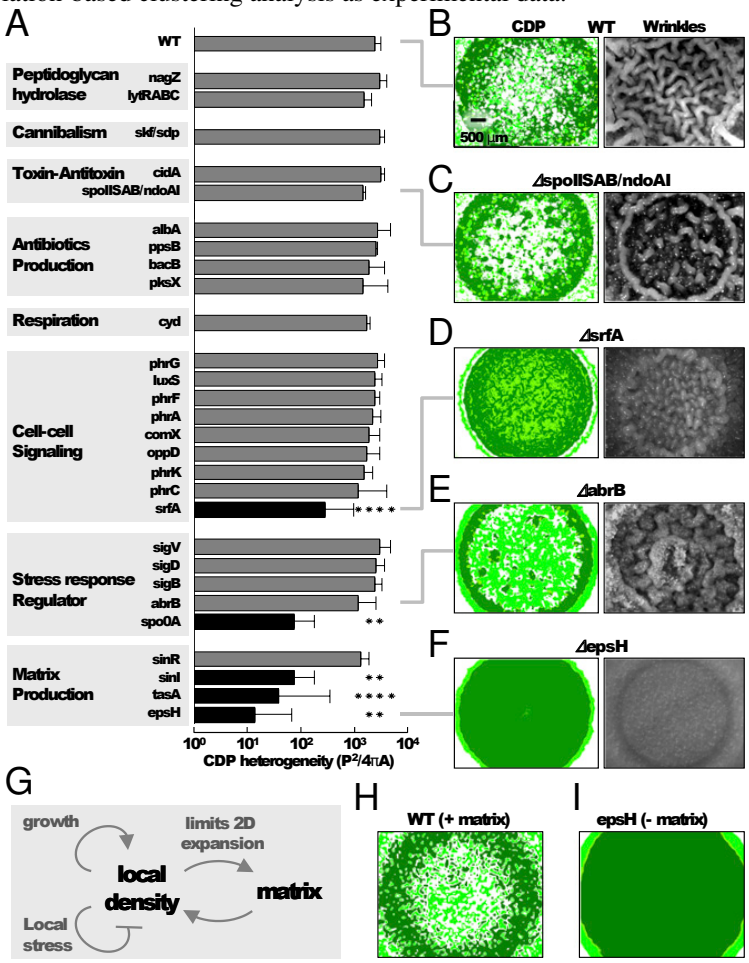
We also identified gene deletions that reduced cell death without changing the spatial pattern. Specifically, deletion of two environmental stress-dependent toxin genes (spoIISAB and ndoAI) reduced cell death by $25 \pm 7\%$ (Fig. 4.2C and Fig. S3), suggesting that death occurs due to build-up of local biochemical stress [87, 88, 89, 90]. The notion of local biochemical stress-induced death is furthermore supported by the following results: (i) Cell death is first observed at regions of initial high cell density (Fig. S1D), and (ii) cell death is heterogeneous and highly localized in space (Fig. 4.1F). Together, these data imply that biochemical stress builds up locally during biofilm development, and thereby gives rise to the observed CDP. Furthermore, this mutant strain with reduced cell death also generated biofilms with visibly diminished wrinkle morphology (Fig. 4.2C). Reduction of cell death thus appeared to have an unfavorable consequence on wrinkle formation, even though the spatial pattern of death was unchanged from WT (Fig. 4.2A). Altogether, the gene deletion results imply that localized cell death may contribute to wrinkle formation during biofilm development.

4.4.3 Population Model Illustrates Effect of ECM on Cell Death Localization

Recently, we have observed a spatially heterogeneous pattern of cell death in Bacillus subtilis biofilms. The heterogeneity manifests itself as localized areas of cell death and I was able to quantify this localization by examining the perimeter to area ratio of the localized regions.

We asked if a mesoscopic mathematical model of population growth, based on our findings described above, could account for the complex CDP. Similar to previous theoretical approaches to microbial pattern formation [91], our model is based on a system of coupled partial differential

Figure 4.2: Gene deletion analysis reveals that ECM production is required for the CDP. (A) KO strains are categorized and listed by their functions. The CDP heterogeneity (perimeter squared over the area) of the late CDP (light green) is graphed for each deletion strain (mean $\pm$ SD). Asterisks and black shading indicate P values (**P $<$ 0.01; ****P $<$ 0.0001) evaluated using Tukeys Honestly Significant Difference test. Representative clustered CDP images as in Fig. 4.1G (Left) and colony morphology (Right) of WT (B), $\Delta spoIISAB/ndoAI$ (C), $\Delta srfA$ (D), $\Delta abrB$ (E), and $\Delta epsH$ (F) strains. The scale for BF is as indicated in B. (G) Schematic of the mathematical model in which local cell density is governed by density-dependent growth and death and the ECM maintains density by preventing expansion (details are provided in SI Text). The CDPs are generated by the mathematical model. Simulations with (H) and without (I) matrix production are processed using the same correlation-based clustering analysis as experimental data.



equations that determine local cell density as a function of space and time (Fig. 4.2G and SI Text). This phenomenological model makes the following four simple assumptions that were based on and constrained by experimental observations (the parameters are listed in Table S1): (i) Local cell density is increased by heterogeneous growth, based on experimental measurements of the spatial distribution of metabolically active cells (Fig. S4C); (ii) cell death occurs when density exceeds the carrying capacity, due to build-up of local biochemical stress, as supported by the experimental observations detailed above; (iii) increase in local cell density by cell replication is counteracted by the tendency of high density to spread out in space; and (iv) expression of ECM counteracts the expansion of local cell density, consistent with comparative measurements in WT and epsH deletion strains (see Fig. S7A).

The model accounts for the basic spatiotemporal dynamics of the WT CDP (Fig. 4.2H), and it does so robustly with respect to parameter variation (Fig. S5 and SI Text). Additionally, when matrix production is turned off in the model, the simulations generate a homogeneous CDP similar to that observed in the epsH deletion strain (Fig. 4.2I). Simulations approximating dilute starting culture conditions are also consistent with experimental observations (Fig. S4 G and H and SI Text). Therefore, this population-level model accounts for the observed heterogeneity of the CDP, and thereby provides a conceptual explanation for how the ECM can promote a localized pattern.

4.4.4 Displacement Measurements Track Wrinkle Formation

In addition to its role in localized cell death presented here, the matrix is required to generate wrinkled biofilms [71, 79, 82]. To determine quantitatively where and when wrinkle formation occurs, we measured movement within the biofilm. Specifically, we tracked movement of fluorescent beads mixed in with the starting culture to measure wrinkle formation dynamics quantitatively

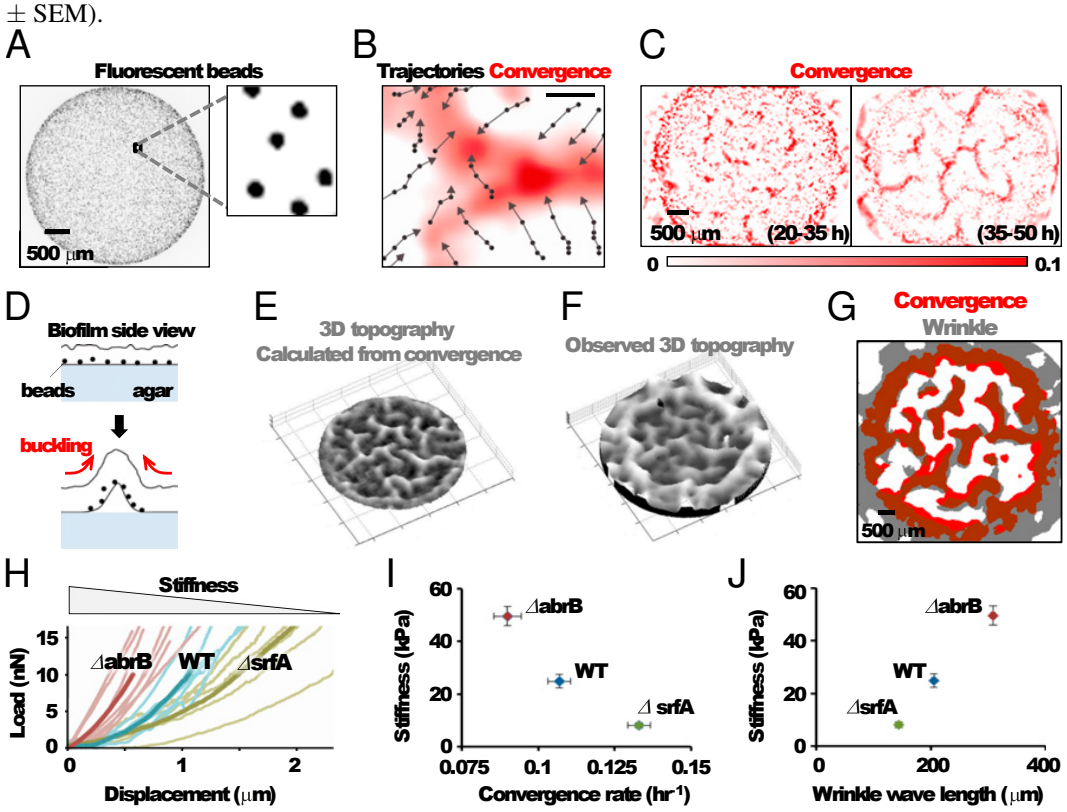
(Fig. 4.3A). We determined the trajectories of beads that were displaced by movement within the developing biofilm (Fig. 4.3B). From these trajectories, we calculated the corresponding velocity vector field of movement (SI Materials and Methods). The resulting data identified local areas at which the velocity field exhibits convergence (or negative divergence) (Fig. 4.3C), which corresponds to velocity vectors pointing toward each other (Fig. 4.3B). Furthermore, we observe disappearance of beads from the focal plane at convergence centers (Movie S1). Together with the convergence of lateral movement, these results suggest that the biofilm is pushed in the z dimension, and thus may be undergoing vertical buckling (Fig. 4.3D).

We investigated if vertical buckling could lead to the formation of wrinkles. To test this prediction, we calculated what the 3D wrinkle pattern would look like based on the measured convergence pattern (Fig. 4.3E). Specifically, we used the magnitude of convergence to assign wrinkle height in the z dimension. We then compared the simulated results with experimentally determined 3D topographies of wrinkle patterns in biofilms that were obtained using stereomicroscopy (Fig. 4.3F). As predicted, we find that the convergence pattern overlaps in space with the experimentally observed wrinkle (Fig. 4.3G). These findings support the idea that wrinkles form through vertical buckling within the biofilm.

4.4.5 Mechanical Stiffness Slows Down Wrinkle Formation Rate

Furthermore, upon examination of displacement within the biofilm we found that the displacement field indicated the formation of three dimensional wrinkles in the biofilm. Furthermore, we found that early areas of cell death corresponded to the location of the wrinkles. Later of areas of cell death occurred more sporadically throughout the biofilm but were only present when the wrinkles formed.

Figure 4.3: Movement reveals that wrinkle formation in the biofilm is generated by mechanical buckling. (A) Fluorescent beads (black) mixed with cells from the beginning of biofilm formation are shown in an inverted fluorescence image. (B) Movements of beads during biofilm formation are tracked, and their trajectories are shown as black arrows and dots. (Scale bar: 50 m.) (B and C) Convergence (negative divergence) field (red) of the interpolated vector field calculated from bead trajectories. Convergence images obtained from movements in the time window are indicated in C. (D) Schematic of the hypothesis shows convergence resulting in buckling and wrinkle formation. (E) Three-dimensional surface plot predicted from a convergence field after spatial averaging. (F) Three-dimensional surface plot of the same biofilm as in E observed from above using stereomicroscopy in the same orientation as in E. (G) Areas from E (red) and F (gray) above the respective thresholds represent wrinkle locations merged in two dimensions. (H) Load-displacement curves of WT, ΔsrfA, and ΔabrB obtained by AFM nanoindentation. The faded lines indicate actual measurements results, and the thick lines represent their mean curves. The stiffness (Youngs modulus: tensile stress divided by tensile strain) of each biofilm strain is calculated using these load-displacement curves. The maximum convergence rate determined from each convergence time trace (Fig. S6C) (I) and the wrinkle width (J) are plotted against stiffness (Youngs modulus) for WT, ΔsrfA, and ΔabrB (mean ± SEM).



Generation of wrinkles by buckling suggests that wrinkle formation would depend on the mechanical properties of the biofilm, which are, at least in part, governed by the ECM. This hypothesis is consistent with studies that have shown ECM production to be necessary for wrinkled biofilm formation [71, 79, 82]. To investigate the possible role of mechanical forces in wrinkle formation, we turned to material sciences studies of wrinkling in the bonded, thin-layered films that are commonly used in optical coatings and electronic devices, such as microprocessors. We used these studies because, as the name implies, biofilms constitute thin biological films composed of bacterial cells that are embedded within an ECM. Material sciences studies have shown that the stiffness (Youngs modulus) of such thin films determines the width of the wrinkle pattern (i.e., how far wrinkles are apart) as well as its formation rate [92, 93].

To test whether this prediction also applies to biofilms, we used AFM nanoindentation measurements to determine mechanical stiffness accurately. We used ΔabrB and ΔsrfA strains suggested to generate biofilms with varying stiffness (stiffer or softer relative to WT) due to differences in ECM production. In particular, the abrB gene product is a well-characterized repressor of ECM production [94]; therefore, its deletion is expected to generate a stiffer biofilm. In contrast, activity of the srf operon promotes ECM production [95]; thus, its deletion is expected to generate a softer biofilm. Unfortunately, biofilms formed by the epsH mutant were too soft (¡3 kPa) to measure with our experimental AFM setup, but we obtained reproducible measurements for the abrB and srfA mutant strains. We quantified the mechanical stiffness (Youngs modulus) of these biofilms using AFM nanoindentation (Fig. 4.3H). As expected, the abrB deletion strain generated the most rigid biofilm (49.6 $\pm$ 3.7 kPa, n = 16) followed by WT (25.0 $\pm$ 2.5 kPa, n = 19) and the srfA deletion strain (8.1 $\pm$ 1.1 kPa, n = 17) (Fig. 4.3H and Fig. S6 A and B). For all three strains, we then quantified convergence of bead movements during biofilm formation (as described above) to determine the

rate of wrinkle formation. Consistent with theoretical predictions developed for layered thin films, we find a negative correlation between mechanical stiffness and the rate of wrinkle formation in biofilms (Fig. 4.3I). Furthermore, we find a positive correlation between stiffness and the width of the wrinkle pattern (Figs. 2 B, D, and E and 3J), which is also consistent with predictions [92, 93]. Together, these results demonstrate that the mechanical stiffness of the biofilm, at least in part determined by the ECM, is a key property that determines wrinkle formation pattern and dynamics during biofilm development.

4.4.6 Cell Death Colocalizes with Convergence of Displacement Field

Next, we simultaneously measured bead trajectories and the CDP, which revealed that convergence of movement and localized cell death overlap in space but are separated in time. Specifically, localized cell death (18 h) occurs first and is then followed by convergence (2035 h). Magnified images clearly show that convergence within the velocity field is directed toward the center of the preceding region of cell death (Fig. 4.4 AC). As expected, convergence was not observed in the epsH deletion strain, which undergoes global and homogeneous cell death (Fig. S7A). Therefore, localized cell death not only precedes convergence of movement (Fig. 4.4D) but, more specifically, appears to focus the velocity field in space. Further analysis confirms that the CDP is predictive of the convergence pattern that emerges with a time delay of 6 ± 0.8 h (based on cross-correlation analysis, n = 6) (Fig. 4.4E). We note that facilitation of wrinkle formation by localized cell death is more pronounced in stiffer biofilms, which would either require larger forces or longer time scales to undergo buckling (Fig. 4.4F). These data suggest that regions of cell death provide a localized outlet for mechanical forces. Therefore, wrinkle formation appears to be initiated by localized cell death, which spatially focuses lateral forces, and thereby promotes vertical buckling of the biofilm (Fig. 4.4G).

4.4.7 Initiation of Cell Death Leads to Local Wrinkles

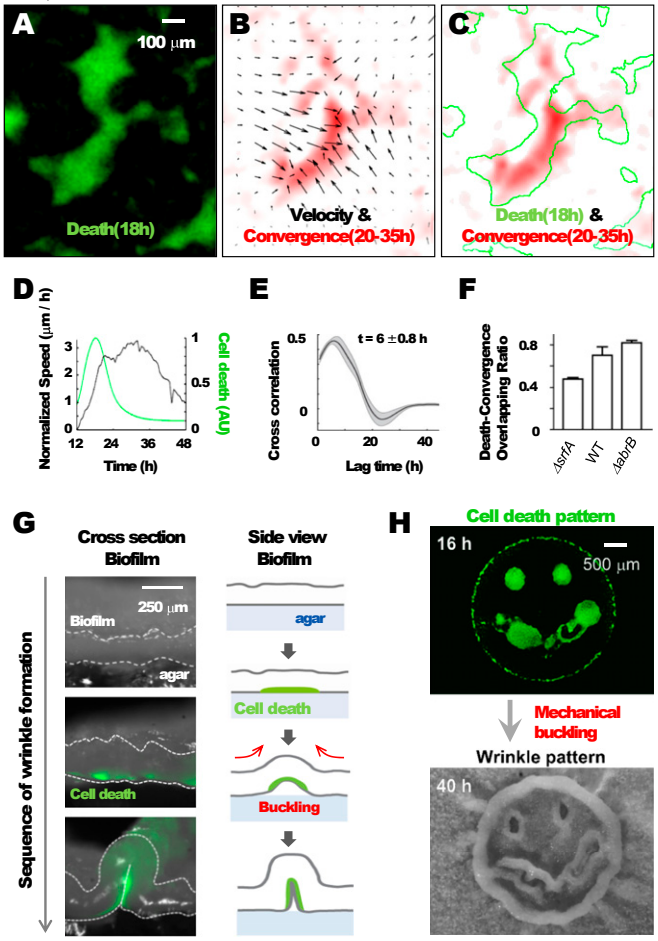
To test our finding that cell death facilitates wrinkle formation, we introduced an artificial pattern of cell density and death to generate synthetic wrinkle patterns (Fig. 4.4H). Specifically, we manually applied cells to the agar plate over the typical starting culture to generate local regions of higher cell density. As expected, these regions of high cell density gave rise to local cell death, which, in turn, generated a matching wrinkle pattern (Fig. 4.4H). Using this approach, we were able to generate large wrinkles within the biofilm reliably at arbitrary sites of our choosing (Fig. S7 C and D). In addition, direct induction of buckling by local mechanical perturbation of the biofilm generated artificial wrinkles (Fig. S7 E and F). These results further suggest that localized cell death provides an outlet for build-up compressive forces and focuses them in space, thereby resulting in buckling and consequent wrinkle formation. Self-assembly of the bacterial biofilm can thus be controlled through these principles to engineer synthetic wrinkle structures.

4.5 Discussion

Another study subsequent discovered that these wrinkles created capillary like channels by which nutrients could be distributed throughout the biofilm.

In this study, we show that death is nonuniform in space and time during biofilm formation, and appears to give rise to mechanical forces that can generate macroscopic 3D structures. Specifically, our quantitative description of the direct interaction between cellular and mechanical processes reveals that localized cell death and wrinkle formation comprise an ordered and controlled process during biofilm development. Any microbial biofilm that exhibits cell growth and death in combination with production of an ECM would be subject to mechanical forces like those characterized here. Furthermore, we note that the nucleation of wrinkle formation by cell death suggests

Figure 4.4: Localized cell death facilitated mechanics of wrinkle formation. Magnified view of localized cell death at 18 h (A), velocity field (arrows) and convergence (red) of the same location (B), and merged image of convergence (red, 2035 h) and cell death outline (green, 18 h) (C). The scale for AC is as indicated in A. (D) Average speed of fluorescent beads in the ΔepsH strain subtracted from that of the WT strain (black line: Speed wt SpeedΔepsH) (n = 3). The Sytox intensity (green) is shown for WT. The individual time traces for WT and ΔepsH are provided in Fig. S4A. AU, arbitrary unit. (E) Cross-correlation curve (between cell death and convergence) of localized cell death regions analyzed with individual biofilms. The calculated time delay (= 6 ± 0.8 h, mean ± SEM, n = 5) is the temporal offset that gives the maximum cross-correlation value between cell death and convergence. (F) Ratio of high-convergence regions that spatially correlate with local early cell death, measured for the WT, ΔsrfA, and ΔabrB strains (mean ± SEM, n = 3). (G) Cross-sectional images and schematic of the wrinkle formation process (green: cell death, light blue: agar medium). (H) Artificial smiley face CDP (Upper) created by painting higher cell density areas and a matching wrinkle pattern (Lower).



that partial killing of bacterial cells in biofilms can enhance wrinkle formation. This notion could explain why increased wrinkle formation is observed in biofilms that experience greater cell death as a result of treatment with an antimicrobial agent [96]. It is possible that partial killing of bacteria in biofilms enhances their resilience against environmental extremes through increased wrinkle formation, because wrinkling contributes to higher resistance against liquid wetting and gas penetration [75]. Therefore, we suggest that localized cell death constitutes a community-level stress response to enhance biofilm resistance under unfavorable conditions.

Mechanical forces arising from cell death reported here for biofilm development can potentially provide a conceptual framework for observations made in other biological systems, including multicellular organisms. For example, recent work has suggested that apoptosis facilitates dorsal closure during Drosophila embryogenesis [97]. Similar to the buckling of the biofilm triggered by local cell death, dorsal closure in Drosophila appears to be aided by the buckling of surrounding tissue toward the site of apoptosis. Even though the mechanism of cell death may differ, in both cases, localized cell death appears to focus mechanical forces in space to drive macroscopic movement. Therefore, utilization of localized cell death as a means to focus and direct mechanical forces spatially may be a general mechanism underlying the generation and alteration of macroscopic 3D structures during development. Furthermore, it may be possible to test and enhance our understanding of development in natural systems by attempting to design and engineer synthetic 3D patterns rationally from living organisms [64, 72, 73]. Synthetic self-organization through control over localized cell death could also serve as a tool to generate mechanical forces and perturb biological systems to investigate the interplay between cellular and mechanical processes directly in biology.

4.6 Materials and Methods

4.6.1 Strains and Growth Conditions

Table S2 lists the B. subtilis strains used in this study. All strains were routinely grown in LB or LB agar plates at 37 C. When appropriate, antibiotics were supplemented in LB at the following concentrations: 100 g/mL ampicillin, 8 g/mL neomycin, 300 g/mL spectinomycin, and 5 g/mL chloramphenicol. All strains were derived from the undomesticated WT B. subtilis strain NCIB3610.

4.6.2 Strain Constructions

Strains were created by a standard method using chromosomal integration vectors after sequence confirmation. All transformations of the NCIB3610 strain were performed by a standard one-step transformation procedure [98]. For the deletion constructs, plus-strand regions of target genes were amplified from NCIB3610 by PCR using specific primers. Amplified fragments were cloned into B. subtilis chromosomal integration vector pER449 (a kind gift from the laboratory of Wade Winkler, University of Maryland, College Park, MD). The resulting target deletion constructs were confirmed by direct sequencing and integrated into the chromosome of NCIB3610 by homologous recombination. The plasmid pSac-Cm (ECE174; Bacillus Genetic Stock Center) was used for integration to the sacA locus. Chromosomal integrations were confirmed by colony PCR using specific primers.

4.6.3 Biofilm Formation

B. subtilis strains were picked from overnight growth on LB plates and cultured in LB at 37 C. Saturated culture was grown in MSgg [5 mM potassium phosphate (pH 7.0), 100 mM 3-(N-morpholino)propanesulfonic acid (pH 7.0), 2 mM MgCl2, 700 M CaCl2, 50 M MnCl2, 100 M

FeCl3, 1 M ZnCl2, 2 M thiamine, 0.5% glycerol, 0.5% glutamate] for 1 h and was then spotted on an MSgg plate (MSgg medium supplemented with 1.5% agar, 3 mm thickness, dried overnight) and incubated at 30 C (10). For the fluorescent bead tracking, cell culture was mixed with beads (3.6 103 per colony) (FluoSpheres Carboxylate-Modified Microspheres, 1.0 m, Yellow-Green Fluorescent; Invitrogen) before plating on MSgg. Mixing the beads was confirmed not to affect the biofilm formation. When appropriate, Sytox dead cell stain (Invitrogen) was supplemented into MSgg medium, with a final concentration of 0.5 M and 1 M for Sytox Green and Blue (Invitrogen), respectively. For cross-section images of biofilm, colonies were manually sliced by razor and imaged by a Retiga 2000R digital camera (QImaging) via an SZX10 fluorescent stereomicroscope (Olympus).

4.6.4 EM Image

An overnight culture of WT B. subtilis was spotted onto MSgg agar plates and incubated at 30 C for 20 h. Bacterial colonies were fixed with 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, postfixed in 1% buffered osmium tetroxide, en bloc-stained in 4% uranyl acetate, dehydrated with a graded series of ethanol, and embedded in EMbed-812 resin (EMbed). Thin sections (80 nm) were cut on an EM UC6 ultramicrotome (Leica) and stained with 2% uranyl acetate and lead citrate. Images were acquired on a Tecnai G2 Spirit transmission electron microscope (FEI) equipped with a LaB6 source and operating at 120 kV.

4.6.5 Time-Lapse Microscopy

Time-lapse movies were obtained with custom software for automated image acquisition and microscope control. B. subtilis strains were applied onto a 1.5% agar pad made with MSgg medium supplemented with Sytox Green or Blue. Time series of images were acquired by an ORCA-ER digital camera (Hamamatsu) via an IX71 fluorescence microscope (Olympus) every 40

min at 30 C. Cells were diluted 50-fold in dDW before application onto the pad for the single-cell movies. Biofilm movies were made from the bottom using an objective lens (MPLFLN 2.5/0.08; Olympus) with a long working distance. Collected images were processed with ImageJ (National Institutes of Health, http://imagej.nih.gov/ij/) [99] and MATLAB (MathWorks).

4.6.6 Three-Dimensional Surface Topology Measurement

To measure the height of the B. subtilis biofilm (4-d-old) structures, we used an OPM-1 stereomicroscope (Zeiss) that is based on a binocular vision system, also known as 3D digital image correlation [100]. After determining the relative positions of the two cameras and intrinsic parameters, the shape of the specimen was reconstructed from the point correspondences using triangulation. Classic camera calibration approaches typically start from the ideal pinhole camera model, to which the parametric of the distortion functions are added to account for the distortion of each optical element. A 0.2-mm square-dimension chessboard (Texas Industrial Optics, Inc.) was used as the calibration board. To generate the mapping function, a flat glass slide with random paint speckles was imaged at fixed translations during in-plane motions on a 6-df stage (Optics Edmunds). Biofilm images were taken with the light source behind the biofilm to avoid the decorrelation effect of light coming from the observing side.

4.6.7 Stiffness Measurement

From nanoindentation measurements, Youngs modulus of the biofilm was calculated by fitting the load-displacement curve with the Hertz model [101]. The Hertz model for indentation depth by a conical indenter is $F = \frac{2E_r\delta^2}{\pi(1-\nu)}tan(\alpha)$, where F is the loading force, ν is Poisson's ratio (assusmed to be 0.3), δ is the indentation depth, α is the half-angle of a conical indenter, and E_r is the reduced modulus. Youngs modulus of the sample can be obtained by $\frac{1}{E_r} = \frac{1-\nu_i^2}{E_i} + \frac{1-\nu_s}{E_s}$, where

E_i and E_s are the moduli of the indenter tip and sample, respectively, and ν_i and ν_s are Poissons ratio of the indenter tip and sample, respectively. Nanoindentation measurements were conducted on an MFP-3D stand-alone AFM machine (Asylum). A conical tip with a half-angle of 30 °and a spring constant of 0.32 N/m was used for nanoindentation. All measurements were conducted in 100% glycerol at room temperature. The indentations were performed in the center area of 7-d-old biofilms. Two rounds of measurements were performed to verify the consistency of these tests. Igor Pro-6.22 (Igor) and its Asylum subroutine installed on the AFM machine were used to do nanoindentation elastic analysis. Statistical analysis was performed using GraphPad Prism.

4.7 Supplement and Supporting Information

Additional details including a movie of beads converging are located at the following URL:

`http://www.pnas.org/content/suppl/2012/09/20/1212429109.DCSupplemental`

My specific contributions to the supplement are included below.

4.7.1 Clustering-Based Image Analysis of the CDP

To analyze the pattern of several movies, we developed a correlation-based technique to group together areas of the microscope field that varied similarly over time relative to the mean. In particular, we were interested in spatiotemporal variations in the CDP as observed by Sytox fluorescence. Initial examination of the CDP revealed that certain areas of the CDP would demonstrate fluorescence at different times in the movie and that many of the fluorescence areas were highly localized. We quantified this phenomenon in both time and space. It is important to note that our primary interest was in the fluorescence of each pixel relative to the mean fluorescence of Sytox in the microscope field over time. The mean fluorescence increased sharply and then exponentially decreased in a smooth, pulse-shaped curve (Fig. 4.5A). Despite this smooth behavior, it was clear

that fluorescence, and thus cell death, was not occurring uniformly within the bacteria colony. To measure this heterogeneity, we sought to identify deviations from the mean behavior of the field and to identify areas that deviated in a correlated fashion in time. For each time-lapse microscopy frame of the Sytox signal, we first subtracted the mean fluorescence intensity at each time point from each pixel at that time point. Pixels with a value greater than the mean would have a positive value, and those with intensity less than the mean would have a negative value (Fig. 4.5B). This would also mean that areas of high initial fluorescence intensity would decrease in value if the entire microscope field increased uniformly in value.

Using this method we were able to see which areas were brighter or dimmer than the mean fluorescence over time. To segment these areas, we employed k-means clustering analysis using a Pearson correlation-based metric $(1 - r)$. Pixels that were more correlated in time were located closer together with regard to the clustering. The k-means clustering requires that the number of clusters be determined. We arbitrarily chose to identify seven clusters based on empirical results that showed this would sufficiently segment the areas into localized regions (Fig. 4.5E). From the seven clusters, we could then evaluate the mean value of the relative intensity to the frame (Fig. 4.5F). From this analysis, we could separate areas of high Sytox intensity earlier in the movie from localized clusters later in the movie. Notably, we observed that some clusters had high relative Sytox intensity earlier in the movie, corresponding to high initial cell density, and that other clusters had high relative Sytox later in the time-lapse sequence, showing a clear temporal separation (Fig. 4.5 C and D).

We conducted clustering analysis first on the entire time-lapse sequence, which produces clusters dominated by the intense earlier fluorescence from Sytox. This initial clustering analysis is used to identify earlier CDPs. An additional round of clustering to segment the late Sytox pattern

was restricted to a shorter time span starting at the inversion time point described below for a period of 16.7 h. The late Sytox pattern clusters are quantified (Fig. 4.5) and summarized in Fig. 4.2A.

In examination of the time-lapse microscopy, this clear temporal separation was usually marked by a period of uniform Sytox fluorescence intensity throughout the frame, which was identifiable by a minimum coefficient of variation (SD divided by the mean) between 16 and 23 h (Fig. 4.5F). We used this temporal inversion point to divide the clusters into early and late Sytox patterns by evaluating if the maxima of the mean relative Sytox intensity in a cluster occurred before or after this inversion point (Fig. 4.5 F and G).

The early clusters were further filtered such that maxima were greater than 0. Similarly, the maxima of the late clusters were further filtered such that they occurred after the minima and were greater than 3.1 arbitrary fluorescence units above the mean frame fluorescence. This threshold was altered to adapt to microscopy conditions of a particular time-lapse series. Once the patterns were identified using the correlation-based clustering analysis, we then wanted to distinguish the patterns quantitatively. We noticed that many of the patterns consisted of many small localized areas, whereas others were uniform throughout the field and connected.

To quantify these patterns, we used the reciprocal of the isoperimetric quotient [102]: the perimeter squared divided by the area (Fig. 4.5G). For a circular disk, this ratio is 4; thus, we normalize this value by 4, such that a circular disk has an isoperimetric quotient of 1. We used the reciprocal as a metric rather than the quotient such that this metric would increase as the perimeter increases relative to the area. This quotient was applied for the total perimeter and total area of the entire cluster. All the areas and all the perimeters of all the disconnected objects within each cluster were summed together before finding the isoperimetric quotient for each cluster. To determine a final value for the late CDP, the quotients for all the clusters meeting the late pattern criteria were

then summed together and colored red (Fig. 4.5H).

4.7.2 Displacement Vector Map

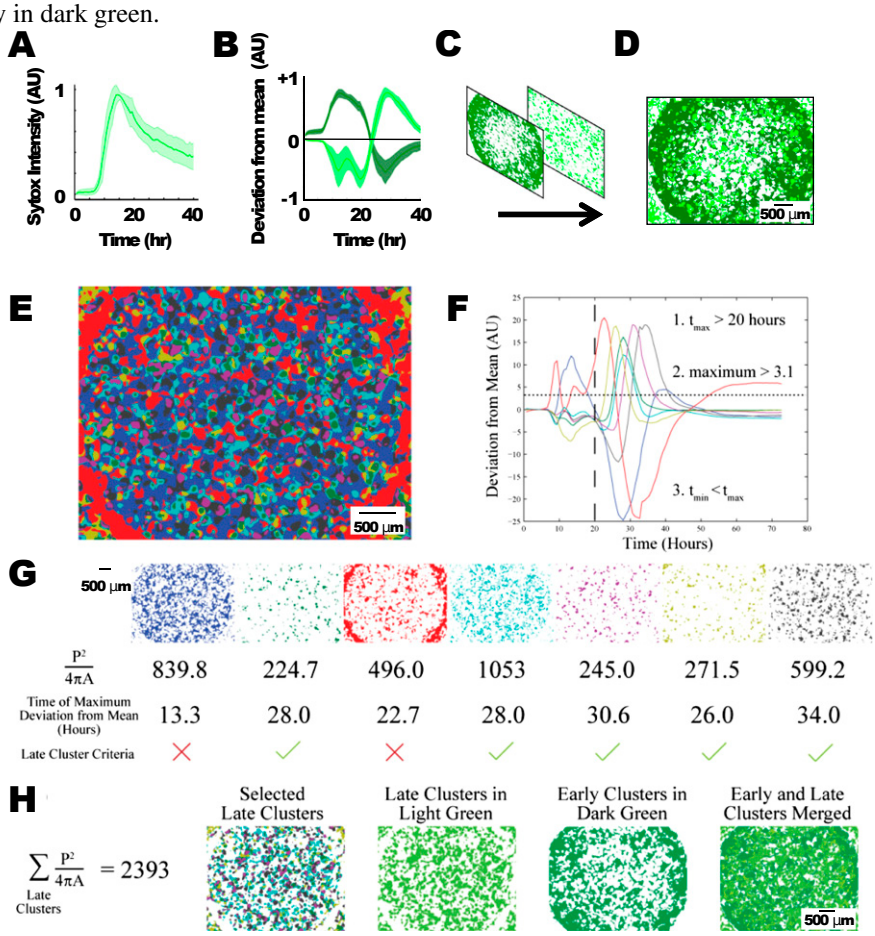
Time-lapse images of fluorescent beads embedded into B. subtilis biofilm were processed through the subtract background function in ImageJ (National Institutes of Health), and fluorescent particles were tracked using the ImageJ plug-in, MOSAIC [103]. The particle linker was modified to track a large number of fluorescent beads from the source code available under the GNU General Public License. The modifications were directed toward optimizing processor and memory use without changing the basic algorithm. Particle trajectories were then processed with MATLAB (MathWorks) to create a velocity field on a regular lattice through interpolation as a function of time. The divergence field was then calculated from this vector field.

4.8 Literature Impact Review

In this chapter, I along with my collaborators on this project identified a cell death pattern created mechanical instability leading to mechanical wrinkling of the biofilm. This mechanism appeared different than previous descriptions of biofilms being built-up to create ridges [104]. The function of these wrinkles was not well known at the time. Wilking et al. suggested shortly after publication of this work that the wrinkles create channels through which liquid transport occurs with the help of evaporative forces [105]. Together with the mechanics and cell death pattern, the biofilm seems like a multicellular community where programmed cell death facilitates the construction of a liquid distribution system.

Indeed Fig. 4.4 suggests that biofilm engineering may be possible as noted in the accompanying commentary of the work [106]. While much of the related biofilm engineering work is

Figure 4.5: Cluster Analysis of Late Cell Death Pattern (Originally Fig. S2 in Asally et al.) (A) Fluorescent time trace of mean Sytox intensity (black line) ± SEM (shaded area) from 12 time-lapse images. AU, arbitrary unit. (B) Shown in dark and light green are time traces depicting the average deviation of the two main image pixel clusters from the global Sytox signal (black line, also shown in A). The mean (solid line) SEM (shaded areas) was determined from 12 WT Sytox time-lapse images. (C) Spatial location of each of the dark-green and light-green image clusters shown in B. (D) Merged image of the dark-green and light-green pixel clusters shown in C. (E) Late Sytox pattern correlation-based k-means clustering of a WT colony. Each cluster is shown in a different color. (F) Mean intensity deviation of each cluster from the mean intensity of each frame is plotted vs. time using the same color as in A. (G) Reciprocal of the isoperimetric quotient is calculated for each individual cluster. Clusters are filtered based on their maxima according to the criteria in B. (H) Final summary metric for the CDP is the sum of quotients selected for the late clusters. All the individual late clusters are colored in light green and combined with the early clusters analyzed similarly in dark green.



directed towards fighting biofilms to fight infection for antibiotic purposes [107] or in municipal infrastructure. This work was awarded the "Bill Characklis Poster Award for Excellence in Engineering in Biofilm Research" at the ASM 6th Conference on Biofilms held in Miami, Florida [108].

Additional, directly resulting from this work, my collaborators in mechanical engineering advanced the use of stereo microscopy to measure the wrinkling deformation process in three dimensions [109]. This was an extension of my work in tracking fluorescent beads embedded into the biofilm.

While I and my collaborators focused on buckling in three dimensions, Rudge et al. discovered that buckling also allows for the creation of fractal patterns in two dimensions [110].

The interdisciplinary nature of this work brought together researchers from several fields and has inspired novel research in various areas.

4.9 Epilogue

The biochemical mechanisms concerning the cell death pattern remains unclear but is the subject of active research in the Süel lab and others. Certainly, the localization of cell death to confined areas and the ability to generate these areas from a single microcolony suggests an underlying stochastic mechanism that is amplified in a spatially limited manner.

Furthermore, it is not clear if the relation between cell death and wrinkling is general mechanism to generate rugose biofilms in other species or whether it is a specific to *Bacillus subtilis*. Recent results have shown that rugose biofilms in *Escherichia coli* UTI89 are related to iron dependent superoxide resistance implying a possible cell death related connection [111]. Related work also demonstrated that redox imbalances may play a critical role in the development of biofilms of *Pseudomonas aeruginosa* [112].

Chapter 5

Stochasticity in Host-Pathogen Interactions

5.1 Introduction

Interactions between pathogens and their hosts can be quite complex since two physiologies are now involved. Can a network of biochemical interactions be isolated such that stochasticity is analyzable?

Robert Orchard in the laboratory of Neal Alto studied a protein called Map expressed by *Enteropathogenic Escherichia coli* (EPEC). This protein is injected into a mammalian epithelial cell within the gut during early pathogenesis. Map contains a WxxxE motif making it part of a family of guanine-nucleotide exchange factor (GEF) proteins that activate Cdc42 activity leading to actin polymerization. In this particular case, Map produced filopodia around the bacteria initiating the infection. These filopodia would then serve to retain the bacteria on the surface of epithelium allowing for further pathogenesis to progress.

5.1.1 Contributions

Experimental research of Map and the polarization of filopodia was conducted by Robert C. Orchard under the supervision of Professor Neal Alto. I developed the mathematical model with the advice of Professor Gürol Süel and by consulting with the expertise of Lani Wu and Steven Altschuler. I also assisted in the interpretation of results, the creation of figures, and writing of the

original manuscript published in the journal Cell[1].

5.2 Biochemical Characterization

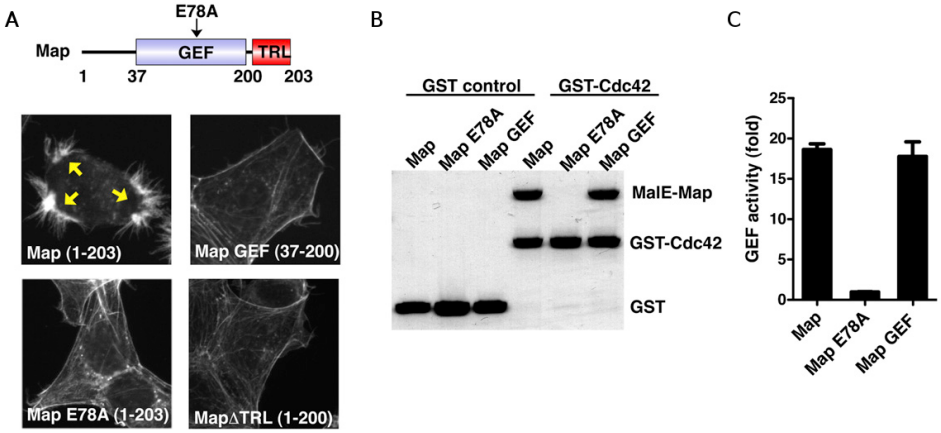
In order to better understand how this bacterium worked, Robert Orchard expressed Map ectoptically in mammalian cells without the use of EPEC. To his surprise, localized pockets of filopodia formed on the surface of the cells without any stimulation or cue.

Robert then began to characterize the biochemistry of this 203 residue protein. The guanine-nucleotide exchange factor domain occupied residues 37 through 200. Upon expressing this domain in isolation, Robert found that the GEF activity alone was necessary but not sufficient to recreate the phenotype of localized filopodia (Figure 5.1). No localized filopodia formed with the GEF domain alone. However, upon disabling the active site of the guanine-nucleotide exchange factor by mutating glumate to alanine at residue 78, he found this also eliminated the filopodia. The last three residues of tyrosine, arginine, and leucine were also found to be necessary. When residues 1 through 200 were expressed without the last three residues, foci of filopodia again failed to form. These three residues formed a ligand for a PSD-95/Disc Large/ZO-1 (PDZ) domain. Therefore, both the catalytic activity of MapGEF as well as the PDZ ligand were necessary for Map to form localized filopodia.

To better understand the function of the PDZ domain ligand, Robert explored the function of association with the PDZ domain of Ezrin binding protein 50 (Ebp50). Ebp50 is known to associate with membrane bound proteins as well as an actin bound protein called Ezrin. Which of these interactions allowed for filopodia to form in small foci? The actin binding relationship turned

[1]Reprinted from Cell, Volume 148, Orchard, R.; **Kittisopikul, M.**; Altschuler, S.; Wu, L.; Süel, G.M. & Alto, N. Identification of F-actin as the Dynamic Hub in a Microbial-Induced GTPase Polarity Circuit, Pages 803-815, with permission from Elsevier under license number 3164720976383.

Figure 5.1: Biochemical characterization of Map. A) The 203 residues of the full Map construct produces localized foci of filopodia. Deletion of either the last three residues, TRL, or disabling the catalytic activity of Map through mutation (E78A) eliminates the localization phenotype. B) GST pulldown of Cdc42 shows that the E78A mutant can no longer bind Cdc42 while MapGEF (residues 37-200) can. C) MapGEF retains catalytic activity



out to be the most significant. Fusions of the MapGEF with the PDZ domains of Ebp50 did not produce filopodia. Palmitolyzation of MapGEF, allowing for direct association with the membrane, produced filopodia along the entire surface of the cell. However, fusion of MapGEF with the actin binding domain (ABD) of Ezrin reproduced the localized filopodia seen with the full Map construct (Figure 5.2). How was actin association able to produce this phenotype?

One possibility was that association with actin allowed the MapGEF to be actively moved away from the membrane as new actin polymers were added near the membrane. Indeed, Robert was able to see Map molecules cascade away from the membrane down straight actin filaments by labeling a portion of them with a fluorescent dye. However, this would not allow for localization of the filopodia. Such a mechanism would decrease the membrane associated concentration of Map at a rate proportional to the concentration. This meant that concentrated areas of Map present at foci would experience greater removal of Map than areas with few Map molecules.

Figure 5.2: Function of the PDZ Ligand. A) Map has binding partners that relate it to both the membrane and actin filaments. B) Synthetic constructs of the GEF domain fused with various binding domains in (A). C) MapGEF fused with PDZ or palmitolyzed do not reproduce the localized filopodia seen naturally. Fusion of MapGEF with the Actin Binding Domain (ABD) of Ezrin does reproduce the localized filopodia. D) Natural localization of filopodia to EPEC. E) Map-ABD and Actin co-localize while catalytic and binding mutants do not show co-coalization of Map and Actin.

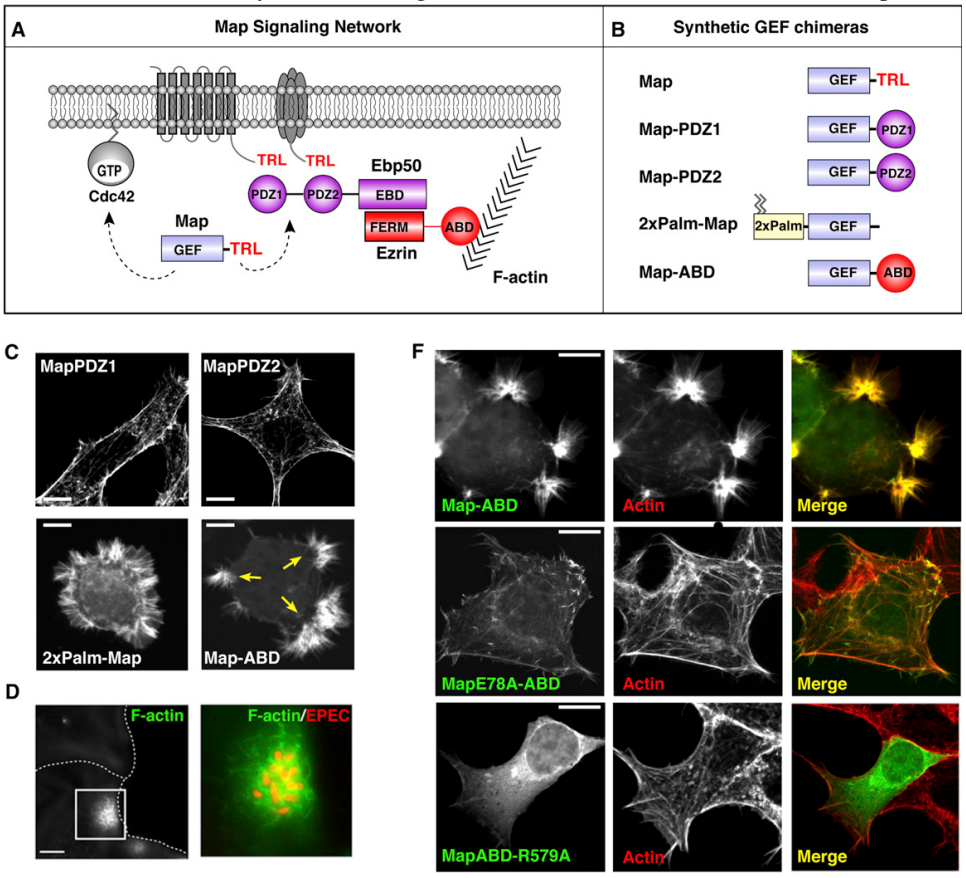
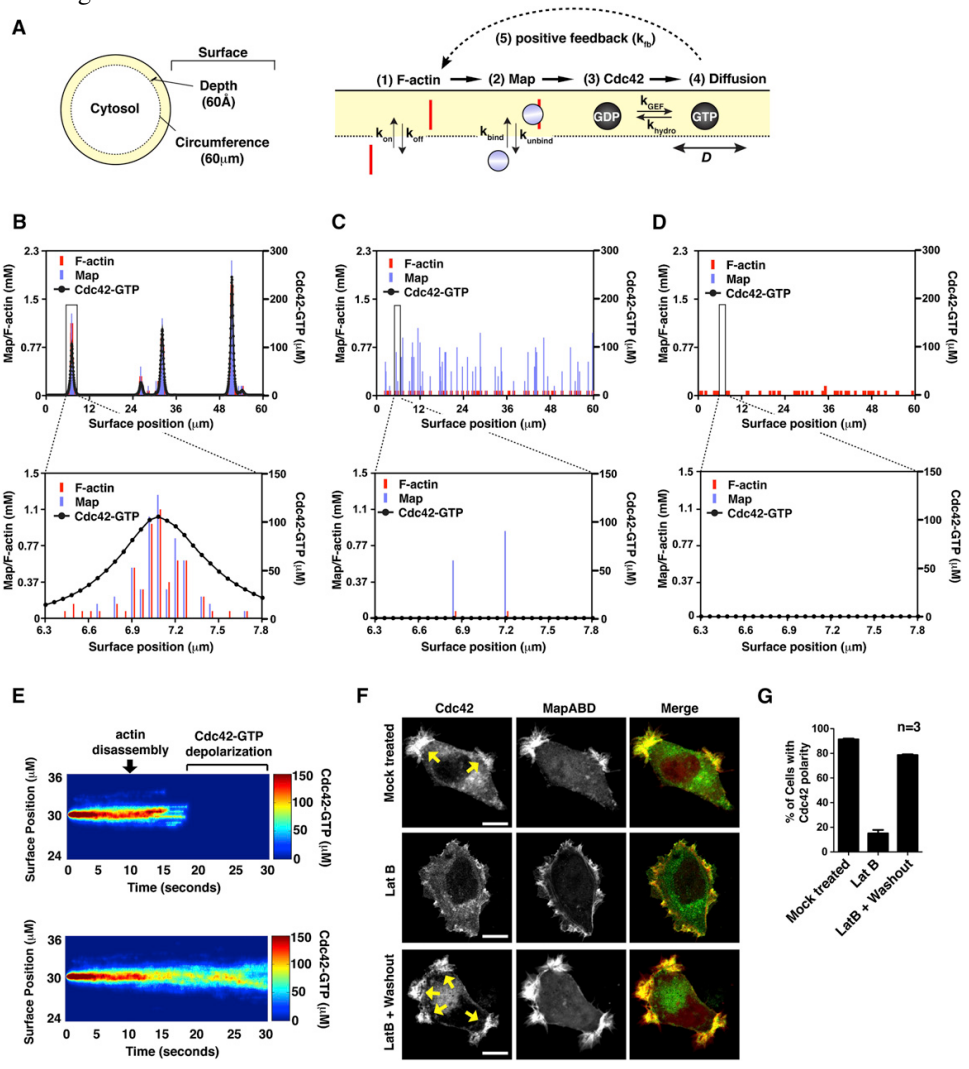


Figure 5.3: Model description and validation. A) The positive feedback model involves F-actin, Map, and Cdc42. F-actin attaches and detaches from the membrane. Map binds and unbinds F-actin. Cdc42 is activated, diffuses, and is hydrolyzed. Active Cdc42 then encourages the polymerization of more F-actin. B) The model shows localization to discrete foci. C) Deletion of GEF activity or D) Map binding to F-actin results in the lack of discrete foci. E) Turning off of actin polymerization results in rapid decay of localized foci (top) as opposed to the unperturbed model (bottom). F) Experimental validation of the model by addition of LatrunculinB that disrupts actin polymerization as in (C). Washing out of LatB restores the foci of filopodia. G) Population study of three experiments showing the effect of LatB.



Another possibility is that association with actin allowed for positive feedback loop to be created. Perhaps the association of MapGEF with actin allowed for local activation of Cdc42. This resulted in increased local additional polymerization of actin leading to more recruitment of Map in that local area as opposed to locations further away. In order to explore this possibility a detailed model was created as detailedi n the next section.

5.3 Model Description

5.3.1 Introduction

I seek to understand if the proposed molecular interactions between Map, actin, and Cdc42 are sufficient to explain the development of localized areas of filopodia on the membrane of eukaryotic cells expressing Map. Notably, the foci of filopodia appear to form both spontaneously and in response to a cue.

5.3.2 Overview

I model the association and dissociation of Actin and Map to and from the membrane associated compartments as well as the spatial distribution of a membrane diffusible species, Cdc42 (Figure 5.3A).

Actin is represented in the model by discrete actin filaments. These are actin polymers which can associate and dissociate from the membrane. Association occurs spontaneously but is also enhanced by Cdc42 signaling due to increased actin polymerization.

Map is a discrete guanine nucleotide exchange factor for Cdc42 in the model. It activates Cdc42 by exchanging GDP for GTP. Map associates to the membrane by binding to actin filaments through an actin binding domain (ABD) from Ezrin. Map is removed from the membrane through

two mechanisms. One is simply unbinding from an actin filament. Unbinding represents any event by which Map is no longer able to function as a GEF for Cdc42. This may include the removal of Map to the cytosol. Another is detachment of an actin filament from the membrane to which Map molecules are bound. When an actin filament detaches, a proportional amount of Map is removed in the relevant compartment.

Activated Cdc42 is modeled as a continuous concentration that can diffuse along the membrane. Cdc42 is activated by Map. It is inactivated via hydrolysis by GAPs that are not explicitly simulated. Cdc42 is able to diffuse laterally within the membrane and in this way provides for lateral communication of molecular signaling along the membrane. Activated Cdc42 signals to a number of downstream effectors which leads to actin polymerization and thus encourages further actin filament association to the membrane.

For the purposes of simulation, I divide the cell into many small compartments along the inner surface of the plasma membrane and a cytosolic region that is functionally away from the membrane. Each membrane surface compartment represents a small volume along the membrane of a cell that contains a discrete number of actin molecules, a discrete number of Map molecules, and a concentration of Cdc42. Actin and Map move between the membrane surface compartments and further into the cytosolic region of the cell, but are only active along the membrane. In contrast, Cdc42 is the only species that directly moves from one membrane surface compartment to another through diffusion. I model the association and dissociation of Actin and Map to and from the membrane associated compartments as well as the spatial distribution of a membrane diffusible species, Cdc42.

Actin is represented in the model by discrete actin filaments. These are actin polymers which can associate and dissociate from the membrane. Association occurs spontaneously but is

also enhanced by Cdc42 signaling due to increased actin polymerization.

Map is a discrete guanine nucleotide exchange factor for Cdc42 in the model. It activates Cdc42 by exchanging GDP for GTP. Map associates to the membrane by binding to actin filaments through an actin binding domain (ABD) from Ezrin. Map is removed from the membrane through two mechanisms. One is simply unbinding from an actin filament. Unbinding represents any event by which Map is no longer able to function as a GEF for Cdc42. This may include the removal of Map to the cytosol. Another is detachment of an actin filament from the membrane to which Map molecules are bound. When an actin filament detaches, a proportional amount of Map is removed in the relevant compartment.

Activated Cdc42 is modeled as a continuous concentration that can diffuse along the membrane. Cdc42 is activated by Map. It is inactivated via hydrolysis by GAPs that are not explicitly simulated. Cdc42 is able to diffuse laterally within the membrane and in this way provides for lateral communication of molecular signaling along the membrane. Activated Cdc42 signals to a number of downstream effectors which leads to actin polymerization and thus encourages further actin filament association to the membrane.

For the purposes of simulation, I divide the cell into many small compartments along the inner surface of the plasma membrane and a cytosolic region that is functionally away from the membrane. Each membrane surface compartment represents a small volume along the membrane of a cell that contains a discrete number of actin molecules, a discrete number of Map molecules, and a concentration of Cdc42. Actin and Map move between the membrane surface compartments and further into the cytosolic region of the cell, but are only active along the membrane. In contrast, Cdc42 is the only species that directly moves from one membrane surface compartment to another through diffusion.

5.3.3 Assumptions

1. Map binds to actin filaments that associate with the cellular membrane.
2. Map acts as guanine nucleotide exchange factor (GEF) for Cdc42.
3. Activation of Cdc42 by Map increases the likelihood of actin filament attachment by encouraging actin polymerization.
4. Polymerization of actin provides more binding partners for Map.
5. Unactivated Cdc42, bound to GDP, is assumed to be in excess such that the rate of Cdc42 activation by Map does not inversely depend on the active Cdc42 concentration.
6. Cdc42 diffuses laterally along the membrane.
7. Cdc42 signaling zones induced by ectopically expressed Map occur spontaneously.
8. Cdc42 signaling zones can be induced by by seeding Map (as when injected by a Type 3 secretion system) or by seeding Actin (through contact with a Fibronectin bead.
9. The number of binding sites for Map on an actin filament is not limiting.
10. Associated Map molecules are removed from the membrane when an actin filament detaches from the membrane.
11. The total amount of Map and actin filaments are considered to be constant over the course of the simulation.

Table 5.1: Variables used in the model

Variable	Description	Units
t	Time	sec
x	Membrane position	$60nm$
$M(t)$	MapGEF in the cytosol	molecules
$m(t) \equiv m(x,t)$	Map near the membrane	molecules
$A(t)$	Unattached actin filaments	filaments
$a(t) \equiv a(x,t)$	Actin filaments attached to the membrane	filaments
$c(t) \equiv c(x,t)$	Active Cdc42 on the membrane	μM

Table 5.2: Parameters used in the model

Parameter	Units	Description
k_{on}	sec^{-1}	Attachment rate of actin filaments to the membrane
k_{off}	sec^{-1}	Detachment rate of actin filaments from the membrane
k_{bind}	$sec^{-1}\, filament^{-1}$	Binding rate of ABD to actin
k_{unbind}	sec^{-1}	Unbinding rate of ABD to actin
k_{gef}	$sec^{-1}\mu M\, molecule^{-1}$	Catalytic rate of Cdc42 activation through GTP exchange
k_{hydro}	sec^{-1}	GAP mediated hydrolysis rate of Cdc42
D	$sec^{-1}\, um^2$	Diffusion constant of Cdc42 on the membrane
k_{fb}	$sec^{-1}\mu M^{-1}$	Cdc42 mediated actin filament attachment

5.3.4 Variables

Time is simulated in discrete and constant time steps such that events are relatively rare for each time step. A spatial aspect along the membrane is introduced by dividing the membrane into many compartments identified by x. $a(t), m(t),$ and $c(t)$ describe the amount of Actin, Map, and Cdc42 functionally associated with each membrane surface compartment at position x, respectively. $A(t)$ and $M(t)$ describe the amount of Actin and Map not functionally associated with the membrane.

5.3.5 Parameters

- k_{on} and k_{off} represent the spontaneous association and dissociation of actin filaments to the membrane, independent of Cdc42.
- k_{bind} and k_{unbind} define the binding and unbinding rates of Map to an actin filament.
- k_{gef} and k_{hydro} describe the activation of Cdc42 by Map and the deactivation of Cdc42 by GAPs, respectively.
- D describes the diffusion of Cdc42 laterally along the membrane.
- k_{fb} represents active recruitment of actin filaments in a Cdc42 dependent fashion.

5.3.6 Physical Basis for Parameters

5.3.6.1 Dimensions of the Cell and Compartments

I estimated the 2D circumference of the cell as 60 μm. Approximating the cell as a disc gives a radius of 9.55 μm (which is within $10.5 \pm 2.2 \mu m$)[2] [113, 114]. The depth of the volume near the membrane by which Map can signal to Cdc42 is approximated as 60 Angstroms or 6 nm as estimated from structural information[3].

Since the membrane is divided up into 1000 compartments, each compartment spans 60 nm. The volume of each compartment, $V_{compartment}$, is therefore $60 \times 60 \times 6 nm^3 = 2.16 \times 10^4 nm^3$.

Since a $nm^3 = 10^{-24} L$, each compartment thus has a volume of $2.16 \times 10^{-20} L$. Therefore a molar concentration in a compartment represents a density of $1 mol/L \times 6.022 \times 10^{23} molecules/mol \times$

[2] BioNumbers ID 103788. [113], Original reference [114]
[3] Figure 4A in the main text

$2.16 \times 10^{-20} L = 1.30 \times 10^4$ molecules per compartment. Thus, a 1 mM concentration in a compartment corresponds to about 13 molecules in that compartment. For the purposes of the stochastic description and simulation, we will describe actin and Map in terms of quantized units of molecules per compartment which corresponds to increments of 77 μM.

1 mM can also be converted into an area density in that 1 mM corresponds to about 13 molecules per 3600 nm^3 of membrane or 1 molecule per 277 nm^3 on average. Thus we can estimate that half of the average distance is the radius of a circle with area 277 nm^2. From $\pi r^2 = 277 nm^3, r = 9.39 nm$ or a mean intermolecular distance of 18.8 nm. The distance scales with the square root of the molecular concentration.

5.3.6.2 Diffusion Constant of Cdc42

The most directly relatable physical constant to the dimensions of the cell is the diffusion constant of Cdc42. This has been measured to be 0.036 $\mu m^2/sec$ in *S. cervisiae* and estimated be about ten times faster in *H. sapiens* due to prenylation [115]: 0.36 $\mu m^2/sec$. This corresponds to the simulation unit of $100 compartments^2/sec$. The diffusion constant is used as per Section 5.3.12.

5.3.6.3 Association and Dissociation Rate of Actin Filaments

To calibrate the association rate of actin filaments, I use the binding kinetics of Arp2/3 to WASP. The K_D has been measured to be $0.25 \mu M$ while the k_{off} has been measured to be $0.6 sec^{-1}$ [116]. This yields a calculated k_{on} rate of $2.4 \mu M^{-1} sec^{-1}$. For the simulation, I use an effective rate of 0.19 sec^{-1} or $1.9 \times 10^{-4} sec^{-1}$ per compartment for 1000 compartments. This also considers the effective amount of WASP per compartment to be approximately six. The rate used dictates that 24% of the available actin filaments in the simulation will be associated to the membrane in the absence of feedback at steady-state conditions.

5.3.6.4 Binding of Map to Actin

The binding rates of Map to Actin are derived from K_D of 500 nM for the Ezrin Actin Binding Domain (ABD) and Actin. In the natural system, Map is associated with Ezrin and its ABD through scaffolding proteins. In the constructed system, Map is tethered directly to an ABD derived from Ezrin. k_{bind} is set to 1 $filament^{-1}sec^{-1}$ and k_{unbind} to 6.5 sec^{-1} per 1000 compartments [117].

5.3.6.5 Activation and Hydrolysis of Cdc42

The estimated k_{cat} of Map is 5-19 sec^{-1} and the estimated K_M is 6-14 μM [118, 119]. I thus estimate the k_{cat} to be 10 sec^{-1} and the K_M to be $10\mu M$. The effective simulation constant for k_{gef} is 77 $\mu M/molecule/sec$ incorporating both the k_{cat} and K_M values since I do not simulate inactive Cdc42. The catalyzed hydrolysis rate has a k_{cat} of 2103.9 min^{-1} or about 35 sec^{-1} [120]. Assuming 0.1 uM GAP present, this leads to a simulated rate of $3.5sec^{-1}$.

5.3.6.6 Feedback Term: Cdc42 to Actin polymerization

The feedback term, k_{fb} is a difficult term to relate as its physical basis depends on a number of species that signal between Cdc42 and the actin polymerization machinery that are not modeled here. This term was determined on an empirical basis based upon the mean number of filopodia foci observed in parameter variation studies. The rate is 0.012 $\mu M^{-1}sec^{-1}$.

5.3.6.7 Number of Foci and Width of Foci

The number and width of Cdc42 signaling zones that form foci of filopodia are measured in conjunction with this work. The number is dependent on how many positive feedback loops can be initiated spontaneously before the available supply of Map and actin filaments is depleted. The number is thus dependent on the k_{on} and k_{fb} rates. High k_{on} rates increase the spontaneous

association of actin filaments to the membrane and thus increases the number of foci. High k_{fb} rates increases the rate at which such an association recruits more actin filaments in competition with other spatially distinct sites. Thus high k_{fb} will eventually decrease the number of foci since foci that form earlier will attract more molecules. This is examined in a parameter variation study as shown in the supplementary figures and discussed below in Section 5.3.13.

Another consideration for the number of foci is the ability to spatially distinguish them, which is a function of foci width. The width of the foci is determined by how far an activated Cdc42 molecule can diffuse before hydrolysis inactivates it. Hydrolysis subjects active Cdc42 to exponential decay with a temporal half-life of $ln(2)/k_{hydro}$. Diffusion distributes active Cdc42 in space with a standard deviation of $\sqrt{2Dt}$.

In order to analyze foci, a low threshold (2 μM) is first used to determine when the Cdc42 concentration exceeds a certain value indicating the beginning and end of focus. The number of compartments for which the concentration exceeds this value is considered the width of the focus. A higher threshold (100 μM) is then used to further screen the maxima of potential foci for areas where Cdc42 is intensely concentrated. In summary, foci of filopodia are first distinguished by a low threshold and then only counted if their maxima exceed a high threshold.

5.3.7 Conservation of Map and Actin

$$M_T = M(t) + \sum_x m(x,t) \tag{5.1}$$

$$A_T = A(t) + \sum_x a(x,t) \tag{5.2}$$

$$M(t) = M_T - \sum_x m(x,t) \tag{5.3}$$

$$A(t) = A_T - \sum_x a(x,t) \tag{5.4}$$

For the purposes of the simulation, the total amount of Actin and Map available in the cell are considered to be fixed. Essentially, I assume that production and degradation of Actin and Map remain constant and that the cell is at or near steady state conditions for these two species. The total number of Actin and Map is thus the sum of the amount that is in equivalent compartments along the inner surface of the membrane and the amount of molecules not functionally associated with the membrane.

5.3.8 Partial Differential Equations

The following is a deterministic approximation of the model. Actin and Map are simulated stochastically as discrete molecules in the next section. Cdc42 is actually modeled as a continuous variable that represents the concentration of Cdc42 near the membrane.

$$\frac{\partial a}{\partial t} = (k_{on} + k_{fb}c)(A) - k_{off}a \tag{5.5}$$

$$\frac{\partial m}{\partial t} = k_{bind}Ma - (k_{unbind} + k_{off})m \tag{5.6}$$

$$\frac{\partial c}{\partial t} = k_{gef}m - k_{hydro}c + D\nabla^2 c \tag{5.7}$$

Actin is added in an intrinsic (k_{on} term) and Cdc42 dependent manner (k_{fb} term) based upon the number of actin filaments not associated with the membrane. Actin is removed by an intrinsic,

Table 5.3: Expressions used in the stochastic model

Expression	Description
$p_{am}(t) = Pr[a(x,t) = a, m(x,t) = m]$	Probability of having a Actin and m Map at time t
$W_{am}(t)$	Transition propensity to a Actin and m Map at time t
$\dot{p}_{am} \equiv \frac{dp_{am}(t)}{dt}$	Time derivative of the probability

linear rate dependent on the amount of actin in each membrane surface compartment (k_{off} term).

Map binds to actin in each membrane surface compartment in such a way that binding sites are not consumed significantly (k_{bind} term). Map can also unbind from actin in a manner proportional to the amount of Map on the membrane (k_{unbind} term). Molecules of Map can also leave the membrane through the loss of an actin filament described by the k_{off} term for actin. A proportional amount of Map is thus removed from the membrane: $k_{off} a \frac{m}{a} = k_{off} m$.

Cdc42 is activated by Map (k_{gef} term) and hydrolyzed at a linear rate that is assumed to be catalyzed by GAPs (k_{hydro} term). Cdc42 is also able to diffuse along the membrane and thus accounts for communication between the different membrane surface compartments.

5.3.9 Stochastic Description

Actin and Map are actually simulated as discrete molecules upon which stochastic Poisson processes act.

$$\begin{aligned} W_{am} &= k_{off}(a+1)p_{a+1,m} - (k_{on} + k_{fb}c)p_{am} \\ &+ (k_{unbind} + k_{off})(m+1)p_{a,m+1} - k_{bind} a p_{am} \end{aligned} \tag{5.8}$$

$$\dot{p}_{am} = W_{am} - W_{a,m-1} - W_{a-1,m} - W_{a-1,m-1} \tag{5.9}$$

$c(x,t)$ is governed by Equation 5.7. Equation 5.9 is the master equation that describes the stochastic evolution of Map and Actin.

5.3.10 Simulation

The simulation implements the above by simulating Map and Actin events as a Poisson random processes and Cdc42 deterministically according the PDE in a specific order:

1. Remove actin as per the k_{off} term
    - Remove Map proportionally with actin k_{off} events
2. Remove Map due to unbinding from actin, k_{unbind} term
3. Add Map due to binding with actin, k_{bind} term
4. Hydrolyze Cdc42 according to exponential decay, k_{hydro} term
5. Activate Cdc42 deterministically with respect to Map, k_{gef} term
6. Diffuse Cdc42 along the membrane, D diffusion term
7. Add actin by nucleation on the membrane, k_{on} and k_{fb} term

Time progresses according to constant, discrete timesteps chosen to minimize the number of events per iteration of the simulation.

5.3.11 Implementation

I implemented a fixed time increment simulation in MATLAB that simulates actin filaments and Map stochastically while treating Cdc42 deterministically. The stochastic events are determined by using a Poisson pseudo-random number with a mean propensity according to the corresponding rate law and time increment.

With this scheme it is possible for more actin or Map to be removed from a compartment than present. This is minimized by using small time increments. In case of such a rare situation, I explicitly cap the amount of a species that can be removed from a membrane surface compartment to the amount present. Similarly, for events where more of a species is moved to a membrane surface compartment from the cytosolic compartment I randomly cancel the excess number of moves. This error correction code is rarely used and does produce warnings when run.

The deterministic terms affecting Cdc42 are integrated per term in a fixed sequence. Hydrolysis is evaluated as an exponential decay. Cdc42 activation occurs deterministically based upon the presence of Map in a compartment. Cdc42 diffusion is calculated based on convolution with a Gaussian kernel as described below.

5.3.12 Note on Diffusion of Cdc42

Cdc42 is able to diffuse laterally between nearby membrane surface compartments. This is simulated by convolution with a Gaussian kernel. The Gaussian kernel is the Green's function of of the 1D diffusion equation[121]. The 1D diffusion equation is represented here:

$$\frac{\partial c}{\partial t} = D\nabla^2 c \tag{5.10}$$

Where D represents the diffusion constant expressed in $\mu m^2/sec$. The Gaussian kernel has standard deviation $\sigma = \sqrt{2Ddt}$. dt is the small time interval used for each iteration of the simulation. Thus, the kernel is expressed as

$$G(x,dt) = \frac{1}{\sqrt{4\pi Ddt}} \exp(-\frac{x^2}{4Ddt}) \tag{5.11}$$

The solution to the diffusion equation (5.10) is the convolution of the Cdc42 with this kernel:

$$c(x,t-t_0) = \int_{-\infty}^{+\infty} G(x-y,t-t_0)c(y,t_0)dy \tag{5.12}$$

However, I note that Eqn. 5.10 is not the solution to the full Cdc42 equation (Eqn. 5.7). Use of the convolution for diffusion in this case is thus an approximation which is only valid for small timesteps.

Depending on dt the standard deviation, σ, may become less than the physical span of one compartment. Thus the simulation provides a facility by which Cdc42 may be monitored at higher spatial resolution than for Map or actin. For interaction with Map or actin the high spatial resolution Cdc42 distribution is converted to a distribution with the lower resolution of the original compartments.

5.3.13 Parameter Variation

5.3.13.1 Non-dimensional steady-state equation

I used a parameter variation study to examine how the parameters k_{on} and k_{fb} affected the number of foci that formed. More specifically, I nondimensionalized the parameters by considering the ratios of the k_{on} and feedback parameters relative to k_{off} parameter. This is justified by dividing equation 5.5 through by k_{off} and A_T:

$$\frac{1}{k_{off}A_T}\frac{\partial a}{\partial t} = (\frac{k_{on}}{k_{off}} + \frac{k_{fb}}{k_{off}}c)(\frac{A}{A_T}) - \frac{a}{A_T} \tag{5.13}$$

$$\alpha(t) \equiv \frac{\sum_x a(t)}{A_T} \tag{5.14}$$

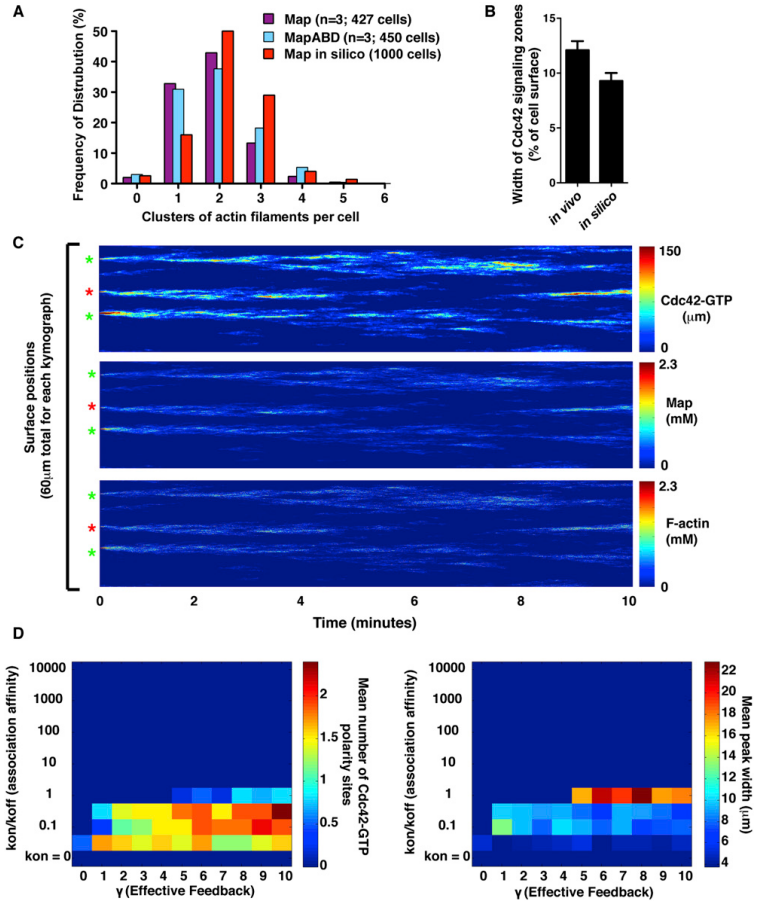
$$\frac{1}{k_{off}}\frac{d\alpha}{dt} = (\frac{k_{on}}{k_{off}} + \frac{k_{fb}}{k_{off}}C)(1-\alpha) - \alpha \tag{5.15}$$

$$\frac{1}{k_{off}}\frac{d\alpha_{ss}}{dt} = (\frac{k_{on}}{k_{off}} + \frac{k_{fb}}{k_{off}}\frac{k_{gef}}{k_{hydro}}\frac{\alpha_{ss}}{K/A_T + \alpha_{ss}}M_T)(1-\alpha_{ss}) - \alpha_{ss} \tag{5.16}$$

$$= 0 \tag{5.17}$$

Where N is the number of compartments, C is the Cdc42 concentration averaged over the compartments on the membrane (the steady state formula for this is derived below) and K is the

Figure 5.4: Compuational modeling of Map-induced polariity. A) Frequency histogram displaying the correlation between the number of Cdc42 signaling zones for Map (purple) and MapABD (cyan) expressing cells (in 3 independent experiments and over 400 cells) and the number of corresponding activity peaks generated computationally from 1000 individual cells (red). See Section 5.3.6.7. (B) Graph showing the average widths of Cdc42 signaling zones determined in vivo and in silico. 55 protrusions from 23 cells were used to calculate the mean width of Cdc42 signaling zones induced by Map in vivo. 33 protrusions from 8 simulations were used to calculate the mean width of Cdc42 signaling zones induced by Map in silico. (C) Kymograph analysis of a simulation in which the distribution of Cdc42-GTP (top), Map (middle), and F-actin (bottom) is monitored over time (x axis) in the 60*mm* virtual cell (y axis). The green asterisks mark Cdc42 signaling zones that persist through the entire 10 min simulation; whereas the red asterisks mark Cdc42 signaling zones that disappear during this time frame. These results are consistent with the longevity and dynamics of Cdc42 signaling zones observed in Map expressing cells (Figure 2A). (D) A parameter scan in which kon/koff (y axis) and g (effective feedback; x axis) were varied. The mean number of Cdc42-GTP polarity sites (left) and the average Cdc42 activity peak width (right) were counted (color bars).



effective dissociation constant for Map-actin binding:

$$C(t) \equiv \frac{1}{N}\sum_x c(t) \tag{5.18}$$

$$K \equiv \frac{k_{unbind}+k_{off}}{k_{bind}} \tag{5.19}$$

The effective feedback scaling factor, γ, is thus:

$$\gamma = \frac{k_{fb}}{k_{off}}\frac{k_{gef}}{k_{hydro}}M_T \tag{5.20}$$

$$\text{such that} \quad \frac{d\alpha_{ss}}{dt} = (\frac{k_{on}}{k_{off}}+\gamma\frac{\alpha_{ss}}{K/A_T+\alpha_{ss}})(1-\alpha_{ss})-\alpha_{ss}=0 \tag{5.21}$$

The actual steady state amount of actin on the membrane still depends on K. However, I can estimate the amount of actin on the membrane assuming total binding, $k_{bind} \gg k_{unbind}+k_{off}$ such that $K \approx 0$:

$$0 \approx N(\frac{k_{on}}{k_{off}}+\gamma)(1-\alpha_{ss})-\alpha_{ss} \tag{5.22}$$

$$\approx N(\frac{k_{on}}{k_{off}}+\gamma)-N\alpha_{ss}(\frac{k_{on}}{k_{off}}+\gamma+1) \tag{5.23}$$

$$\alpha_{ss} \approx \frac{\frac{k_{on}}{k_{off}}+\gamma}{\frac{k_{on}}{k_{off}}+\gamma+\frac{1}{N}} \tag{5.24}$$

The main difference between the k_{on} parameter and γ is that k_{on} applies equally to all membrane associated compartments, while γ is is modulated by the state of each compartment. k_{on} affects the initialization of polarity at a compartment. γ describes the strength at which a focus develops once initiated. The rate constants here are scaled for each compartment which is why the $1/N$ term is present (e.g., k_{on} is the sponteaneous binding rate for a single compartment whereas Nk_{on} is the spontaneous binding rate for the entire membrane). I directed the parameter variation efforts on understanding how k_{on} and γ affect the number of prominent foci.

5.3.13.2 Derivation of total steady state active Cdc42

Equation 5.16 is derived at steady state at conditions by first evaluating Map at steady state:

$$\frac{dM}{dt} = -k_{bind}M(A_T - A) + (k_{unbind} + k_{off})(M_T - M) \tag{5.25}$$

$$\frac{1}{k_{unbind} + k_{off}}\frac{\partial M}{\partial t} = -\frac{k_{bind}}{k_{unbind} + k_{off}}M(A_T - A) + (M_T - M) \tag{5.26}$$

$$= -\frac{1}{K}M(A_T - A) + (M_T - M) \tag{5.27}$$

$$= M_T - (1 + \frac{1}{K}(A_T - A))M \tag{5.28}$$

$$= 0 \tag{5.29}$$

$$\frac{M_{eq}}{M_T} = 1/(1 + \frac{1}{K}(A_T - A)) \tag{5.30}$$

$$= K/(K + (A_T - A)) \tag{5.31}$$

$$\frac{M_{memb}}{M_T} = \frac{M_T - M_{ss}}{M_T} = (A_T - A)/(K + (A_T - A)) \tag{5.32}$$

$$= \frac{\alpha_{ss}}{K/A_T + \alpha_{ss}} \tag{5.33}$$

Once I have the steady state amount of Map on the membrane, I then also derive the total steady state concentration of Cdc42 on the membrane:

$$\frac{dC}{dt} = k_{gef}(M_T - M) - k_{hydro}C \tag{5.34}$$

$$= 0 \tag{5.35}$$

$$C_{ss} = \frac{k_{gef}}{k_{hydro}}(M_T - M_{ss}) \tag{5.36}$$

$$= \frac{k_{gef}}{k_{hydro}}\frac{\alpha_{ss}}{K/A_T + \alpha_{ss}}M_T \tag{5.37}$$

5.3.13.3 Parameter variation results

I varied $\frac{k_{on}}{k_{off}}$ and γ, effective feedback, as explained above with Equation 5.16. The results of the effects on number of polarity sites, foci, and foci width are shown in Figure 5.4. γ was varied

by changing the k_{fb} term. Ten simulations were run for each pair of parameters using a time step of 10^{-4} seconds for 10^4 timesteps. This equates to a simulation time of one second. As expected, the number of actin filaments associated with the membrane increased directly with either $\frac{k_{on}}{k_{off}}$ or γ. I also then counted the number of foci formed as discussed above in Section 5.3.6.7. Distinct and prominent foci formed at low levels of k_{on} and increased in number with γ. In this parameter regime, few foci are initiated but those do form are able to become prominent.

At high levels of k_{on} relative to k_{off} prominent foci failed to form since many foci are initiated but they fail to become prominent or distinct. At very high levels of γ not shown in the parameter variation, the number of foci begin to decrease as one or two foci quickly become prominent and out-compete subsequent foci that may be initiated later.

At $k_{on} = 0$, no foci were initiated and thus no actin filaments associated with the membrane. At $\frac{k_{on}}{k_{off}} \times 1000 = 1$ about half of the actin filaments are associated with the membrane. The factor of 1000 is multiplied since the k_{on} rate is always evaluated for 1000 compartments on the membrane, whereas k_{off} only applies to compartments which have actin associated with them.

Foci width increases with k_{on} and decreases slightly with γ. The widths increase with k_{on} is mostly due to an increased likelihood of two foci being close together in space. The two foci are counted as one foci with greater width. The slight decrease in width with increasing γ is due to greater feedback intensity at the center of foci which are more concentrated in the middle. Foci width are mainly dependent on the diffusion constant, D, and the hydrolysis rate, k_{hydro}, as explained above in Section 5.3.6.7.

Map accumulated on the membrane in significant numbers because of high binding affinity as derived from the literature. The amount of active Cdc42 mainly increases with k_{on}. With higher

k_{on} active Cdc42 is more evenly spread out over the membrane. This prevents GAPs from reaching V_{max} and thus decreases hydrolysis in total, allowing for more Cdc42 overall.

Overall, the parameter variation shows that distinct foci of filopodia form with a low to intermediate spontaneous association rate, k_{on}, and a high effective feedback rate, γ, dependent on k_{fb} and Cdc42 dynamics relative to the spontaneous dissociation rate, k_{off}.

5.3.14 Ten minute simulation

The ten minute simulation shown in Supplemental Figure 5C was done by running the simulation at a time step of 6×10^{-4} for 10^6 timesteps yielding a total simulation time of 600 seconds or 10 minutes. Three foci spontaneously form at the beginning of the simulation. The foci are shown to be dynamic over this timespan, but are relatively stable.

5.3.15 Distribution of Number of Foci

To determine the distribution of the number and width of foci as shown in Supp. Figure A and B, 1000 simulations were run with the parameters detailed above with a time step of 10^{-4} seconds.

5.4 Model Validation

The model above derived many parameters from prior physical measurements not including in this study. The positive feedback loop model generates a number of testable predictions including:

1. The positive feedback loop can be activated by any of three molecular species: MapGEF, Cdc42, or actin.

2. Eliminating a reaction of the feedback loop would cause the foci to disappear.

3. The number of the filopodia foci present should follow the predicted distribution.

4. The width of the filopodia foci is predicted.

5.4.1 Activation of Filopodia through Actin Stimulation

An unanswered question about the MapGEF system is that in the natural system where Map is injected into the mammalian cell via a Type 3 secretion system only a single focus of filopodia is typically seen. The ability to cue a focus, however, is not contradictory to the model.

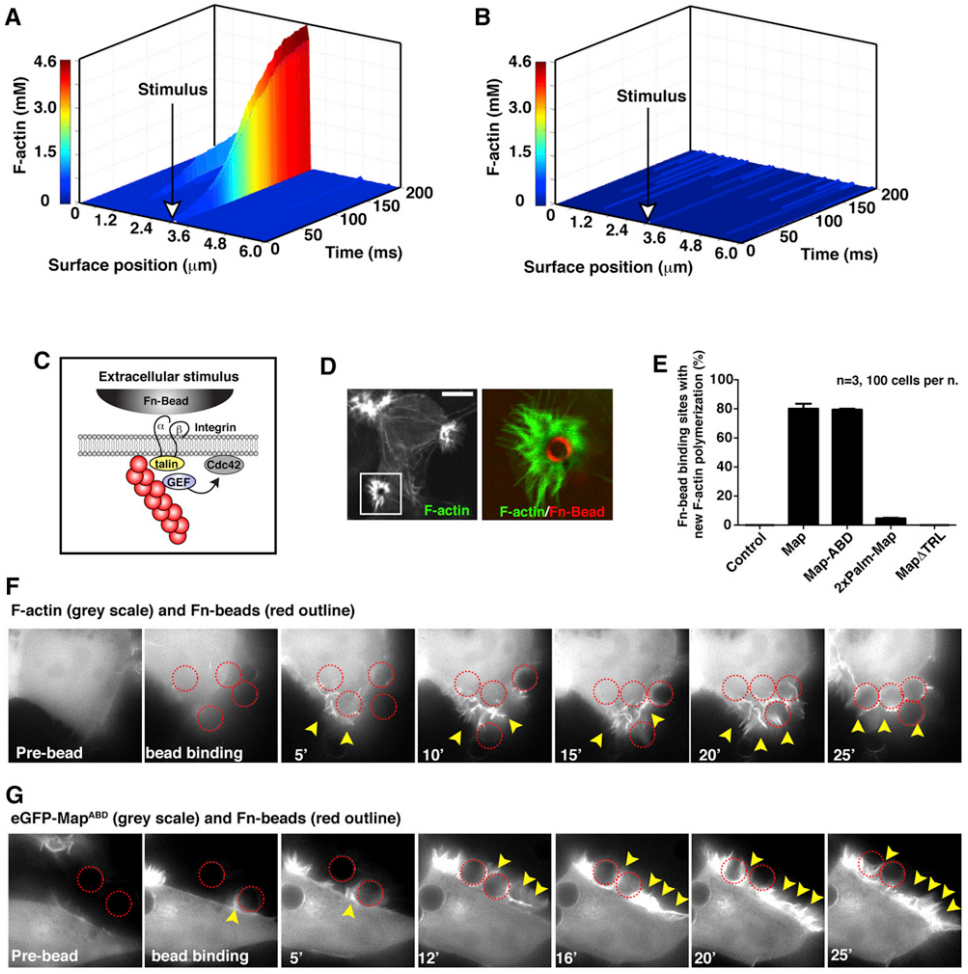
If actin polymerization were initiated at a particular location along the membrane, the model predicts that a foci would be encouraged to form at that location. In order to test this model, Robert Orchard proceeded to introduce fibronectin coated beads into a population of transfected cells (Figure 5.5). The fibronectin is known to stimulate integrins which would encourage actin polymerization. Indeed, upon contact of the beads with the membranes of the transfected cells, filopodia formed at the location and the foci persisted after the bead moved to other locations along the membrane.

Perhaps then this method of cueing a focus is used by the bacterium in order to reinforce the cue at the site of injection. EPEC could stimulate integrins in a similar manner.

5.4.2 Disruption and Rescue of Foci Using LatrunculinB

Another prediction of the model is that if the feed forward loop were disrupted then the foci would no longer be able to form in a localized fashion. A drug called LatrunculinB is able to inhibit the polymerization of actin. When Robert Orchard introduced the drug to cells with foci, those foci then disappeared. Upon subsequent washing out of the drug, the filopodia then returned in localized foci.

Figure 5.5: Cueing of the model and experimental system. A) Localization occurs when actin is seeded along the membrane. B) Localization does not occur without positive feedback. C) Schematic of fibronectin bead stimulation experiment. Fibronectin stimulates actin polymerization seeding the positive feedback loop. D) Localized filopodia triggered by fibronectin bead contact. E) Fibronectin bead stimulation of filopodia requires actin binding. F) Cells expressing Map and eGFP-ABD show actin polymerization during time-lapse microscopy. G) Cells expressing eGFP-Map shows localized recruitment of Map to sites where the fibronectin beads contacted the cell surface.

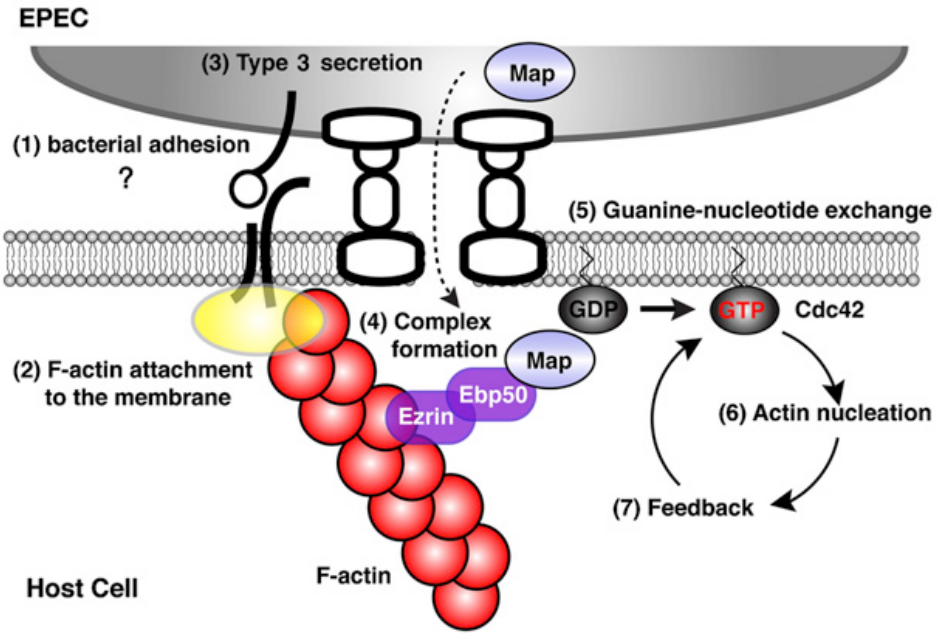


5.5 Discussion

What is the role of stochasticity in this system? Upon ectopic expression filopodia form on the membrane in stochastic and localized manner. Yet, the system is also cueable in that any stimulation of the positive feedback loop would encourage the formation of a focus at that location (Figure 5.6).

This system seems to allow for the localized foci and the cue to evolve independently. That is the formation of local pockets of filopodia is not dependent on a particular cue and the cue can be any mechanism that eventually influences one of the three key molecular species. The ability of the system to partially function in the absence of a cue is another example of robustness in the low information state.

Figure 5.6: Summary of Map model. 1) EPEC adheres to the cell membrane of an epithelial cell. 2) F-actin filaments attach to the membrane in response to integrin stimulation. 3) Map is secreted into the cytoplasm of the host cell via a Type-3 secretion system. 4) Map is associated to F-actin filaments through Ebp50 and Ezrin. 5) Guanine-nucletide exchange activity of Map activates Cdc42 in a G-protein dependent manner. 6) Cdc42 encourages actin nucleation. 7) Positive feedback occurs through the creation of local binding sites for the Map-Ebp50-Ezrin complex.



Chapter 6

Is stochasticity a selectable property?

6.1 Introduction

6.1.1 Background

Is stochasticity an evolutionary selectable trait? Probabilistic outcomes have been noted in biological systems from the molecular scale up to communities of organisms or even between cells of different organisms. Stochasticity has allowed a population of cells to have many possible phenotypes given a single microenvironment. This property diversifies the population and may allow parts of the population to be better adapted to future unknown circumstances. In other situations, such diversification is not useful and is in fact harmful to critical biological processes that require development along a strict genetic program. For example, embryogenesis of humans and other multicellular organisms follows a developmental program that allows for structured differentiation at programmed intervals and is largely not subject to stochasticity. It seems then that stochastic diversification of phenotypes is beneficial to some processes while it is deleterious to others. However, it is unclear if such differences are great enough to affect the ecological fitness of an organism. In order for organisms to evolve networks of interactions that are or are not stochastic then stochasticity must be a selectable, encodable property that confers a fitness advantage.

Horizontal gene transfer (HGT) is a phenomena that occurs in many organisms that allows them to better adapt to their environments. In *Bacillus subtilis*, a gram positive rod shaped bacterium, HGT is conducted through a transient phenotype called competence. Competence is stochastically

activated and exited permitting a portion of a population to enter competence for varying durations. While previous work has demonstrated that stochasticity may allow for greater HGT efficiency for varying concentrations of exogenous genes, it is unclear whether this could affect the fitness and furthermore the evolution of an organism.

The development of antibiotic resistance is one such way that bacteria can adapt to their environments. Meanwhile, it also poses a significant problem for human healthcare where the use of antibiotics has become a common treatment to fight dangerous infections.

6.1.2 Problem

Can encoded stochasticity of horizontal gene transfer affect the ability of *Bacillus subtilis* to evolve antibiotic resistance?

6.1.3 Hypothesis

Stochastic expression of competence in *Bacillus subtilis* will confer a greater advantage when less information about stressful conditions is known to the organism than when more information is interpretable by an organism. In contrast, deterministic expression of competence will confer advantage when information about the environment is greater.

6.1.4 Approach

We will compare a strain containing the native competence circuit of *Bacillus subtilis* with a strain containing a synthetic, less stochastic competence circuit (SynEx). Information will be provided to each strain through an inducible promoter that will be indicative of the severity of a stressful condition, the concentration of antibiotic.

6.2 Aims

6.2.1 Aim 1: Determine if there is a pre-existing fitness advantage between the native and SynEx strains without antibiotic challenge.

The construction of strains of *Bacillus subtilis* requires a series of gene additions and deletions through synthetic biology which may make one more ecologically more fit than the other even without a selection factor such as antibiotic challenge. In order to determine this, I propose measuring the growth of each strain and directly competing them within a well-mixed liquid culture. We will use a chemostat or turbidostat to determine the relative fitness of each strain. Should the strains differ significantly then modifications will be made to attempt to put their relative fitness levels on parity.

6.2.2 Aim 2: Compete native and SynEx strains directly within an environment containing antibiotics to determine if low or high information environments favors either strain.

In order to encourage the development of antibiotic resistance in the native and SynEx strains of *Bacillus subtilis* we will challenge them with antibiotics in a type of turbidostat called a morbidostat. The morbidostat attempts to maintain the concentration of bacteria at a certain level by modulating the concentration of antibiotic present rather than modulating the flow rate as in a traditional turbidostat. Along with the antibiotics, I propose to also induce competence using IPTG either in or out of phase with the antibiotic. The IPTG inducer will provide a cue to enter competence that will be interpreted stochastically or deterministically by the native and SynEx strains respectively.

6.2.3 Aim 3: Use high throughput sequencing analysis to determine if HGT is more effective in either strain towards the development of antibiotic resistance.

The development of antibiotic resistance in either the native or SynEx strains could have developed fully spontaneously rather than being assisted by the use of horizontal gene transfer. The

accumulation of spontaneous mutations within a single lineage is a possible mechanism to develop antibiotic resistance. Another mechanism using HGT would be for spontaneous but complementary mutations to occur in different lineages but then be merged within a single lineage through horizontal gene transfer. Meanwhile, there may be multiple possible mutations that could achieve similar overall functionality. Since it is not known before hand where these mutations might occur, high throughput gene sequencing will be required to locate the mutations that occur. The patterns of mutations will then be analyzed for evidence of HGT by comparing the likelihood of similar spontaneous mutations occurring in multiple lineages.

6.3 Research Strategy

In order to examine stochasticity in an evolutionary context we will examine a phenotype, competence, that is known to be stochastic but that also may have impact on the ability of an organism to adapt to a toxic environment containing antibiotics. This can be accomplished through synthetic biology since the phenotype can be triggered in an analogous but alternate manner that is similar but not stochastic. This has been implemented in the SynEx strain containing a more deterministic competence circuit. While this strategy does not evolve stochastic or deterministic properties *de novo*, it attempt to address whether stochasticity is selectable via competition.

One way that stochasticity can help an organism adapt is through bet-hedging. By diversifying the phenotypes of a population within an environment the organism is not depending on a single phenotype. Rather each phenotype might be each adapted better for a possible future environmental condition. This bet-hedging strategy only seems useful if the future environmental condition is uncertain. If an organism can accurately predict a future environmental condition then it makes more sense to solely prepare for that condition alone rather than partially committing part of the

population to the best possible phenotype. This research proposal seeks test this through the use of an inducible promoter $P_{hyperspank}$ that is used with a lactose inducer or an analog such as IPTG. The concentration of this inducer will be a surrogate for environmental cues and thus will be the interpretable information that *B. subtilis* could use. The inducible promoter will be used to drive the expression of ComK, the master regulator of competence in *B. subtilis*. This information can then be used to convey when a stressful environment is approaching allowing the bacterial population to enter into competence in anticipation.

The information conveyed by the inducer does not have to be completely accurate or useful, however. In fact, it could also be detrimental. As with natural environmental cues, the information being conveyed may not fully determine a future state or may even be unreliable. Thus the amount of information of the inducer can be varied by altering the correlation of the inducer with a stressful condition. In this way, we can create environments that either contain high or low information about impending antibiotic concentrations. If the inducer and antibiotic had a strong positive correlation, then the engineered strains could use this to enter competence at higher frequency. The difference between the strains would be whether they interpreted the information stochastically or deterministically.

6.4 Significance

The development of antibiotic resistance among pathogenic microorganisms is a severe health risk that could hamper the ability to treat infections. While the specific mutations that might permit resistance could be rare, the ability to transfer mutations or genes between lineages or species may significantly accelerate the ability of microorganisms to adapt to new antibiotics. This research would provide a way to study that phenomena in non-pathogenic bacteria with a well defined and

innate horizontal gene transfer and homologous gene recombination mechanism.

Furthermore, this study will explore the relationship between external environmental cues, stochasticity, and impact of stochasticity on evolution. Organisms must be able to interpret their environment in order to adapt to changing circumstances. The inability to adapt would decrease fitness of an organism. While adaptation may just involve changing phenotypes, it could also involve a mechanism such as horizontal gene transfer in order to acquire new abilities. HGT, however, is risky since exogenous genes are not guaranteed to be beneficial. In fact, since many of them may be sourced from dead cells, many of the of the genes of genetic alleles present may not offer a complete solution.

By understanding how bacteria adapt to the environments including the role of HGT and HGR in the development of antibiotic resistance, better strategies of antibiotic dosing can be developed.

Chapter 7

Conclusions

In this work, I have shown how stochasticity can be encoded within a biochemical network and delved into examples of how stochasticity plays a role at different scales of biology from transcription factors to biofilms. Additionally, I have proposed how stochasticity encoded within a network may be selected for by using information within the environment. In the example of feed-forward loops, I have shown how analogous circuits may provide the flexibility needed to select between regulatory networks that are more stochastic or deterministic in their behavior. Furthermore, I have provided bioinformatic evidence that certain physiological functions may use certain network patterns dependent on their noise requirements. These principles were further extended to the biochemical network governing genetic competence in *B. subtilis* where an analogous, synthetic circuit encodes different noise properties than the natural one. In that example, it is shown that the natural circuit has a wider dynamic response range that may be a tradeoff for being more stochastic.

Expanding in scale, I explored how the biofilms of *B. subtilis NCBI 3610* and the interaction between *EPEC* and mammalian host cell are subject to stochastic phenomena but are also cueable. In the case of biofilms, heterogeneity in the cell death pattern generated by stochasticity appears to create mechanical instabilities allowing for the creation of liquid channels in wrinkles. Yet the initial density of cells is able to strongly influence the initial cell death pattern. During the pathogenesis of EPEC, localized pockets of filopodia are created around the infecting *E. coli*. In the absence of the bacteria themselves, however, foci of filopodia are spontaneously formed. The ability to cue these

systems allows them to be responsive to external inputs. In both cases, localization is necessary for the patterns to be functional. In the *Bacillus* biofilm, localization allows for the mechanical forces to converge to a specific location. In the case of the filopodia, localization allows for the use of a limited number of Map molecules. Stochasticity in these cases allows for the localization to occur independent of a cue, possibly allowing for a cue to evolve separately from the interactions necessary to create localization. Thus stochasticity allows for robust localization regardless of cueing.

In the competence networks and cueable systems, the biochemical networks involved were processing environmental information. Stochasticity in that processing is advantageous in low information or uncertain environments. For the competence circuit, a bet-hedging strategy allows for the population of *Bacillus* to anticipate an uncertain future environment by not committing the entire population to a certain stress response. However, a more deterministic circuit may be more advantageous where a stressful environment appropriate for genetic competence is more certain. For the cueable systems, the cues contain information and the response to the cues are more deterministic. The stochastic properties occur in the absence of a cue, or when information from the environment is low. Stochasticity is thus advantageous in low information environments.

Since the advantage of stochasticity is related to the information content of the environment, stochasticity may then be selectable in low information environments while deterministic responses are selectable in high information environments that are interpretable. For environmental signals that are clear and reliable, a deterministic response could take greater advantage of the circumstances than a stochastic response. Returning to the example of the roulette wheel referred to in the introduction, consider the situation where the operator has rigged the roulette wheel and is giving a hint that the ball may land on black. If the operator is consistently a reliable source of information, then always following that hint would be a good betting strategy. However, if the operator

is unreliable, then it may be a better strategy to continue to hedge your bets. Interestingly, cueable systems allow for a default strategy of being stochastic in the relative absence of information while being deterministic when information is detected above a threshold.

In conclusion, organisms may take advantage of stochastic biochemical networks when faced with uncertainty in their environments. In that situation, stochasticity may allow for bet-hedging resulting in a robust adaptation to the environment. This improves the fitness of the organism. Additionally, since the stochastic property is encodable in the network of biochemical interactions that are in turn encoded in the genome, stochasticity may be a selectable property based on information in the environment. Being responsive to a cue, however, allows stochastic networks to act determininistically when given information and robustly in the absence of information. A fully deterministic system, however, does not allow for bet-hedging in the absence of information and is therefore not robust in that circumstance. Stochasticity thus allows for robustness when information is not detectable or otherwise not reliable.

Bibliography

[1] Tsai, T. Y.-C. *et al.* Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **321**, 126–129 (2008). URL http://dx.doi.org/10.1126/science.1156951.

[2] Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* **100**, 11980–11985 (2003). URL http://dx.doi.org/10.1073/pnas.2133841100.

[3] Kollmann, M., Lvdok, L., Bartholom, K., Timmer, J. & Sourjik, V. Design principles of a bacterial signalling network. *Nature* **438**, 504–507 (2005). URL http://dx.doi.org/10.1038/nature04228.

[4] Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450–461 (2007). URL http://dx.doi.org/10.1038/nrg2102.

[5] Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008). URL http://dx.doi.org/10.1038/nature07389.

[6] Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008). URL http://dx.doi.org/10.1126/science.1161427.

[7] Cagatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. & Süel, G. M. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139**, 512–

522 (2009). URL http://dx.doi.org/10.1016/j.cell.2009.07.046.

[8] Igoshin, O. A., Brody, M. S., Price, C. W. & Savageau, M. A. Distinctive topologies of partner-switching signaling networks correlate with their physiological roles. *J Mol Biol* **369**, 1333–1352 (2007). URL http://dx.doi.org/10.1016/j.jmb.2007.04.021.

[9] Klemm, K. & Bornholdt, S. Topology of biological networks and reliability of information processing. *Proc Natl Acad Sci U S A* **102**, 18414–18419 (2005). URL http://dx.doi.org/10.1073/pnas.0509132102.

[10] Wall, M. E., Hlavacek, W. S. & Savageau, M. A. Design principles for regulator gene expression in a repressible gene circuit. *J Mol Biol* **332**, 861–876 (2003).

[11] Savageau, M. A. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* **11**, 142–159 (2001). URL http://dx.doi.org/10.1063/1.1349892.

[12] Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006). URL http://dx.doi.org/10.1038/nature04588.

[13] Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719 (2007). URL http://dx.doi.org/10.1126/science.1137455.

[14] Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002). URL http://dx.doi.org/10.1126/science.298.5594.824.

[15] Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325**, 429–432 (2009). URL http://dx.doi.org/10.

1126/science.1171347.

[16] Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Boca Raton: Chapman & Hall/CRC, 2007).

[17] Mangan, S., Itzkovitz, S., Zaslaver, A. & Alon, U. The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *J Mol Biol* **356**, 1073–1081 (2006). URL http://dx.doi.org/10.1016/j.jmb.2005.12.003.

[18] Turcotte, M., Garcia-Ojalvo, J. & Sel, G. M. A genetic timer through noise-induced stabilization of an unstable state. *Proc Natl Acad Sci U S A* **105**, 15732–15737 (2008). URL http://dx.doi.org/10.1073/pnas.0806349105.

[19] Hornos, J. E. M. *et al.* Self-regulating gene: an exact solution. *Phys Rev E Stat Nonlin Soft Matter Phys* **72**, 051907 (2005).

[20] Schultz, D., Onuchic, J. N. & Wolynes, P. G. Understanding stochastic simulations of the smallest genetic networks. *J Chem Phys* **126**, 245102 (2007). URL http://dx.doi.org/10.1063/1.2741544.

[21] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69–73 (2002). URL http://dx.doi.org/10.1038/ng869.

[22] Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006). URL http://dx.doi.org/10.1126/science.1119623.

[23] Ingram, P. J., Stumpf, M. P. H. & Stark, J. Nonidentifiability of the source of intrinsic noise in gene expression from single-burst data. *PLoS Comp Biol* **4**, e1000192 (2008). URL http://dx.doi.org/10.1371/journal.pcbi.1000192.

[24] Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in bacillus subtilis. *Science* **317**, 526–529 (2007). URL http://dx.doi.org/10.1126/science.1140818.

[25] Shahrezaei, V., Ollivier, J. F. & Swain, P. S. Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* **4**, 196 (2008). URL http://dx.doi.org/10.1038/msb.2008.31.

[26] Dunlop, M. J., Cox, R. S., Levine, J. H., Murray, R. M. & Elowitz, M. B. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat Genet* **40**, 1493–1498 (2008). URL http://dx.doi.org/10.1038/ng.281.

[27] Ghosh, B., Karmakar, R. & Bose, I. Noise characteristics of feed forward loops. *Phys Biol* **2**, 36–45 (2005). URL http://dx.doi.org/10.1088/1478-3967/2/1/005.

[28] Hansen, S., Lewis, K. & Vulić, M. Role of global regulators and nucleotide metabolism in antibiotic tolerance in escherichia coli. *Antimicrob Ag Chemo* **52**, 2718–2726 (2008). URL http://dx.doi.org/10.1128/AAC.00144-08.

[29] Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem US* **81**, 2340–2361 (1997).

[30] Gillespie, D. T., Lampoudi, S. & Petzold, L. R. Effect of reactant size on discrete stochastic chemical kinetics. *J Chem Phys* **126**, 034302 (2007). URL http://dx.doi.org/10.1063/1.2424461.

[31] Keseler, I. M. *et al.* Ecocyc: a comprehensive view of escherichia coli biology. *Nucleic Acids Res* **37**, D464–D470 (2009). URL http://dx.doi.org/10.1093/nar/gkn751.

[32] Gama-Castro, S. *et al.* Regulondb (version 6.0): gene regulation model of escherichia coli k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation. *Nucleic Acids Res* **36**, D120–D124 (2008). URL http://dx.doi.org/10.1093/nar/gkm994.

[33] Kang, Y., Weber, K. D., Qiu, Y., Kiley, P. J. & Blattner, F. R. Genome-wide expression analysis indicates that fnr of escherichia coli k-12 regulates a large number of genes of unknown function. *J Bacteriol* **187**, 1135–1160 (2005). URL http://dx.doi.org/10.1128/JB.187.3.1135-1160.2005.

[34] Freed, N. E. *et al.* A simple screen to identify promoters conferring high levels of phenotypic noise. *PLoS Genet* **4**, e1000307 (2008). URL http://dx.doi.org/10.1371/journal.pgen.1000307.

[35] Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008). URL http://dx.doi.org/10.1038/nature07292.

[36] Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40**, 471–475 (2008). URL http://dx.doi.org/10.1038/ng.110.

[37] Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *J Theor Biol* **234**, 227–253 (2005). URL http://dx.doi.org/10.1016/j.jtbi.2004.11.020.

[38] Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530 (2004).

[39] Savageau, M. A. Demand theory of gene regulation. ii. quantitative application to the lactose and maltose operons of escherichia coli. *Genetics* **149**, 1677–1691 (1998).

[40] Savageau, M. A. Demand theory of gene regulation. i. quantitative development of the theory. *Genetics* **149**, 1665–1676 (1998).

[41] Schultz, D., Jacob, E. B., Onuchic, J. N. & Wolynes, P. G. Molecular level stochastic model for competence cycles in bacillus subtilis. *Proc Natl Acad Sci U S A* **104**, 17582–17587 (2007). URL http://dx.doi.org/10.1073/pnas.0707965104.

[42] Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005). URL http://dx.doi.org/10.1016/j.cell.2005.09.031.

[43] Wang, Y., Guo, L., Golding, I., Cox, E. C. & Ong, N. P. Quantitative transcription factor binding kinetics at the single-molecule level. *Biophys J* **96**, 609–620 (2009). URL http://dx.doi.org/10.1016/j.bpj.2008.09.040.

[44] Leisner, M., Kuhr, J.-T., Rdler, J. O., Frey, E. & Maier, B. Kinetics of genetic switching into the state of bacterial competence. *Biophys J* **96**, 1178–1188 (2009). URL http://dx.doi.org/10.1016/j.bpj.2008.10.034.

[45] Leisner, M., Stingl, K., Frey, E. & Maier, B. Stochastic switching to competence. *Curr Opin Microbiol* **11**, 553–559 (2008). URL http://dx.doi.org/10.1016/j.mib.2008.09.020.

[46] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004). URL http://dx.doi.org/10.1126/science.1099390.

[47] Pineda-Krch, M. Gillespiessa: Implementing the stochastic simulation algorithm in r. *J Stat Softw* **25(12)**, 1–18 (2008).

[48] Wernicke, S. & Rasche, F. Fanmod: a tool for fast network motif detection. *Bioinformatics* **22**, 1152–1153 (2006). URL http://dx.doi.org/10.1093/bioinformatics/btl038.

[49] Ma, H.-W., Buer, J. & Zeng, A.-P. Hierarchical structure and modules in the escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics* **5**, 199 (2004). URL http://dx.doi.org/10.1186/1471-2105-5-199.

[50] AC Davidson, D. H. *Bootstrap Methods and Their Applications* (Cambridge Univ Press, 1997).

[51] Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *MOLECULAR SYSTEMS BIOLOGY* **7** (2011).

[52] Frigola, D., Casanellas, L., Sancho, J. M. & Ibanes, M. Asymmetric Stochastic Switching Driven by Intrinsic Molecular Noise. *PLOS ONE* **7** (2012).

[53] Wang, P., Lu, J. & Ogorzalek, M. J. Global relative parameter sensitivities of the feed-forward loops in genetic networks. *NEUROCOMPUTING* **78**, 155–165 (2012).

[54] Rue, P. & Garcia-Ojalvo, J. Gene circuit designs for noisy excitable dynamics. *MATHEMATICAL BIOSCIENCES* **231**, 90–97 (2011).

[55] Tieri, P. *et al.* Network, degeneracy and bow tie. Integrating paradigms and architectures to grasp the complexity of the immune system. *THEORETICAL BIOLOGY AND MEDICAL MODELLING* **7** (2010).

[56] Koh, R. S. & Dunlop, M. J. Modeling suggests that gene circuit architecture controls phenotypic variability in a bacterial persistence network. *BMC SYSTEMS BIOLOGY* **6** (2012).

[57] So, L.-H. *et al.* General properties of transcriptional time series in Escherichia coli. *NATURE GENETICS* **43**, 554–U84 (2011).

[58] Yosef, N. & Regev, A. Impulse Control: Temporal Dynamics in Gene Transcription. *CELL* **144**, 886–896 (2011).

[59] Chalancon, G. *et al.* Interplay between gene expression noise and regulatory network architecture. *TRENDS IN GENETICS* **28**, 221–232 (2012).

[60] Sterlini, J. M. & Mandelstam, J. Commitment to sporulation in bacillus subtilis and its relationship to development of actinomycin resistance. *Biochem J* **113**, 29–37 (1969).

[61] Harwood, C. & Cutting, S. *Molecular Biological Methods for Bacillus* (John Wiley & Sons Ltd, West Sussex, England, 1990).

[62] Whitesides, G. M. & Grzybowski, B. Self-Assembly at All Scales. *Science* **295**, 2418–21 (2002). URL `http://www.ncbi.nlm.nih.gov/pubmed/11923529`.

[63] Camazine, S. *et al.* *Self-Organization in Biological Systems* (Princeton University Press, 2003). URL `http://press.princeton.edu/titles/7104.html`.

[64] Woodford, C. & Zandstra, P. W. Tissue engineering 2.0: guiding self-organization during pluripotent stem cell differentiation. *Current Opinion in Biotechnology* **23**, 1–10 (2012). URL http://www.ncbi.nlm.nih.gov/pubmed/22444525.

[65] Mammoto, T. & Ingber, D. E. Mechanical control of tissue and organ development. *Development* **137**, 1407–20 (2010). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2853843&tool=pmcentrez&rendertype=abstract.

[66] Savin, T. *et al.* On the growth and form of the gut. *Nature* **476**, 57–62 (2011). URL http://www.nature.com/doifinder/10.1038/nature10277.

[67] Uyttewaal, M. *et al.* Mechanical Stress Acts via Katanin to Amplify Differences in Growth Rate between Adjacent Cells in Arabidopsis. *Cell* **149**, 439–51 (2012). URL http://linkinghub.elsevier.com/retrieve/pii/S0092867412002899.

[68] Kierzkowski, D. *et al.* Elastic Domains Regulate Growth and Organogenesis in the Plant Shoot Apical Meristem. *Science* **335**, 1096–99 (2012). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1213100.

[69] Wang, Y.-C., Khan, Z., Kaschube, M. & Wieschaus, E. F. Differential positioning of adherens junctions is associated with initiation of epithelial folding. *Nature* **484**, 390–393 (2012). URL http://dx.doi.org/10.1038/nature10938.

[70] Webb, J. S., Givskov, M. & Kjelleberg, S. Bacterial biofilms: prokaryotic adventures in multicellularity. *Current Opinion in Microbiology* **6**, 578–85 (2003). URL http://linkinghub.elsevier.com/retrieve/pii/S1369527403001504.

[71] Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. Fruiting body formation by Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **98**, 11621–26 (2001). URL http://ukpmc.ac.uk/articles/PMC58779.

[72] Wood, T. K., Hong, S. H. & Ma, Q. Engineering biofilm formation and dispersal. *Trends in Biotechnology* **29**, 87–94 (2011). URL http://www.ncbi.nlm.nih.gov/pubmed/21131080.

[73] Markx, G. H., Andrews, J. S. & Mason, V. P. Towards microbial tissue engineering? *Trends in Biotechnology* **22**, 417–22 (2004). URL http://www.ncbi.nlm.nih.gov/pubmed/15283987.

[74] López, D. & Kolter, R. Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. *FEMS Microbiology Reviews* **34**, 134–49 (2010). URL http://www.ncbi.nlm.nih.gov/pubmed/20030732.

[75] Epstein, A. K., Pokroy, B., Seminara, A. & Aizenberg, J. Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. *Proc. Natl. Acad. Sci. USA* **108**, 995–1000 (2011). URL http://www.ncbi.nlm.nih.gov/pubmed/21191101.

[76] Bayles, K. W. The biological role of death and lysis in biofilm development. *Nature Reviews Microbiology* **5**, 721–6 (2007). URL http://www.ncbi.nlm.nih.gov/pubmed/17694072.

[77] Webb, J. S. *et al.* Cell death in pseudomonas aeruginosa biofilm development. *J Bacteriol* **185**, 4585–4592 (2003).

[78] Seminara, A. *et al.* Osmotic spreading of Bacillus subtilis biofilms driven by an extracellular matrix. *Proc. Natl. Acad. Sci. USA* **109**, 1116–21 (2012). URL http://www.ncbi.nlm.nih.gov/pubmed/22232655.

[79] Romero, D., Aguilar, C., Losick, R. & Kolter, R. Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *Proc. Natl. Acad. Sci. USA* **107**, 2230–4 (2010). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2836674&tool=pmcentrez&rendertype=abstract.

[80] Marvasi, M., Visscher, P. T. & Martinez, L. C. Exopolymeric substances (eps) from bacillus subtilis: polymers and genes encoding their synthesis. *FEMS Microbiol Lett* **313**, 1–9 (2010). URL http://dx.doi.org/10.1111/j.1574-6968.2010.02085.x.

[81] Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nature Reviews Microbiology* **8**, 623–33 (2010). URL http://www.nature.com/doifinder/10.1038/nrmicro2415.

[82] Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. & Losick, R. A master regulator for biofilm formation by bacillus subtilis. *Mol Microbiol* **55**, 739–749 (2005). URL http://dx.doi.org/10.1111/j.1365-2958.2004.04440.x.

[83] Gonzy-Tréboul, G., Karmazyn-Campelli, C. & Stragier, P. Developmental Regulation of Transcription of the Bacillus subtilis ftsAZ Operon. *Journal of Molecular Biology* **224**, 967–79 (1992). URL http://www.sciencedirect.com/science/article/pii/002228369290463T.

[84] Roth, B. L., Poot, M., Yue, S. T. & Millard, P. J. Bacterial Viability and Antibiotic Susceptibility Testing with SYTOX Green Nucleic Acid Stain. *Applied and Environmental Microbiology* **63**, 2421–31 (1997). URL http://aem.asm.org/cgi/content/abstract/63/6/2421.

[85] Lamsa, A., Liu, W.-T., Dorrestein, P. C. & Pogliano, K. The Bacillus subtilis cannibalism toxin SDP collapses the proton motive force and induces autolysis. *Molecular Microbiol-*

ogy **84**, 486–500 (2012). URL http://doi.wiley.com/10.1111/j.1365-2958.2012.08038.xhttp://www.ncbi.nlm.nih.gov/pubmed/22469514.

[86] Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development* **22**, 945–53 (2008). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2279205&tool=pmcentrez&rendertype=abstract.

[87] Wang, X. & Wood, T. K. Toxin/Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response. *Applied and Environmental Microbiology* **77**, 5577–83 (2011). URL http://aem.asm.org/cgi/doi/10.1128/AEM.05068-11.

[88] Reetrov, S., Florek, P., Muchov, K., Wilkinson, A. J. & Bark, I. Expression and localization of spoiisa toxin during the life cycle of bacillus subtilis. *Res Microbiol* **161**, 750–756 (2010). URL http://dx.doi.org/10.1016/j.resmic.2010.09.005.

[89] Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L. & Condon, C. The Bacillus subtilis ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. *Molecular Microbiology* **56**, 1139–48 (2005). URL http://www.ncbi.nlm.nih.gov/pubmed/15882409.

[90] Wu, X., Wang, X., Drlica, K. & Zhao, X. A Toxin-Antitoxin Module in Bacillus subtilis Can Both Mitigate and Amplify Effects of Lethal Stress. *PLoS one* **6**, e23909 (2011). URL http://www.ncbi.nlm.nih.gov/pubmed/21897862.

[91] Fu, X. *et al.* Stripe Formation in Bacterial Systems with Density-Suppressed Motility. *Physical Review Letters* **108**, 198102 (2012). URL http://link.aps.org/doi/10.1103/PhysRevLett.108.198102.

[92] Huang, R. & Im, S. H. Dynamics of wrinkle growth and coarsening in stressed thin films. *Physical Review E* **74**, 026214 (2006). URL http://link.aps.org/doi/10.1103/PhysRevE.74.026214.

[93] Cerda, E. & Mahadevan, L. Geometry and physics of wrinkling. *Phys Rev Lett* **90**, 074302 (2003).

[94] Hamon, M. A., Stanley, N. R., Britton, R. A., Grossman, A. D. & Lazazzera, B. A. Identification of AbrB-regulated genes involved in biofilm formation by Bacillus subtilis. *Molecular Microbiology* **52**, 847–60 (2004). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1409746&tool=pmcentrez&rendertype=abstract.

[95] Camp, A. H. & Losick, R. A feeding tube model for activation of a cell-specific transcription factor during sporulation in bacillus subtilis. *Genes Dev* **23**, 1014–1024 (2009). URL http://dx.doi.org/10.1101/gad.1781709.

[96] López, D., Vlamakis, H., Losick, R. & Kolter, R. Cannibalism enhances biofilm development in Bacillus subtilis. *Molecular Microbiology* **74**, 609–18 (2009). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2983100&tool=pmcentrez&rendertype=abstract.

[97] Toyama, Y., Peralta, X. G., Wells, A. R., Kiehart, D. P. & Edwards, G. S. Apoptotic force and tissue dynamics during Drosophila embryogenesis. *Science* **321**, 1683–86 (2008). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2757114&tool=pmcentrez&rendertype=abstract.

[98] Jarmer, H., Berka, R., Knudsen, S. & Saxild, H. H. Transcriptome analysis documents induced competence of bacillus subtilis during nitrogen limiting conditions. *FEMS Microbiol*

Lett **206**, 197–200 (2002).

[99] Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with imagej. *Biophotonics international* **11**, 36–42 (2004).

[100] Orteu, J.-J. 3-D computer vision in experimental mechanics. *Optics and Lasers in Engineering* **47**, 282–291 (2009).

[101] Last, J. A., Russell, P., Nealey, P. F. & Murphy, C. J. The applications of atomic force microscopy to vision science. *Invest Ophthalmol Vis Sci* **51**, 6083–6094 (2010). URL http://dx.doi.org/10.1167/iovs.10-5470.

[102] Maceachren, M. Compactness of geographic shape: Comparison and evaluation of measures. *Geogr Ann Ser B* **37**, 53–67 (1985).

[103] Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J Struct Biol* **151**, 182–195 (2005). URL http://dx.doi.org/10.1016/j.jsb.2005.06.002.

[104] Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the bacillus subtilis way. *Nature Reviews Microbiology* (2013).

[105] Wilking, J. N. *et al.* Liquid transport facilitated by channels in bacillus subtilis biofilms. *Proceedings of the National Academy of Sciences* **110**, 848–852 (2013).

[106] Schultz, D., Onuchic, J. N. & Ben-Jacob, E. Turning death into creative force during biofilm engineering. *Proceedings of the National Academy of Sciences* **109**, 18633–18634 (2012).

[107] Tanouchi, Y., Lee, A. J., Meredith, H. & You, L. Programmed cell death in bacteria and implications for antibiotic therapy. *Trends in microbiology* (2013).

[108] Haussler, S. & Fuqua, C. Biofilms 2012: New discoveries and significant wrinkles in a dynamic field. *Journal of bacteriology* **195**, 2947–2958 (2013).

[109] Hu, Z., Luo, H., Du, Y. & Lu, H. Fluorescent stereo microscopy for 3d surface profilometry and deformation mapping. *Optics express* **21**, 11808–11818 (2013).

[110] Rudge, T. J., Federici, F., Steiner, P. J., Kan, A. & Haseloff, J. Cell shape-driven instability generates self-organised, fractal patterning of cell layers. *ACS synthetic biology* (2013).

[111] DePas, W. H. *et al.* Iron induces bimodal population development by escherichia coli. *Proceedings of the National Academy of Sciences* **110**, 2629–2634 (2013).

[112] Dietrich, L. E. *et al.* Bacterial community morphogenesis is intimately linked to the intracellular redox state. *Journal of bacteriology* **195**, 1371–1380 (2013).

[113] Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750–D753 (2010). URL http://dx.doi.org/10.1093/nar/gkp889.

[114] Zhao, L. *et al.* Intracellular water-specific mr of microbead-adherent cells: the hela cell intracellular water exchange lifetime. *NMR Biomed* **21**, 159–164 (2008). URL http://dx.doi.org/10.1002/nbm.1173.

[115] Wedlich-Soldner, R., Altschuler, S., Wu, L. & Li, R. Spontaneous cell polarization through actomyosin-based delivery of the cdc42 gtpase. *Science* **299**, 1231–1235 (2003). URL http://dx.doi.org/10.1126/science.1080944.

[116] Marchand, J. B., Kaiser, D. A., Pollard, T. D. & Higgs, H. N. Interaction of wasp/scar proteins with actin and vertebrate arp2/3 complex. *Nat Cell Biol* **3**, 76–82 (2001). URL

http://dx.doi.org/10.1038/35050590.

[117] Roy, C., Martin, M. & Mangeat, P. A dual involvement of the amino-terminal domain of ezrin in f- and g-actin binding. *J Biol Chem* **272**, 20088–20095 (1997).

[118] Friebel, A. *et al.* Sope and sope2 from salmonella typhimurium activate different sets of rhogtpases of the host cell. *J Biol Chem* **276**, 34035–34040 (2001). URL http://dx.doi.org/10.1074/jbc.M100609200.

[119] Huang, Z. *et al.* Structural insights into host gtpase isoform selection by a family of bacterial gef mimics. *Nat Struct Mol Biol* **16**, 853–860 (2009). URL http://dx.doi.org/10.1038/nsmb.1647.

[120] Zhang, B., Wang, Z. X. & Zheng, Y. Characterization of the interactions between the small gtpase cdc42 and its gtpase-activating proteins and putative effectors. comparison of kinetic properties of cdc42 binding to the cdc42-interactive domains. *J Biol Chem* **272**, 21999–22007 (1997).

[121] WA, S. *Partial Differential Equations: An Introduction* (John Wiley & Sons, Inc, 1992).

[122] Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999). URL http://dx.doi.org/10.1038/16483.

[123] Auffray, C., Imbeaud, S., Roux-Rouqui, M. & Hood, L. Self-organized living systems: conjunction of a stable organization with chaotic fluctuations in biological space-time. *Philos Transact A Math Phys Eng Sci* **361**, 1125–1139 (2003). URL http://dx.doi.org/10.1098/rsta.2003.1188.

[124] Batada, N. N. & Hurst, L. D. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* **39**, 945–949 (2007). URL http://dx.doi.org/10.1038/ng2071.

[125] Bell-Pedersen, D. *et al.* Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet* **6**, 544–556 (2005). URL http://dx.doi.org/10.1038/nrg1633.

[126] Bolouri, H. & Davidson, E. H. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proc Natl Acad Sci U S A* **100**, 9371–9376 (2003). URL http://dx.doi.org/10.1073/pnas.1533293100.

[127] Dekel, E., Mangan, S. & Alon, U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys Biol* **2**, 81–88 (2005). URL http://dx.doi.org/10.1088/1478-3975/2/2/001.

[128] Dobrzynski, M. & Bruggeman, F. J. Elongation dynamics shape bursty transcription and translation. *Proc Natl Acad Sci U S A* **106**, 2583–2588 (2009). URL http://dx.doi.org/10.1073/pnas.0803507106.

[129] Eichenberger, P. *et al.* The program of gene transcription for a single differentiating cell type during sporulation in bacillus subtilis. *PLoS Biol* **2**, e328 (2004). URL http://dx.doi.org/10.1371/journal.pbio.0020328.

[130] Gibson, M. A. & Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A* **104**, 1876–1889 (2000). URL http://dx.doi.org/10.1021/jp993732q.

[131] Gillespie, D. T. Deterministic limit of stochastic chemical kinetics. *J Phys Chem B* **113**, 1640–1644 (2009). URL http://dx.doi.org/10.1021/jp806431b.

[132] Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* **58**, 35–55 (2007). URL http://dx.doi.org/10.1146/annurev.physchem.58.032806.104637.

[133] Gillespie, D. T., Cao, Y., Sanft, K. R. & Petzold, L. R. The subtle business of model reduction for stochastic chemical kinetics. *J Chem Phys* **130**, 064103 (2009). URL http://dx.doi.org/10.1063/1.3072704.

[134] Hayot, F. & Jayaprakash, C. A feedforward loop motif in transcriptional regulation: induction and repression. *J Theor Biol* **234**, 133–143 (2005). URL http://dx.doi.org/10.1016/j.jtbi.2004.11.010.

[135] Iborra, F. J., Jackson, D. A. & Cook, P. R. The case for nuclear translation. *J Cell Sci* **117**, 5713–5720 (2004). URL http://dx.doi.org/10.1242/jcs.01538.

[136] Iborra, F. J., Jackson, D. A. & Cook, P. R. Coupled transcription and translation within nuclei of mammalian cells. *Science* **293**, 1139–1142 (2001). URL http://dx.doi.org/10.1126/science.1061216.

[137] Ingram, P. J., Stumpf, M. P. H. & Stark, J. Network motifs: structure does not determine function. *BMC Genomics* **7**, 108 (2006). URL http://dx.doi.org/10.1186/1471-2164-7-108.

[138] Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U S A* **100**, 7714–7719 (2003). URL http://dx.doi.org/10.1073/pnas.1332628100.

[139] Kaneko, K. & Furusawa, C. Consistency principle in biological dynamical systems. *Theory Biosci* **127**, 195–204 (2008). URL http://dx.doi.org/10.1007/s12064-008-0034-z.

[140] Kashtan, N., Itzkovitz, S., Milo, R. & Alon, U. Topological generalizations of network motifs. *Phys Rev E Stat Nonlin Soft Matter Phys* **70**, 031909 (2004).

[141] Kuzawa, C. W. Fetal origins of developmental plasticity: are fetal cues reliable predictors of future nutritional environments? *Am J Hum Biol* **17**, 5–21 (2005). URL http://dx.doi.org/10.1002/ajhb.20091.

[142] Liu, C., Weaver, D. R., Strogatz, S. H. & Reppert, S. M. Cellular construction of a circadian clock: period determination in the suprachiasmatic nuclei. *Cell* **91**, 855–860 (1997).

[143] Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773 (2009). URL http://dx.doi.org/10.1016/j.cell.2009.06.013.

[144] Maughan, H. & Nicholson, W. L. Stochastic processes influence stationary-phase decisions in bacillus subtilis. *J Bacteriol* **186**, 2212–2214 (2004).

[145] McAdams, H. H. & Arkin, A. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet* **15**, 65–69 (1999).

[146] McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* **94**, 814–819 (1997).

[147] Mihalcescu, I., Hsing, W. & Leibler, S. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430**, 81–85 (2004). URL http://dx.doi.org/10.1038/nature02533.

[148] Pomerening, J. R., Kim, S. Y. & Ferrell, J. E. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell* **122**, 565–578 (2005). URL http://dx.doi.org/10.1016/j.cell.2005.06.016.

[149] Ratushny, A. V. *et al.* Control of transcriptional variability by overlapping feed-forward regulatory motifs. *Biophys J* **95**, 3715–3723 (2008). URL http://dx.doi.org/10.1529/biophysj.108.134064.

[150] Rust, M. J., Markson, J. S., Lane, W. S., Fisher, D. S. & O'Shea, E. K. Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* **318**, 809–812 (2007). URL http://dx.doi.org/10.1126/science.1148596.

[151] Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. *Sci STKE* **2006**, re17 (2006). URL http://dx.doi.org/10.1126/stke.3662006re17.

[152] Tabor, J. J., Bayer, T. S., Simpson, Z. B., Levy, M. & Ellington, A. D. Engineering stochasticity in gene expression. *Mol Biosyst* **4**, 754–761 (2008). URL http://dx.doi.org/10.1039/b801245h.

[153] Tripathi, T. & Chowdhury, D. Transcriptional bursts: A unified model of machines and mechanisms. *EPL (Europhysics Letters)* **84**, 68004 (6pp) (2008). URL http://stacks.iop.org/0295-5075/84/68004.

[154] Voigt, C. A., Wolf, D. M. & Arkin, A. P. The bacillus subtilis sin operon: an evolvable network motif. *Genetics* **169**, 1187–1202 (2005). URL http://dx.doi.org/10.1534/genetics.104.031955.

[155] Wall, M. E., Dunlop, M. J. & Hlavacek, W. S. Multiple functions of a feed-forward-loop gene circuit. *J Mol Biol* **349**, 501–514 (2005). URL http://dx.doi.org/10.1016/j.jmb.2005.04.022.

[156] Yamaguchi, S. *et al.* Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science* **302**, 1408–1412 (2003). URL http://dx.doi.org/10.1126/science.1089287.

[157] Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nat Rev Microbiol* **8**, 623–633 (2010). URL http://dx.doi.org/10.1038/nrmicro2415.

[158] Florek, P., Muchová, K., Pavelcíková, P. & Barák, I. Expression of functional Bacillus SpoIISAB toxin-antitoxin modules in Escherichia coli. *FEMS Microbiology Letters* **278**, 177–84 (2008). URL http://www.ncbi.nlm.nih.gov/pubmed/18096016.

[159] Garca-Ojalvo & Sancho. Colored noise in spatially extended systems. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **49**, 2769–2778 (1994).

[160] López, D., Vlamakis, H., Losick, R. & Kolter, R. Paracrine signaling in a bacterium. *Genes & Development* **23**, 1631–38 (2009). URL http://genesdev.cshlp.org/content/23/14/1631.short.

[161] Liu, W.-T. *et al.* Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **107**, 16286–90 (2010). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1008368107.

[162] Luo, P. F., Chao, Y. J., Sutton, M. A. & Peters, W. H. Accurate measurement of three-dimensional deformations in deformable and rigid bodies using computer vision. *Experimental Mechanics* **33**, 123–32 (1993). URL http://www.springerlink.com/index/10.1007/BF02322488.

[163] Monod, J. The growth of bacterial cultures. *Annual Review of Microbiology* **3**, 371–94 (1949).

[164] Nariya, H. & Inouye, M. MazF, an mRNA interferase, mediates programmed cell death during multicellular Myxococcus development. *Cell* **132**, 55–66 (2008). URL http://www.ncbi.nlm.nih.gov/pubmed/18191220.

[165] Thompson, D. W. & Architecture, O. *On Growth and Form* (Cambridge University Press, 1992). URL http://books.google.com/books?id=7_F4OUJmLFcC&lpg=PP1&pg=PP1#v=onepage&q&f=false.

[166] Yunker, P. J., Still, T., Lohr, M. A. & Yodh, A. G. Suppression of the coffee-ring effect by shape-dependent capillary interactions. *Nature* **476**, 308–311 (2011). URL http://dx.doi.org/10.1038/nature10344.