Boston University

OpenBU **http://open.bu.edu**

Theses & Dissertations Boston University Theses & Dissertations

2023

Biophysical methods bridging signal pathway architecture and dynamics in multigenerational bacterial processes

https://hdl.handle.net/2144/46262

Boston University

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

BIOPHYSICAL METHODS BRIDGING SIGNAL PATHWAY ARCHITECTURE AND DYNAMICS IN MULTIGENERATIONAL BACTERIAL PROCESSES

by

MARK SAMUEL ARONSON

B.S., Purdue University, 2017
M.S., Boston University, 2021

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2023

© 2023 by
MARK SAMUEL ARONSON
All rights reserved, except for portions of Chapter 1, which is ©2020 Current Opinion in Biomedical Engineering and Appendix B, which is ©2021 Molecular Biology of the Cell

Approved by

First Reader

Allyson E. Sgro, Ph.D.
Group Leader in Computation and Theory and 4D Cellular Physiology
Janelia Research Campus
Howard Hughes Medical Institute

Second Reader

Mary J. Dunlop, Ph.D.
Associate Professor of Biomedical Engineering

Third Reader

Ahmad S. Khalil, Ph.D.
Professor of Biomedical Engineering

Fourth Reader

Joseph W. Larkin, Ph.D.
Assistant Professor of Biology and Physics

Fifth Reader

Zeba Wunderlich, Ph.D.
Assistant Professor of Biology

“Professor,” replied Captain Nemo.
“You must not confuse static and dynamic situations,
or else you will fall into serious errors.”

Jules Verne, *20,000 Leagues Under the Sea*

“Of course you know. You’re brilliant. Everyone says so.”
“What else can they say? I do neurochemistry.
No one knows what that is.”

Don DeLillo, *White Noise*

Acknowledgments

In terms of this document, I would like to thank Emily Hager for helpful feedback and conversations, especially for Chapter 1. You always find a way to cut through and give structure to my messy thoughts. I would like to thank Allyson Sgro for editing on all of the chapters, thank you for helping me delete the second sentence every time I repeat myself, but with switched phrasing and thank you for helping make specific my writing about broad topics.

For the work here, I would like to every member of my lab over the years: Sam Ghilardi, Shwan Javdan, Achini Opathalage, Breanna O'Reilly (thank you especially for joining the lab on the same day I did), Chuqiao Huyan (for rounding out the three of us joining from the Fall 2017 cohort), Xinwen Zhu (little did I know my first-year mentee would end up becoming a labmate when we were first paired), Chiara Ricci-Tam, Noshin Nawar, Emily Hager, Maya Peters Kostman, and Sophia Kuipa. Every one of you played a role in helping craft these projects as they stand and making me the scientist that I am as I write these words.

From the lab, I would also like to thank the undergraduate research assistants who I had the honor of mentoring over the years: Vivian Shi (proud to say is now a BU BME PhD), Cullen Paulisick, Dev Mehrotra (also proud to say another BU BME PhD), and Brian Zhou. Your collective patience with me as a mentor is infinite, thank you for exploring every rabbit hole and every half-baked idea for a side project throughout the years, and thank you for helping me hone my own understanding of genetic cloning, *Bacillus subtilis*, microscopy, and coding as I worked to understand them enough to convey the ideas to you.

I was fortunate enough to have the opportunity to learn from many scientific mentors outside of the lab as well. Thank you to Allyson for sending me off to Woods

Hole my first summer in the lab so that I had the opportunity to learn from the Physical Biology of the Cell team: Rob Phillips, Jane Kondev, Julie Theriot, and Hernan Garcia. Thank you for showing me the power of quantitative hypothesis and for helping me build the foundations of my projects. I would also like to thank my committee: Mary Dunlop, Mo Khalil, Joe Larkin, and Zeba Wunderlich. Mary and Mo, thank you for allowing me to rotate through your labs and to learn your flavors of science. Joe, thank you for teaching me so much about *B. subtilis* and allowing me to be an honorary member of the lab. To James Fitzgerald, thank you for welcoming me to the world of collaborations at Janelia and for all your work in elevating the nongenetic similarity project.

Of course none of this would have happened without the person that started the lab in the first place: my advisor, Allyson Sgro. Allyson, thank you for taking a huge risk on a naive 23-year-old who had a little bit of synthetic biology experience, a little bit of coding experience, and a lot to learn about science, the world, and himself. Thank you for trusting me with the space to take risks and fail, and thank you for being there when I slipped too far. Thank you for being someone who has the courage to constantly strive to live up to their values and bringing me along for the ride to your new career stage at Janelia.

Beyond the world of science, I was surrounded with wonderful friends and family throughout my PhD journey, who all helped in their own way to see me through. To Jess, Mike, Jess, Adam, Abe, Lisa, Nicolette, Audra, Una, and Kara, thank you for being the best group of friends, for making Boston a perpetual adventure and now a city full of many memories. To Charlie, Matt, Alex, and Andrew, thank you for always knowing how to pull me out of me head for a movie or a camping trip or a road trip. To my Boston roommate Josh, thank you for agreeing to room with a friend of a friend in the same club at Purdue, and for all of the morning and evening and late night conversations about grad school and science and life. To my program cohort, especially Michael and Becks, for making the transition to grad school a joy, filled with trips to beaches and skiing, lots of dinner parties, and many nights watching The Bachelor. To the friends afar, Sam, Fouch, Hana, Lauren, Alyssa, Kate, Lauren, Beef, Delaney, and Ruby, thank you for all of the adventures criss-crossing the country over the past six years.

And last, and most importantly, the people who made it possible for me to be on this planet at all. Thank you Mom and Dad for your unending support over the years and listening to me whinge about the smallest possible things. Thank you for being excited about my move to Virginia, even though it meant moving a flight away from you. Thank you to my grandfather, Arthur Aronson, who has served as a scientific role model, especially since the summer I lived with him during my first research internship. Thank you also to my brother Joe and my (future) sister-in-law Maggie for your support and Maggie, for your grad school commiseration. Thank you finally to my partner, Blythe, who has shown me how days on days spent together full of laughter, teasing, thoughtful and challenging conversations make you know that a lifetime together will never be enough time.

Mark
February 2023

BIOPHYSICAL METHODS BRIDGING SIGNAL PATHWAY ARCHITECTURE AND DYNAMICS IN MULTIGENERATIONAL BACTERIAL PROCESSES

MARK SAMUEL ARONSON

Boston University, College of Engineering, 2023

Major Professor: Allyson E. Sgro, Ph.D.
Group Leader in Computation and Theory and 4D Cellular Physiology
Janelia Research Campus
Howard Hughes Medical Institute

ABSTRACT

Cells sense their environment and process changes through intracellular signaling networks to coordinate behavioral changes, such as cell fate decisions. In bacterial systems, these changes often occur over time periods longer than a single cell cycle. While we are now able to experimentally track and monitor these behavioral changes over multiple generations, we have a limited conceptual understanding of how these decisions are mediated by signaling pathways. Here, I present two projects that build predictive frameworks for understanding signaling pathway dynamics over multiple generations informed by the signal network architectures. In the first section, I use computational simulations to understand how signaling pathway architecture controls the duration over which related cells maintain similar concentrations of signaling pathway components following division from a common mother cell. I find that signal amplification is a requirement for similarity between related cells. In the second section, I take a joint theory-experiment approach to analyze the accumulation timescale of the signaling molecule cyclic di-GMP during biofilm initiation in the soil bacterium *B. subtilis.* Here I predict that the accumulation occurs over many generations, suggesting the possibility cyclic di-GMP is used as a cellular timer mechanism during biofilm initiation. These results both explain previous experimental findings as well as generate new predictions for how signaling pathways mediate single-cell behaviors in bacterial populations. Together, my work demonstrates the power of a joint theory-experiment approach to understand the long-term, dynamical behavior of intracellular signaling pathways by linking their architecture to their dynamical function.

Contents

1 Introduction **1**

1.1 Disclosure and Copyright Statement 1

1.2 Overview 1

1.3 Signaling networks process information between cellular inputs and outputs 2

1.4 Signaling network architecture defines timing of gene activation in bacteria 3

1.5 Biological noise 5

1.6 Network design and cellular noise 5

1.7 Noise arising from cell division 6

1.8 Outline: Examination of signal network architecture governing multigenerational processes in bacterial cells 7

2 A toy model describes the requirements for nongenetic similarity in dividing cells **10**

2.1 Introduction 10

2.2 Results 12

2.2.1 Construction of a stochastic model to simulate a minimal signaling circuit 12

2.2.2 Nongenetic similarity is present only in signaling molecule concentration 16

2.2.3 Nongenetic similarity magnitude is sensitive to reaction rate and tied to cell cycle length 19
2.2.4 Nongenetic similarity in signal molecule concentration requires signal amplification 23
2.2.5 Cell divisions bound the duration of nongenetic similarity . . . 26
2.2.6 Degradation enzyme activity decreases duration of nongenetic similarity 29
2.2.7 Inverse relationship between nongenetic similarity duration and population noise can be overcome with low enzyme noise . . . 33
2.3 Discussion 34
2.4 Methods 36

3 Biophysical investigation of cyclic di-GMP as a temporal filter for biofilm formation in *Bacillus subtilils* 39
3.1 Introduction 39
3.1.1 Bacterial biofilms as a model system for multigenerational behavior 39
3.1.2 Biofilm formation is a multigenerational process 40
3.1.3 The role of cyclic nucleotides in bacterial biofilm formation . . 40
3.1.4 *Bacillus subtilis* as a model organism for cellular decision-making 41
3.1.5 The understudied role of cyclic di-GMP in *Bacillus subtilis* biofilm formation 42
3.1.6 Existing approaches for live-cell imaging of cyclic di-GMP . . 44
3.1.7 A joint theory-experiment approach to c-di-GMP dynamics in *B. subtilis* 45
3.2 Results 46

3.2.1 Construction and parameterization of a model to simulate the c-di-GMP pathway in *Bacillus subtilis* 46

3.2.2 Quantitative prediction of c-di-GMP accumulation time in *Bacillus subtilis* 52

3.2.3 Imaging single-cell biofilm activation in *Bacillus subtilis* 55

3.2.4 Biofilm phenotype activation is stochastic and difficult to predict 57

3.2.5 Adaptation of a split GFP sensor for c-di-GMP narrowly distinguishes control strains 61

3.2.6 A novel circularly permuted GFP sensor is entangled with cell growth dynamics 66

3.3 Discussion 69

3.4 Methods and Materials 72

4 Conclusions **78**

4.1 Summary 78

4.2 Future Directions 80

4.3 Outlook 82

A Supplemental Information for Chapter 2 **84**

A.1 Acknowledgements 84

A.2 Derivations of variance variables as functions of model parameters . . 84

A.3 Supplemental Figures 102

B Quantitative Hypothesis Testing of Stress Fiber-Induced Plasma Membrane Deformation **113**

B.1 Disclosure and Copyright Statement 113

B.2 Introduction 113

B.3 Results 116
B.3.1 Estimation of energy required for membrane curvature 116
B.3.2 Estimation of energy required for fiber contraction 119
B.4 Discussion and conclusions 120
B.5 Materials and Methods 121

Bibliography **125**

Curriculum Vitae **142**

List of Tables

3.1 Knockout strains from *Bacillus* Genetic Stock Center 74

3.2 *Bacillus strains* generated for this study 74

List of Figures

2·1 Minimal signaling circuit components result in different underlying distributions . . . 13
2·2 Nongenetic similarity is present only in signaling molecule concentration 17
2·3 Nongenetic similarity extent changes with reaction rate and duration changes with cell cycle time . . . 22
2·4 Nongenetic similarity in signal molecule requires amplification 25
2·5 Nongenetic similarity decay is geometric along cell generational time . 27
2·6 Ratio of formation and degradation enzyme activity tunes duration of nongenetic similarity . . . 31
2·7 Short nongenetic similarity duration corresponds to wide population noise parameter space . . . 38

3·1 DGC production simulation tuned to single molecule data 47
3·2 Parameters for cdG-degrading enzyme tuned to match experimental data 49
3·3 Sample time traces for three simulated strains . . . 51
3·4 cyclic di-GMP accumulates in *B. subtilis* over six generations following Spo0A Activation . . . 53
3·5 Effects of free model parameters on c-di-GMP accumulation time . . 54
3·6 Example microcolony of biofilm matrix production at the single-cell level 56
3·7 Example single-cell behaviors from biofilm phenotype reporter 58
3·8 Biofilm activation and deactivation is infrequent on MSgg agarose pads 59

3·9 Pregrowth conditions are not predictive of microcolony biofilm phenotype behavior 60
3·10 Split GFP sensor distinguishes knockout strains in static imaging condition 62
3·11 Split GFP sensor output shows narrow range in timelapse context . . 65
3·12 A novel circularly permuted GFP sensor for c-di-GMP exhibits signal modulation over physiologic range in vitro 66
3·13 cpGFP sensor signal is dominated by growth rate impacting sensor concentration 68

A·1 Initial model time trace 102
A·2 Signal molecule memory result robust to random seed of model 103
A·3 Nongenetic memory in signal molecule occurs in all partition models . 104
A·4 Memory encoded in signal molecule concentration occurs under variable cell cycle time 105
A·5 Comparison of analytically-derived and numerical simulations for variable means across parameter sweeps 106
A·6 Comparison of analytically-derived and numerical simulations for variance of M-related variables across parameter sweeps 107
A·7 Comparison of analytically-derived and numerical simulations for variance of A-related variables across parameter sweeps 108
A·8 Geometric decay fits for all tested cycle times 109
A·9 Sweep of breaker enzyme parameters alone entangles nongenetic similarity duration with concentration of signal molecule 110
A·10 Sweep of V_{max} ratio while maintaining concentration of signal molecule 111
A·11 CV values of enzymes and signal molecule A over V_{max} ratio sweep . 112

B·1 Fluorescent Structures are Visible in Human Dermal Myofibroblasts Loaded with a Cytosolic Fluorophore 114
B·2 Fluorescent Structures Colocalize with Ventral Actin Stress Fibers . . 115
B·3 Schematic of a proposed mechanism for the development of the observed fluorescent structures 116
B·4 Model conceptualization of stress fiber-induced membrane deformation. 117
B·5 Calculation results from stress fiber contraction calculation. 120

List of Abbreviations

ATP	....	Adenosine triphosphate
BGSC	....	*Bacillus* Genetic Stock Center
cdG	....	cyclic diguanylate
CMV	....	cytomegalovirus
CV	....	Coefficient of Variation
DeLTA	....	Deep Learning for Time-lapse Analysis
DGC	....	diguanylate cyclase
DMEM	....	Dulbecco's Modified Eagle Medium
DNA	....	deoxyribonucleic acid
ERK	....	Extracellular signal-Regulated Kinase
ERM	....	Ezrin-Radixin-Moesin
FGM	....	Fibroblast Growth Medium
FRET	....	Förster resonance energy transfer
GFP	....	Green Fluorescent Protein
HEK	....	Human Embryonic Kidney
IDE	....	Integrated Development Environment
KTR	....	Kinase Translocation Reporter
LB	....	Lysogeny Broth
MAPK	....	Mitogen-activated protein kinase
MSgg	....	Minimal Salts Glutamate Glycerol
NEAA	....	Non-Essential Amino Acid
OD	....	Optical Density
PBS	....	Phosphate-Buffered Saline
PCR	....	Polymerase Chain Reaction
PDE	....	phosphodiesterase
pH	....	Potential Hydrogen
PMF	....	Probability Mass Function
RFP	....	Red Fluorescent Protein
RNA	....	Ribonucleic Acid
TGF-B1	....	Transforming Growth Factor Beta-1
YFP	....	Yellow Fluorescent Protein

Chapter 1

Introduction

1.1 Disclosure and Copyright Statement

This chapter adapts portions of "Exploiting noise to engineer adaptability in synthetic multicellular systems" by Mark S. Aronson, Chiara Ricci-Tam, Xinwen Zhu, and Allyson E. Sgro, 2020 Current Opinion in Biomedical Engineering. 2020; 16 (52-60). © The Authors

1.2 Overview

A defining feature of cellular life is a cell's ability to sense environmental inputs and respond with some output. The connection between these sensed inputs and outputs is a processing step, which, in cells, is done by intracellular signaling pathways. The specific architectures of these signaling pathways dictates their properties, such as the time it takes for an input signal to elicit a response. Often such a response is gene expression which, in rapidly dividing cells like bacteria, occurs on a timescale longer than a single cell cycle. In the case of such multigenerational processes, cell signaling is complicated by two factors: first, cellular noise arising from the stochastic nature of biochemical interactions and second, how the signaling is maintained upon cell division. Different network architectures can limit, increase, or exploit cellular noise, but how signal network architecture affects signaling on a multigenerational timescale remains an open area of inquiry. In this dissertation, I take a theoretical approach to understand how network architecture can enable cell similarity over multiple generations and how that duration can be altered with different network architectures. I then take a joint theory-experiment approach to quantitatively hypothesize and lay the groundwork to test if a natural pathway related to biofilm formation operates as a cellular timer mechanism, integrating information over multiple cell generations. These studies serve to further our knowledge in how cell signaling network design functions in the context of multigenerational processes.

1.3 Signaling networks process information between cellular inputs and outputs

Bacteria make decisions about how to behave by sensing changes in their environment, often through membrane-bound receptor proteins [1]. Bacterial receptors are capable of sensing a variety of environmental changes, including chemical [2], electrical [3], and mechanical [4]. These receptors then transduce the sensed input through signaling networks in order to alter the cell's behavior, whether that is to tune flagellar turn bias so bacteria may chemotax up a gradient [5], to initiate group behaviors through quorum sensing [6], or to self-program cell death [7]. But signal pathways are more than just conduits of information; their diverse structures and dynamic properties enable them to perform more complex functions than simple relay [8]. Feedback loops enable pathways to respond differently to different exposures of the same ligand [9]. Other network architectures enable a cell to not only change behavior upon exposure to an input signal, but to maintain that behavior after the signal has been withdrawn, a form of "cellular memory" [10]. To understand how cells sense and respond to their environment requires a deep comprehension of the signal pathways connecting the sense and response: their static structure along with their dynamical properties [11]. Doing this experimentally requires a suite of tools, from effective biosensors [12] to high spatiotemporal resolution imaging [13] to well-developed algorithms for segmenting and tracking the resulting image data [14]. Even with many advances in these areas, from improved sensors to novel deep-learning algorithms, interpreting experimental data from these pathways remains challenging [11]. To address these challenges, joint theory-experiment approaches have proven effective at dissecting between possible different mechanisms and structures underlying experimental results [15, 16, 17] and have served as an effective strategy to advance our understanding of signal pathway dynamical structure and function. To illustrate the utility of joint-theory approaches for dissecting the impact of signal network architecture on signaling dynamics, let's first examine, as a useful case study, a collection of studies mapping out the impact of network architecture for the timing of simple gene activation. While my work focuses on second messenger signal molecules and their pathways, not transcriptional gene networks, this example will serve as a methodological framing for the approaches I take in the later chapters.

1.4 Signaling network architecture defines timing of gene activation in bacteria

The simplicity of single gene activation in a bacterial system makes it a useful example for understanding how signal pathway structure alters dynamics. This is because proteins in growing bacteria are rarely actively degraded [18], resulting in the concentration of protein being determined only by gene expression and dilution due to cell division. In this example, consider a bacterial cell responding to some environmental signal. A common signal motif in bacteria is the two-component system [19], in which a transmembrane histidine kinase senses an external ligand, autophosphorylates, and then transfers the phosphyl group to a response regulator protein. This response regulator, when phosphorylated, acts as a transcription factor and proceeds to turn on expression of a target gene. Nearly all of the steps of this pathway happen quickly: ligand binding to the membrane receptor, autophosphorylation of the histidine kinase, transphosphorylation of the response regulator, and binding of the response regulator to the promoter all happen on the order of milliseconds to seconds. Even the mechanics of gene expression, transcription and translation, occur on the order of five minutes for a single protein. What does set the timescale of gene activation is the amount of time it takes to accumulate the newly expressed protein to steady-state levels. In fact, for any protein activated with this architecture, accumulation of new protein to half of its new level takes an entire cell cycle [20], which, in growing *E. coli*, can range from 30 min to hours. This time duration of one cell cycle serves as a benchmark for what a typical response time is in a bacterial gene activation network.

Different signal network architectures can modify the dynamics around this benchmark. One modification often seen in nature is to have the protein product regulate itself [21]. If the protein product negatively regulates itself, that is, represses its own production, it is possible to have the response time be faster than a cell cycle [22, 23]. This simple change in how the network is structured leads to a different timing of the output. On the other hand, positive regulation enables a slower response time [24]. By adding different versions of autoregulation, that is, changing the network architecture, the dynamics of the response in simple gene activation is changed.

While these insights into the relationship between network architecture and dynamics are useful, they rely on assumptions that the molecules involved behave in a deterministic and continuous manner. In reality, bacterial systems are composed of a discrete number of molecules whose interactions are driven by probabilistic, thermodynamic laws. The randomness of these interactions leads to variations across systems and across time, variations that have been coined biological noise.

1.5 Biological noise

The term 'noise' originates from electrical engineering and broadly refers to random disturbances to a signal. Noise has been well studied and characterized in neuroscience, where it is recognized to play a prominent role in determining the structure and function of nervous systems [25]. As appreciation for the information processing abilities of single cells grew, the concept of noise was ported over to cell biology, where noise refers to variations in the numbers of molecules, such as mRNAs or proteins, that arise from biochemical reactions. A landmark quantitative examination of noise in cell biology classified it as coming from two sources: extrinsic noise, variations arising between cells, and intrinsic noise, variations arising within a cell [26]. This phenomenon has been incredibly well studied over the past twenty years [27, 28, 29, 30, 31, 32, 33], but exact definitions of biological noise and its origins are the source of some debate in the field [33]. Regardless of its exact nature, cellular noise, or at least variation, is a real phenomenon, one that often has negative connotations. But bacterial systems, in some cases, have evolved to use noise as a feature, especially when driving differentiation of isogenic populations.

1.6 Network design and cellular noise

Bacterial systems can use noise to drive phenotypic differentiation among genetically-identical cells in a population. One canonical case is the *Bacillus subtilis* competence circuit [34, 35]. Competence is a bacterial survival strategy in which cells readily take up DNA from the surrounding environment. However, the strategy is risky, as DNA in the surrounding environment could include hostile entities such as bacteriophages. To balance the risk of uptaking surrounding DNA with the potential benefit of gaining a gene useful for survival, this circuit couples noise and feedback loops in a way such that only a subset of a population will enter the compentent state at a given time. As a growing culture reaches high density, quorum sensing plays a key role in driving noisy ComK protein expression that dictates whether a cell enters into a state of competence. If ComK levels with a single cell exceed a certain threshold, the cell transiently enters a competent state, as enforced by feedback loops, before reverting back to a noncompetent state. This circuit, using noisy levels of one protein coupled to feedback loops that create stability, is an example of signal network architecture can exploit noise to achieve many behaviors when viewed in a dynamical context.

Returning to the example of gene activation, different signal network architectures are more prone to noise than others [36]. Negative autoregulation, in addition to speeding up response time, has been shown to reduce noise in gene expression [37], while positive autoregulation has been shown to increase noise. Additionally, the common bacterial sensory motif of the two component relay is also an effective noise filter [38]. While different signal network architectures can minimize the effect of cellular noise, there is another entirely different source of variation that bacterial cells contend with: the variations arising from cell division.

1.7 Noise arising from cell division

When a bacterial cell divides, its volume is split in half and its internal contents are divided between the resulting daughter cells. Single molecule tracking has shown that these contents are divided according to binomial probability [39, 40]. As this binomial partitioning is not an exact 50-50 split of the contents, one daughter cell will inherit more of a given molecular species than the other. This difference in number of a molecular species inherited between daughter cells, or partition error, when quantified, is on the same scale as fluctuations in stochastic gene expression [41]. Additionally, the random nature of cell division timing adds an additional layer of noise [42]. Cells may employ mechanisms for ordered partitioning, such as volume exclusion, spindle formation, or pair formation, but the parameter demands for these strategies to significantly reduce partition error largely lie outside of the physiologic range, with the notable exception of chromosomal segregation, where it is rational to invest large amounts of cellular resources [43]. Noise arising from cellular division becomes a significant limitation when considering cellular processes that occur on timescales longer than a single cell cycle.

1.8 Outline: Examination of signal network architecture governing multigenerational processes in bacterial cells

These two sources of noise, the stochastic nature of biochemical reactions and the partition error from cell division, affect cell signaling in the case of signal networks responsible for driving multigenerational processes. Which network architectures are capable of resulting in signaling that is robust to these sources of noise, as well as the design parameters that define that robustness remains an open question. Simulation and theoretical analysis can be used to broadly explore possible network architectures and their behavior over many generations.

In the first part of this dissertation, I develop a theoretical and computational approach to explore some basic governing rules of how network architecture impacts the state of a cell signaling network over multiple generations (**Chapter 2**). By simulating toy models of signaling circuits and examining in which components related cells are more similar to each other than to any random cell in the population, I find that doubly stochastic processes, and the variance they generate, is essential to generating a signal that is maintained in related cells. I also investigate the kinetic boundaries of this behavior as well as how the duration of nongenetic similarity can be decreased by network architecture. I end with an analysis of how the duration of this nongenetic similarity connects to the ability to modulate noise levels amongst a population of cells.

While a computational and theoretical approach offers the ability to unearth possible rules governing multigenerational processes, an experimental study dissecting an existing cellular network that drives a multigenerational process can examine a case where evolution has actually implemented one possible incarnation of these design rules. In bacterial systems, multigenerational processes are commonly employed in cell fate decisions, in which cells opt between phenotypes or lifestyles to adapt to environmental conditions [44]. A model system to study such cell fate decisions is the aforementioned soil microbe *Bacillus subilits*. One of its possible cell fate decisions is sporulation, where, under extreme or starvation conditions, *B. subtilis* will form a dormant but durable spore that has the ability to germinate in more favorable conditions [45]. This sporulation decision has been shown to be made over multiple generations [46, 47]. In addition to sporulation, *B. subtilis* has other cellular processes that occur over multiple generations [48]. The uniting network architecture between these sporulation and competence, among other cell fates, is the master regulator Spo0A. An underexplored aspect of Spo0A signaling is its impact on cyclic di-GMP [49], a second messenger commonly implicated in biofilm formation [50].

In **Chapter 3**, I then investigate the dynamics of the cyclic di-GMP pathway in *B. subtilis* with a joint theory and experiment approach. I develop a theoretical model that shows how the kinetics of the enzymes regulating cyclic di-GMP concentration, along with the network architecture, enables accumulation of c-di-GMP to occur over multiple generations. I then proceed to develop an approach for experimentally testing this prediction, first by developing an assay for imaging biofilm phenotype activation in single *B. subtilis* cells using timelapse fluorescence microscopy and then by testing a series of genetically-encoded biosensors, including a newly-designed circularly permuted GFP sensor, for their sensitivity to c-di-GMP dynamics in *B. subtilis*. Combining these approaches enables the ability to measure c-di-GMP accumulation in single cells and to validate if the c-di-GMP circuit in *B. subtilis* is functioning as a part of a multigenerational signal network.

Finally, in **Chapter 4**, I reflect on future directions suggested by this work. On the theory and computation side, I consider elements that could move these models beyond toy circuits and incorporate sufficient complexity for predictive power. On the experimental work examining biofilm formation, I consider how understanding of individual single cell pathways could be combined to understand development on the scale of the whole biofilm. I then end on a note reflecting on how theoretical design of new circuits and their behavior could be experimentally implemented for advancing the design space of synthetic biological systems.

Chapter 2

A toy model describes the requirements for nongenetic similarity in dividing cells

2.1 Introduction

While cell identity is often thought of in terms of genetic identity, the actual behavior of a cell is driven by the myriad of RNAs, proteins, ions, and other molecules contained within a cell. From sensing and responding to environmental signals to metabolism, the processes of a cell are driven by these nongenetic components. Indeed, this idea of the variation of nongenetic components leading to different behaviors of isogenic cells has a long history [51]. Different expression levels of different proteins can lead to different behavior, as can different metabolic states. For instance, bacterial cells with certain expression levels of efflux pumps can survive an antibiotic challenge while the remaining cannot [52].

When a cell divides, great care is taken to ensure its genetic contents, the DNA composing its genome, are copied and divided equally between the resulting daughter cells. The remaining contents of the cell are also divided: the proteins, ions, RNAs, and lipids that make up the remainder of the cellular infrastructure [53]. As cell behavior stems from specific concentrations of these components, the division of said components between resulting daughter cells leads to similar concentrations being passed down a generation, and thus similar behavior [54]. This idea has been recently studied and quantified at a single cell level. One experimental study demonstrated that 70% of the motility behavior in *E. coli* daughter cells can be explained by nongenetic inheritance [55]. Another experimental study showed that, in an isogenic population of *E. coli*, sister cells had correlations of cell size and cycle time beyond correlations between unrelated cells in the population [56]. However, while the sister cells were correlated immediately post-division, the correlation decayed after a six or eight generations. Other studies have found similar timing for the phenotypic divergence of *E. coli* [57]. One explanation for this decay is the random nature in which nongenetic factors are partitioned between daughter cells. Similar ideas hold when examining multigenerational correlations of cycle time in both human and mouse cancer cell lines [58]. Other work has shown that there is an optimal level of nongenetic inheritance to aid in bacterial chemotaxis [59, 60].

The nongenetic factors of bacterial cells, in absence of subcellular organelles, have been shown to be partitioned according to binomial probability, as shown in studies tracing single molecules of RNA [39] and fluorescent proteins [40]. With binomial probability, there is some deviation away from a perfect halving of the contents into each daughter cell. The deviation from this perfect split is referred to as partition error. These partition errors can accumulate over a few generations to yield large heterogeneities in protein count from a single cell, akin to the variation explained by stochastic gene expression [41]. Other theoretical work has shown that overcoming partition error requires extreme parameters, unrealistic in a biological context [43]. Further studies have quantified the levels of noise of a single protein in a system with stochastic production and cell cycle timing [42].

To investigate over what timescales the accumulation of differences in nongenetic factors as a result of binomial partitioning occur, we turned to toy model simulations of signaling circuits. These simple models are a tractable system in which to understand the minimal requirements for nongenetic similarity in related cells. We found that not a single protein, but a simple pathway, in which a protein produces a signal molecule, is sufficient for inherited nongenetic similarity. We then found that adding an additional component, a degrading enzyme, was sufficient to alter the duration of this nongenetic similarity. These two or three component, one or two interaction systems, provide a base of understanding of how inheritance of nongenetic factors can generate similarity in related cells that persists for multiple generations.

2.2 Results

2.2.1 Construction of a stochastic model to simulate a minimal signaling circuit

The simplest signaling circuit one can create to explore the minimal requirements for multigenerational similarity is a “Production Only” circuit, composed of an enzyme, M, which makes a signal molecule, A, at a constant rate, $k_{cat,M}$ (Fig 2·1A). In this architecture, the production probability of the signal molecule A is described using Equation 2.1:

$$P_{prod,A} = k_{cat,M} * M \tag{2.1}$$

where M is the number of M in a given cell. The enzyme M is produced with a fixed probability, $P_{prod,M}$, where one M is produced every $1/P_{prod,M}$ time units on average [61, 62]. Note that this functions as the definition of a Poisson random variable. The intracellular events, specifically production of the enzyme M or signal molecule A, were simulated using a Gillespie algorithm [63]. This was chosen in order to examine

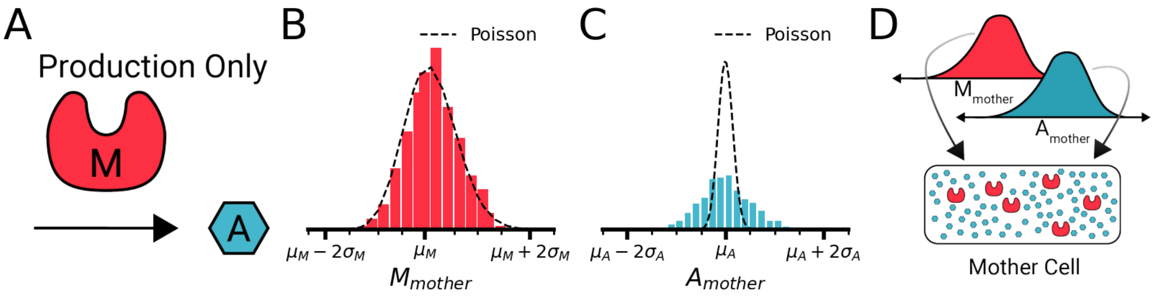

Figure 2·1: Minimal signaling circuit components result in different underlying distributions — **(A)** The Production Only circuit architecture used in this study. An enzyme M, produced at rate $P_{prod,M}$, reacts to form a signal molecule A at a fixed rate $k_{cat,M}$. Decreases in counts of enzyme M and signal molecule A in each cell come solely from dilution due to cell division. **(B)** Distribution of counts of the M enzyme in cells immediately prior to division. The model was run for 1000 cell cycles and the counts of M enzyme prior to each of the 1000 division events is shown here (red histogram). The distribution fits a Poisson model based on the mean M enzyme count (dotted black line). **(C)** Distribution of counts of the signal molecule A in cells immediately prior to division. The same model run of 1000 cell cycles was used and the counts of signal molecule A immediately prior to division are plotted here (blue histogram). A Poisson model based on the mean value was attempted to fit the distribution, but the variance was much wider than a Poisson (black dotted line). **(D)** Each instance of a mother cell in this work can be considered as a cell with counts of enzyme M and signal molecule A independently pulled from these distributions.

the performance of a circuit stochastically in a manner that is robust at all levels of enzyme counts. Following the algorithm, we first calculated the sum of probabilities for this Production Only circuit, as shown in Equation 2.2:

$$R_{total} = P_{prod,A} + P_{prod,M} \tag{2.2}$$

The length of the time step between events is then given by Equation 2.3:

$$\tau = (1/R_{total}) * \ln 1/r_1 \tag{2.3}$$

where r_1 is a random number pulled from a uniform random distribution spanning 0 to 1.

To simulate cell growth we preset a fixed cell cycle time, T_{cc}. At each event time determined from the Gillespie simulation, we then calculated the new cell volume using Equation 2.4:

$$V_{new} = V_{old} + \frac{\tau}{T_{cc}} \tag{2.4}$$

Cell division is triggered when $V > 2$, upon which the intracellular contents are divided up between daughter cells using a binomial random variable and the volume is reset to 1.

For the initial run of the model, we took inspiration for our parameters from the c-di-GMP circuit in *B. subtilis*. The typical k_{cat} of the c-di-GMP forming DGC enzymes is 1 min^{-1} [64] and their copy numbers are in the tens per cell, or in the 10 nM range [65]. Given our time units approximate seconds, we set $k_{cat,M} = 10^{-2}\ s^{-1}$ and $P_{prod,M} = 10^{-2}\ s^{-1}$. Lastly, given *E. coli*'s doubling time of 20 minutes in rich conditions [66], we set our initial cycle time to 1000 time units.

Running the model for n cell generations produces a series of n pre-division conditions, all of which can be seen as possible Mother Cell states (Fig A·1A, enzyme concentration trace; B, signal molecule concentration trace). To generate a large pool fully sampling the possible pre-division parameter space, we ran the model for 1000 generations. This initial condition set of the number of enzyme M molecules at all timepoints immediately prior to division falls on a distribution, indicating the effects of stochastic gene production and binomial partitioning (Fig 2·1B, red bars). Given the simulation for M follows a stochastic process with a fixed rate, we analytically derived the mean as a function of the model variables (see A2, Part 3) and found a Poisson distribution with this mean matches the simulated distribution (Fig 2·1B, black dotted line).

Given that possible values of M follow a Poisson distribution, we wanted to explore what this means for the distribution of possible signal molecule A values. As the signal molecule is produced stochastically with a rate proportional to the stochastic production of the enzyme M, the production of A falls under a category of doubly stochastic processes known as Cox Processes [67]. A known property of these doubly stochastic systems is that the variance of the doubly stochastic variable is greater than that of the singly stochastic variable. To confirm if this was true for the production of signal molecule A, we compared the distribution of the counts of A (Fig 2·1C, blue bars) to a Poisson distribution with the a mean derived from our analytical solutions (see A2, Part 5) (Fig 2·1C, dotted black line). As expected from the Cox Process, the variance of counts of A was wider than a Poisson.

We can now think of any Mother Cell in this model as having its counts of enzyme M or signal molecule A as realizations of these distributions (Fig 2·1D), allowing us to explore how this impacts nongenetic similarity in the system.

2.2.2 Nongenetic similarity is present only in signaling molecule concentration

This simple model not only allows us to explore how a single cell lineage varies over time, but also how similar related cells are to other cells in the population. To quantify the presence of nongenetic similarity, we constructed a metric inspired by experimental findings that properties such as cell cycle time and cell size appear to be more similar for multiple generations of a lineage of cells than would be expected between two random cells in a population [56]. Specifically, we simulated divisions from 1000 different mother cells and calculated pairwise differences between the resulting sister and random cells (Fig 2·2A, top). In each case we then also calculate the variance of the distribution of these differences (Fig 2·2A, bottom). If the variance of the differences between sister cells is less than the variance of the differences between random cells, the sister cells are more similar to each other than to a random cell in the population, and thus nongenetic similarity exists in the system. We then ran the model with the resulting sister cells and random pairs for 10 generations, as this was the time period in which it was found experimentally all nongenetic similarity would decay [56]. We performed this variance of difference calculation at each time point for both the concentration of enzyme M and signal molecule A.

Looking at the variance of the differences in the enzyme M, we found the sister pairs and random pair variances well-tracked one another (Fig 2·2B). Each division doubles the variance value, which then decays downwards until the following division raises the variance between the cells again, creating a sawtooth-like pattern in the variance trace (Fig 2·1B). This sawtooth comes the fact the variance of number of signal molecule doubles while the variance of the concentration is proportional to the variance of the number of signal molecules over the volume squared. As the volume

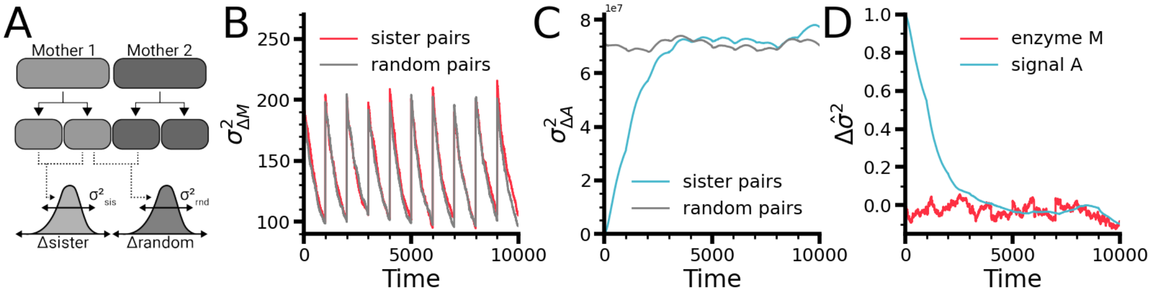

Figure 2·2: Nongenetic similarity is present only in signaling molecule concentration — (A) Schematic of nongenetic similarity metric. To see if nongenetic similarity is present in the system, variance of the distribution of pairwise differences between sister cells ($\Delta sister$, left) were compared to the variance of the distribution of pairwise differences between unrelated cells ($\Delta random$, right). Nongenetic similarity is present when $\sigma^2_{sis} < \sigma^2_{rnd}$. **(B)** Variance of distributions of pairwise differences in counts of enzyme M. 1000 pairs of sister cells and random cells were simulated for 10 generations and the variances of their distributions of pairwise differences in counts of M were calculated at each time point. The values were identical between the sister (red line) and random pairs (gray line) over the ten simulated cell cycles. **(C)** Variance of the distributions of pairwise differences in counts of the signal molecule A. 1000 pairs of sister cells and random cells were simulated for 10 generations and the variances of their distributions of pairwise differences in counts of A were calculated at each time point. The random pairs had a constant variance over the ten simulated generations (gray line) while the sister pairs started at a low value, which then increased until saturating at the value of the variance of the random pairs (blue line). **(D)** Normalized difference of variances plot for enzyme M and signal molecule A. To compare nongenetic similarity across values, the difference between the variance of the random pairs and sister pairs was divided by the difference between random pairs (see Equations 2.5 and 2.6). While the value for the enzyme M hovered near 0 (red line), the value for the signal molecule A started high before decreasing down to the 0 (blue line). Nongenetic similarity is present anytime this normalized difference value is greater than 0. n = 1000 pairs of cells for each calculation.

doubles, the denominator of the variance of the concentration quadruples, creating the decay by one half each generation pattern in both sisiter and random pairs. This matching between the sister and random pairs suggests that there is no point in which the variance of the sister cell differences is less than that of the random pairs, which means the sister cells are just as dissimilar to each other as to any random cell in the population. This lack of similarity between sister cell pairs suggests that no nongenetic similarity exists with respect to the concentration of the enzyme M.

This changes when we examine the variance of the pairwise differences in the signal molecule A (Fig 2·2C). The variance of A of the random pairs stays at a consistent and high level (Fig 2·2C, gray line). This suggests a consistent level of variance in the system, consistent with the variance of the mother cell distribution (Fig 2·1C). Meanwhile, the sister cell pairs start at a low level of variance, which then steadily increases over time until it levels out at the random pair variance (Fig 2·2C, cyan line). The low level of variance in A in the sister cells means that the sister cells are more similar to each other than other random cells in the population. This suggests that nongenetic similarity exists in the concentration of the signal molecule A. To confirm this is not due to our choice of partitioning model, we additionally tested two other partition models: (1) a "correlated binomial" partition model, in which the same binomial random variable is used to partition both the enzyme M and the signal molecule A, which mimics the clustered volume partition model described in [43] (see Fig A·3, second column) and (2) a "perfect partition" model, in which the subcellular contents were divided perfectly in half upon each division (Fig A·1, third column). In both alternate models, we also found nongenetic similarlity to be present in the concentration of the signaling molecule A (Fig A·3C).

To better visualize and compare this nongenetic similarity, we defined a normalized metric using the variances of the sister and random pairs by taking the difference in variances and dividing it by the variance of the random pairs (defined for the enzyme M in Equation 2.5 and the signal molecule A in Equation 2.6):

$$\Delta\hat{\sigma}^2_{\Delta[M]}(t) = \frac{\sigma^2_{\Delta[M],rnd}(t) - \sigma^2_{\Delta[M],sis}(t)}{\sigma^2_{\Delta[M],rnd}(t)} \tag{2.5}$$

$$\Delta\hat{\sigma}^2_{\Delta[A]}(t) = \frac{\sigma^2_{\Delta[A],rnd}(t) - \sigma^2_{\Delta[A],sis}(t)}{\sigma^2_{\Delta[A],rnd}(t)} \tag{2.6}$$

Using this metric to consider the multigenerational behavior of M and A, we see the normalized difference of variances value for the enzyme M hovers around 0 (Fig 2·2D, red line) while the value for the signal molecule starts near 1 before decreasing down to 0 (Fig 2·2D, cyan line). Any time this metric is greater than zero is when the sister cells are more similar to each other than the random pairs, suggesting the presence of nongenetic similarity. To ensure this was not simply a product of a single run of the simulation, we ran the simulation sequentially ten times from the same random seed, achieving similar results each time (Fig A·2). Furthermore, as experimental findings have demonstrated heterogenity in bacterial growth rates (i.e. *Salmonella enterica* [68]), we also explored how the uniform cell cycle time influences nongenetic similarity by simulating cells whose cycle times changed each generation, pulled from a distribution around the set T_{cc} value. We found this did not impact the presence of nongenetic similarity in the signal molecule concentration nor its duration (Fig A·4).

2.2.3 Nongenetic similarity magnitude is sensitive to reaction rate and tied to cell cycle length

Our minimal model has three tunable parameters: the enzyme reaction rate ($k_{cat,M}$), the enzyme production rate ($P_{prod,M}$) and the cell cycle time (T_{cc}). To understand which of and how these parameters affect the presence/duration of nongenetic similarity, we bounded the parameter ranges with physiologically-relevant values and varied them across orders of magnitude.

For the k_{cat} values, a study of *E. coli* enzyme max turnover rates *in vivo* found a range of $10^{-3}s^{-1}$ - $10^{3}s^{-1}$ [69]. In varying the value of the enzyme reaction rate ($k_{cat,M}$) over these six orders of magnitude, we found nongenetic similarity present at all values except the lowest value tested ($k_{cat,M} = 10^{-3}s^{-1}$, Fig 2·3A, darkest green line). Above this value, the magnitude did not change the duration of the nongenetic similarity.

For the enzyme production rate ($P_{prod,M}$), one large-scale measurement of protein copy number in *E. coli* found copy numbers ranging from 1 to 300,000 [70]. To translate this into a production rate, we analytically solved our the mean counts of M as a function of our model parameters (see Supplemental Note 1) and found we could relate the equilibrium copy number of enzyme M ($[M]_{eq}$) to the $P_{prod,M}$ value using Equation 2.7:

$$[M]_{eq} = P_{prod,M} * T_{cc} \tag{2.7}$$

This results in a range for $P_{prod,M}$ of 10^{-3} - 10^{3} s^{-1}. However, given that we saw no nongenetic similarity present at the lower end of the $k_{cat,M}$ range, we used the first magnitude value of $k_{cat,M}$ where nongenetic similarity was present for this sweep. As both $P_{prod,M}$ and $k_{cat,M}$ increase, so too does the equilibrium concentration of signal molecule A. To identify an upper bound for $P_{prod,M}$, an additional study quantified the concentration of all secondary metabolites in *E. coli* and found the largest value in 100 μM range [71]. As we had done with the mean values of enzyme M, we analytically-derived an expression for the mean concentration of signal molecule A as a function of the model parameters (see A2, Part 5) using Equation 2.8:

$$[A]_{eq} = 3/2 * k_{cat,M} * P_{prod,M} * T_{cc}^2 \tag{2.8}$$

Therefore, we set an upper bound of $P_{prod,M}$ to values that would result in an $[A]_{eq}$ value of 100 μM or less. This left us with a range of $P_{prod,M} = 10^{-3} - 10^{2} s^{-1}$. Across all five orders of magnitude, the enzyme production rate did not appear to have an effect on the presence of nongenetic similarity nor its duration (Fig 2·3B).

For the cell cycle time, we again looked to the literature for lower and upper bounds. For the lower bound, we took inspiration from *Pseudomonas natriegens*, the fastest-recorded growing bacteria, with a cell cycle time under 10 minutes [72], and set the lower bound of T_{cc} to 500s. For the upper bound, we were inspired by the 2-6 hour doubling time of *Sinorhizobium meliloti* [73], and set the maximum screened value of T_{cc} to 10,000s. Varying the cell cycle time over this range we observed a change in the nongenetic similarity duration (Fig 2·3C), with longer cell cycle times leading to longer nongenetic similarity duration. To understand how absolute time of nongenetic similarity duration was related to the number of cell cycles the similarity persisted over, we also calculated the duration in cell cycles for each T_{cc} value (Fig 2·3C′). Interestingly, the normalized differences of variance curves collapsed on one another when calculated for cell cycles, suggesting that number of cell cycles, not absolute amount of time, is the critical time parameter in the system. We additionally found this duration constant across the other two partition models (Fig A·3D).

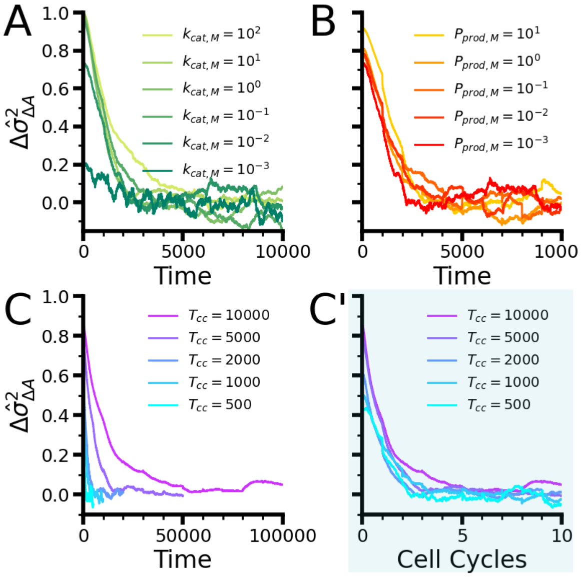

Figure 2·3: Nongenetic similarity extent changes with reaction rate and duration changes with cell cycle time — **(A)** Normalized difference of variances plot for enzyme reaction rate ($k_{cat,M}$) parameter sweep. At the lowest tested value of $k_{cat,M}$, the difference of variance line stayed near 0, suggesting little nongenetic similarity present. All tested values above $k_{cat,M} = 10^{-3}$, spanning five additional orders of magnitude, had high values of variance difference which then decayed similarly, suggesting consistent presence and duration of nongenetic memory above this threshold. **(B)** Normalized difference of variances plot for enzyme production rate ($P_{prod,M}$) parameter sweep. The presence/absence and duration of nongenetic memory was entirely unchanged over five orders of magnitude tested. **(C)** Normalized difference of variances for different cell cycle times. Memory duration was longer with longer cell cycle times. To see if it was a function of number of cell cycles, the data were replotted in **(C′)**. Here, the curves collapsed to the same duration, suggesting that it is number of cell cycles, or generational time, that sets the duration of nongenetic memory, rather than the absolute number of time units. n=1000 pairs of sister and random cells for each difference of variance calculation.

2.2.4 Nongenetic similarity in signal molecule concentration requires signal amplification

To understand why the reaction rate ($k_{cat,M}$) parameter sets a bound on the presence of nongenetic similarity in the system, we first needed to define the presence of nongenetic similarity mathematically. Using our normalized difference of variance metric (Equations 2.5 and 2.6), we defined nongenetic similarity as existing when the value of the normalized pairwise difference was greater than 0. In order to understand how our model parameters shape the boundaries of nongenetic similarity, we first needed to solve the equations for the variances of these comparison distributions, for both the enzyme M and the signal molecule A, and for both the sister and random pairings of cells. We did this using an analytical derivation (see Appendix A) and found the following for the variance of differences in the enzyme M:

$$\sigma^2_{\Delta M,sister} = 2P_{prod,M}T_{cc} \tag{2.9}$$

$$\sigma^2_{\Delta M,random} = 2P_{prod,M}T_{cc} \tag{2.10}$$

Comparing our derivations to variance values calculated from numerical simulations shows good agreement across orders of magnitude of our parameter values (Fig 2·4A, Fig A·6). In examining these derivations, we immediately noticed that the variance equations are the same for the sister and the random cells (Equations 2.9 and 2.10). This indicates that, over all possible parameter combinations for the Production Only circuit, the variances will be equal and our normalized difference of variance metric will always be 0 (Fig 2·4B). This calculation further validates our earlier finding of identical variances of differences in concentration of M over the simulated time in Fig 2·2B.

Moving over to the signal molecule A, a different story plays out. Our analytical derivation (see A2, Part 10) found the following equations describing the variance of the pairwise differences for signal molecule A:

$$\sigma^2_{\Delta A,sister} = 3k_{cat,M}P_{prod,M}T^2_{cc} \tag{2.11}$$

$$\sigma^2_{\Delta A,random} = 3k_{cat,M}P_{prod,M}T^2_{cc} + \frac{20}{9}k^2_{cat,M}P_{prod,M}T^3_{cc} \tag{2.12}$$

The scaling of these equations is such that, at low values of $k_{cat,M}$ or T_{cc}, the variances are similar, suggesting a decrease of nongenetic similarity in the system (Fig 2·4C, lower values), but at increasing values of $k_{cat,M}$ or T_{cc}, the variances begin to diverge, with the variance of random pairs increasing more rapidly than the variance of sister pairs (Fig 2·4C, larger values). We can better visualize this by using the analytical equations to calculate the normalized difference of variance value at each parameter combination of $k_{cat,M}$ and T_{cc} (Fig 2·4D). Here we see a region of parameter space where little to no nongenetic similarity exists (Fig 2·4D, red triangle in the lower left corner) and a region where nongenetic similarity does exist (Fig 2·4D, blue region). The boundary of white values, indicating when $\Delta\hat{\sigma}^2_{\Delta A} = 0.5$ is marked by when $k_{cat,M}T_{cc} = 3/2$. This boundary means that, as long as there is on average more than 3/2 of a reaction per enzyme occurs per cell cycle, nongenetic similarity will be present in the system. This mathematically underlies the idea of signal amplification, that each step of a signaling pathway results in amplification of a signal. We see this here in that, as long as each enzyme produces more than one signal molecule A per cell cycle, the signal of the enzyme concentration is amplified in the signal molecule. We found this result particularly interesting, as it gave a mathematical explanation behind the

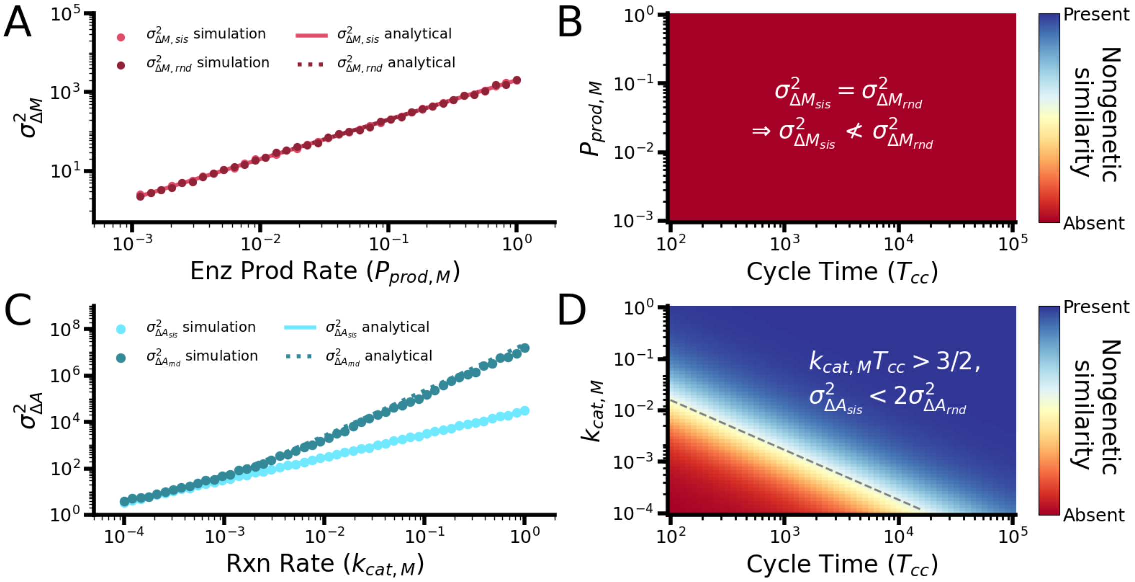

Figure 2·4: Nongenetic similarity in signal molecule requires amplification — **(A)** Comparison of analytically-derived solutions (solid, light red line, sister pairs; dotted, dark red line, random pairs) with numerically-calculated values (light red dots, sister pairs; dark red dots, random pairs) for variance of difference of cell pairs in concentrations of enzyme M. The enzyme production rate parameter ($P_{prod,M}$) was swept over three orders of magnitude and, at each value, 1000 pairs of cells were simulated. Other parameters were held constant over this sweep ($k_{cat,M} = 10^{-2}$, $T_{cc} = 1000$). **(B)** Nongenetic similarity in concentration of enzyme M over two-dimensional parameter sweep. As the variance of differences for sister and random pairs are equal, the variance of the difference between sister pairs will never be less than the variance of difference between random pairs, therefore nongenetic similarity will never exist in the concentration of the enzyme M. **(C)** Comparison of analytically-derived solutions (solid, light blue line, sister pairs; dotted, dark blue line, random pairs) with numerically-calculated values for variance of difference of cell pairs (light blue dots, sister pairs; dark blue dots, random pairs) in concentrations of enzyme A. The reaction rate parameter ($k_{cat,M}$) was swept over five orders of magnitude. At each value, 1000 pairs of sister and random cells were simulated and the variance of their difference in concentration of A was calculated. The difference in variances is negligible at small values of $k_{cat,M}$ and diverges at larger values. Other parameters were held constant over this sweep ($P_{prod,M} = 10^{-2}$, $T_{cc} = 1000$). **(D)** Nongenetic similarity in concentration of signal molecule A over two-dimensional parameter sweep. For each parameter combination of $K_{cat,M}$ and T_{cc}, the analytically-derived formulas were used to calculated the $\Delta\hat{\sigma}^2_{[A]}$ value. The values are plotted here where absent is a $\Delta\hat{\sigma}^2_{[A]}$ of 0 and present is a $\Delta\hat{\sigma}^2_{[A]}$ of 1. A transition between presence and absence occurs where $k_{cat,M}T_{cc} = 3/2$, indicated by the dashed grey line (where $\Delta\hat{\sigma}^2_{[A]} = 1/2$).

usefulness of signal amplification. Signal amplification is often noted as a key feature of signal pathways, such as in the MAPK mating response pathway in *Saccharomyces cerevisiae*, where the final transcriptional factor, Fus3, has a copy number of 20,000, compared to the first protein in the cascade, Ste11, with a copy number of 4000 [74]. With an understanding of the basis behind the amplification requirement, we turn our attention to what sets the duration of the nongenetic similarity.

2.2.5 Cell divisions bound the duration of nongenetic similarity

Given that signaling molecule amplification is important for the presence of nongenetic similarity, it suggests that the level of signaling molecule may be important for tuning the timescale of this similarity. However, our parameter sweep over different cell cycle times (Fig 2·3C′), which correspond to different concentrations of signal molecule A (as set by Equation 2.8), have similar durations of nongenetic similarity when examined in units of cell cycles. This suggests a relationship between the duration of nongenetic similarity and the number of cell cycles. To explore this idea, we developed a conceptual and mathematical hypothesis for the nongenetic similarity duration based on dilution over cell divisions and compare this model to our simulated data from the cell cycle time sweep.

The conceptual hypothesis (diagrammed in Fig 2·5A) is as follows: as any given mother cell in our simulation has counts of A pulled from the A_{mother} distribution (as seen in Fig 2·1C), any given mother cell's counts of A represent some offset from the mean of this distribution (μ_A). When a mother cell divides, its resulting daughter cells inherit this offset from the mother (Fig 2·5A, gen 1). However, over the subsequent cell cycle (gen 1), the two daughter cells produce their own new signal molecule A (whose expected value corresponds to half of the mean of the mother cell distribution), effectively diluting the offset from the mother by one half. This

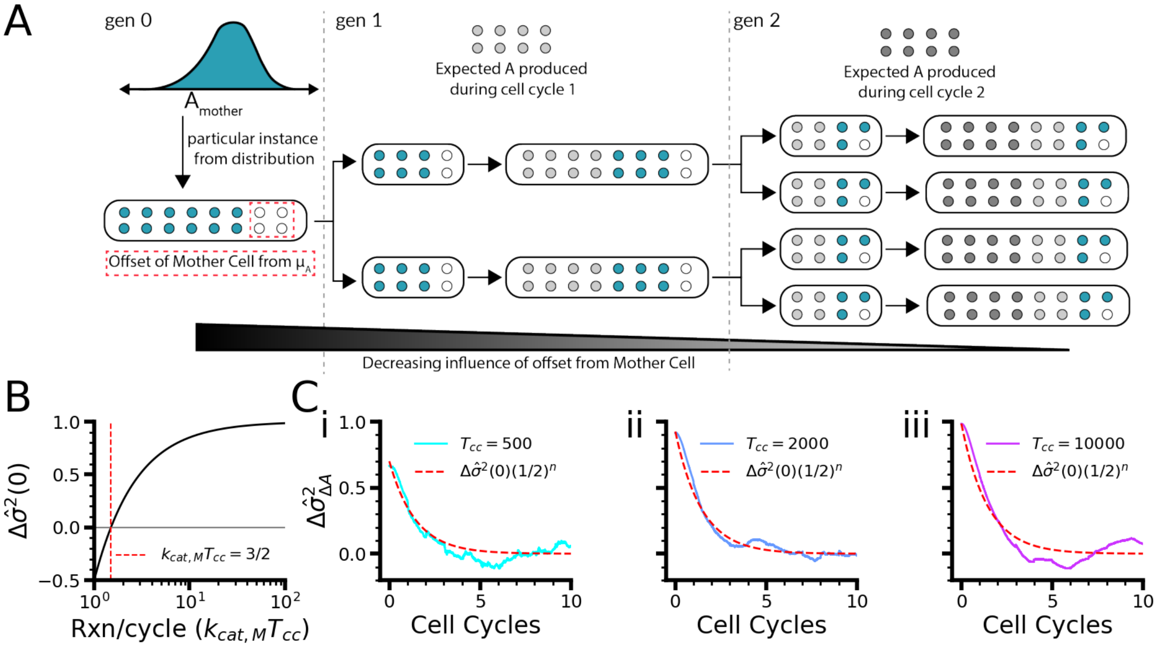

Figure 2·5: Nongenetic similarity decay is geometric along cell generational time — (A) Cartoon of conceptual model for nongenetic memory duration. Any given mother cell has counts of the signal molecule A pulled from the A_{mother} distribution (Fig 2·1C). Any particular instance will have some offset from the mean count of A (open circles in red dashed box). This offset is then passed to the next generation. During growth, the daughter cells will, on average, accumulate the expected amount of A over the first generation (light grey circles). This dilutes the relative influence of the offset from the mother (now two open circles instead of four). This generation then undergoes its own division, producing daughter cells that then go on to accumulate the expected counts of A (dark grey circles). This further dilutes the original offset from the original mother cell again by half. Under this model, it is hypothesized the memory will decay by half each generation. **(B)** Initial value for the normalized difference of variances metric as a function of reactions/cell cycle ($k_{cat,M}T_{cc}$). This relationship was solved using the derived formulas for variances of the pairwise differences (Equation 2.14) and gives the initial value for the memory decay curves as a function of the reaction rate ($k_{cat,M}$) and cell cycle time (T_{cc}). **(C)** Normalized difference of variance curves for different cell cycle times along with the conceptual model plotted on top. 1000 pairs of sister and random cells were run for each cell cycle time and the normalized difference of variances curve was plotted for $T_{cc} = 500$ **(i)**, $T_{cc} = 1000$ **(ii)**, and $T_{cc} = 2000$ **(iii)**. The conceptual model describing memory decay was then plotted on top (Equation 2.16, red dashed lines). Agreement between model and simulated data indicate the memory does follow a geometric decay as a function of the number of elapsed cell cycles, with the model parameters only adjusting the initial value.

is repeated in the following generation (Fig 2·5A, gen 2), in which the cells again divide and again, produce their own signal molecule A over the following cell cycle, again diluting the original offset from the Mother Cell by one half. This suggests the nongenetic similarity decays follows a geometric function as described by Equation 2.13:

$$\Delta\hat{\sigma}^2_{\Delta A}(n) = (1/2)^n \tag{2.13}$$

where n is number of elapsed cell cycles. While this hypothesis describes the dynamics of the decay, it is missing a description of the magnitude of that original, inherited offset. This offset can be described by the value of the normalized difference of variances immediately following division from the mother cell (Equation 2.14).

$$\Delta\hat{\sigma}^2_{\Delta A}(n=0) = 1 - \frac{\sigma^2_{\Delta A,sis}}{\sigma^2_{\Delta A,rnd}} \tag{2.14}$$

Thus, we use our derived equations for the variance formulas to obtain a function describing the initial value of the normalized difference of variances as a function of the model parameters. Plugging in Equations 2.13 and 2.14, we get Equation 2.15:

$$\Delta\hat{\sigma}^2_{\Delta A}(n=0) = 1 - \frac{3}{2k_{cat,M}T_{cc}} \tag{2.15}$$

Here, we see that the initial normalized difference of variances is a function of the enzyme reaction rate and the cell cycle time (Fig 2·5B). It follows that the initial value would change as we sweep through different values of the cell cycle time. We see that the function increases until saturating at a value of 1. This saturation makes sense, given that at high values of $k_{cat,M}T_{cc}$, the concentration of A is so high that the relative impact from the initial binomial partition approaches zero. It also makes sense that this value becomes positive at the critical value ($k_{cat,M}T_{cc} = 3/2$), as that was the minimal requirement for nongenetic similarity we saw earlier (Fig 2·4D). With the original offset quantified, we then derived an equation that combines a geometric decay function and the original offset using Equation 2.16:

$$\Delta\hat{\sigma}^2_{\Delta A}(n) = (1 - \frac{3}{2k_{cat,M}T_{cc}})(1/2)^n \tag{2.16}$$

We could then compare our mathematical hypothesis to the simulated model data with respect to duration of nongenetic similarity (Fig 2·5C i, ii, and iii). There is good agreement between the simulated model data (colored solid lines) and our mathematical hypothesis (red dashed lines), suggesting that all of the nongenetic similarity decays by one half each generation with an initial offset based on the reaction rate ($k_{cat,M}$) and the cell cycle time (T_{cc}) (model curves for all cell cycle sweep data plotted in Figure A·8). Interestingly, as the duration is set by cell division cycles, there is nothing in the model parameters that suggests the ability to extend duration of this type of nongenetic similarity within this Production Only circuit architecture.

2.2.6 Degradation enzyme activity decreases duration of nongenetic similarity

To explore more complex signaling circuit architectures and how they may tune nongenetic similarity, we also considered a "Production and Degradation" model in which we added an enzyme, B, that breaks the molecule A (Fig 2·6A). The degradation of A by the enzyme B follows Michaelis-Menten Kinetics [75], as described by Equation 2.17:

$$P_{break,A} = \frac{k_{cat,B}[B][A]}{K_{M,B} + [A]} \tag{2.17}$$

in which [B] is the concentration of the breaker enzyme B, $k_{cat,B}$ is its maximum reaction rate, and $K_{M,B}$ is its half-saturation constant. We wanted B to operate in its linear range, so we set $K_{M,B} = [A]_{eq}$ calculated using Equation 2.18:

$$[A]_{eq} = K_{M,B} = k_{cat,M} P_{prod,M} T_{cc}^2 - \frac{k_{cat,B}}{2} P_{prod,B} T_{cc}^2 \tag{2.18}$$

Equation 2.18 also reveals the requirements to ensure B does not degrade all of A. In order to have $[A]_{eq} > 0$, Equation 2.19 must hold:

$$k_{cat,M} P_{prod,M} > \frac{k_{cat,B}}{2} P_{prod,B} \tag{2.19}$$

We can define $V_{max} = k_{cat} P_{prod}$ for each enzyme and simplify this to Equation 2.20:

$$V_{max,M} > \frac{V_{max,B}}{2} \tag{2.20}$$

An additional way to formulate this boundary is by defining a maximum on the ratio of V_{max} values, as in Equation 2.21:

$$\frac{V_{max,B}}{V_{max,M}} < 2 \tag{2.21}$$

With that bound set, we can then define the production of the breaker enzyme B. Similar to the maker enzyme M, B is produced stochastically with probability, $P_{prod,B}$. The rate total equation for the Production and Degradation model is shown by Equation 2.22:

$$R_{total} = P_{form,A} + P_{break,A} + P_{prod,M} + P_{prod,B} \tag{2.22}$$

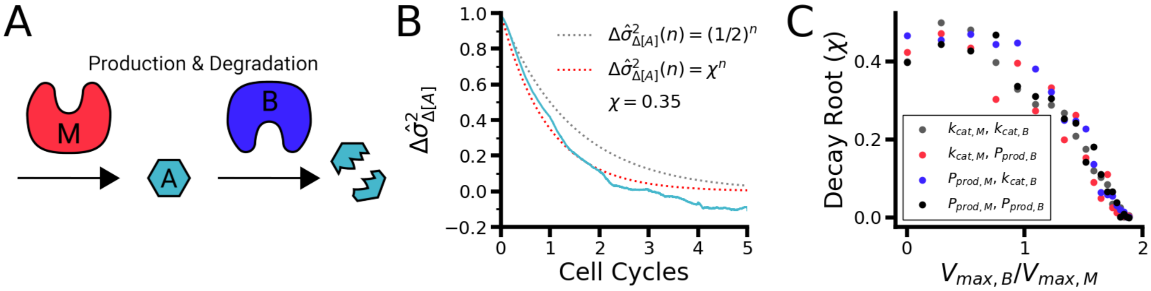

Figure 2·6: Ratio of formation and degradation enzyme activity tunes duration of nongenetic similarity — (A) Schematic of the Production and Degradation circuit. Here, we add a degrading enzyme that breaks down the signal molecule A according to Michaelis-Menten kinetics. **(B)** Quantifying decay of nongenetic similarity in Production and Degradation circuit. Decay of nongenetic similarity in Production and Degradation curve (cyan line) is faster than the $y = (1/2)^n$ curve from the decay timing of the Production Only circuit (dotted grey line). Instead, a new curve is fit following the form of $y = \chi^n$. For this parameter combination, the decay root value $\chi = 0.35$ (dotted red line). **(C)** Decay root quantification over V_{max} ratio sweeps that hold signal molecule concentration constant. To analyze the memory duration of signal molecule A without changing its equilibrium concentration, pairs of kinetic parameters for the maker enzyme M and the breaker enzyme B were changed in tandem (see legend). Regardless of which parameters were altered, increasing the V_{max} ratio decreased the decay root (χ) and thus, the nongenetic similarity duration. n=1000 pairs of cells for each parameter combination.

Intuition suggests that a degradation enzyme would more quickly dilute the inherited offset in the original mother cell (Fig 2·5A), decreasing the duration of nongenetic similarity. Indeed, in our initial run of this model where we selected parameters in the middle of our physiologic range estimates, setting $k_{cat,M}$, $P_{prod,M}$, $k_{cat,B}$, and $P_{prod,B} = 10^{-1}$, we saw that the nongenetic similarity decayed faster in the Production and Degradation circuit (Fig 2·6B, cyan line) compared to the the $y = (1/2)^n$ curve from the Production Only circuit (Fig 2·6B, gray dotted line). To quantify how much faster, we fit a $y = \chi^n$ curve to the data (Fig 2·6B, red dotted line) and backed out the value of the base constant, χ. For this parameter set, the fitted value was $\chi = 0.35$, demonstrating that the addition of a degrading enzyme leads to faster decay of nongenetic similarity.

To see how the kinetic parameters of the degrading enzyme B affected this decay rate, we performed this analysis over a broad parameter range. To ensure the faster decay rate was due not to lower concentration of the signal molecule, but to faster turnover of A, we swept our parameters but maintained a constant level of signal molecule (a sweep where only the parameters for the breaker enzyme B were altered was also run, Fig A·9). To do this, as we increased the value for $k_{cat,M}$ or $P_{prod,M}$, we also increased the value of $k_{cat,B}$ or $P_{prod,B}$ such that the equilibrium concentration of signal molecule was constant at 5 μM (Fig A·10). We found that, over 1000 sister cell simulations, the decay root (χ) of the normalized difference of variances of the signal molecule A ($\Delta\hat{\sigma}^2_{\Delta[A]}$) decreased as the V_{max} ratio increased (Fig 2·6C). This indicates that, as the activity of the degrading enzyme B increases, the duration of similarity between sister cells decreases until it essentially eliminates all duration of nongenetic similarity.

2.2.7 Inverse relationship between nongenetic similarity duration and population noise can be overcome with low enzyme noise

So far, our analysis has been restricted to comparisons between pairs of cells. In the Production Only circuit, we demonstrated nongenetic similarity between sister cells that decays with each generation. We then demonstrated this duration can be decreased with increasing activity of a degrading enzyme. With this in mind, we wanted to explore the consequences of this difference of nongenetic similarity duration at the population level (Fig 2·7A).

To evaluate the consequences of nongenetic similarity duration on cellular populations, we considered the coefficient of variation (CV) of the concentration of signal molecule A, which has been used as a measure of population heterogeneity or noise in other studies [26, 76, 77]. We find low values of $CV_{[A]}$ at long durations of nongenetic similarity, regardless of the parameter pairing (Fig 2·7B, large values of χ). This result is in concurrence with our previous finding (Fig 2·6C), where the choice of parameter pairing did not change the inverse relationship between the nongenetic similarity duration and the V_{max} ratio. However, at shorter durations of nongenetic similarity, we find different relationships between $CV_{[A]}$ and the nongenetic similarity duration (Fig 2·7B, smaller values of χ). When changing the k_{cat} values, we find shorter duration of nongenetic similarity corresponds to a more heterogeneous, or noisier, population (Fig 2·7B, grey dots). This inverse relationship between population noise and nongenetic similarity duration is seen to a lesser degree when only one of the parameters being changed is a $k_{cat,}$ value (Fig 2·7B, red and blue dots). Finally, we see a decrease in $CV_{[A]}$ values when both parameters being tuned at P_{prod} values (Fig 2·7B, black dots). This divergence can be explained by examining the underlying CV values of the enzymes themselves (Fig A·11). At high P_{prod} values, where the enzyme counts are higher, the CV values decrease (as has been seen elsewhere [41]). These less noisy enzyme concentrations then "trickle up" to less noise across the concentrations of A (Fig 2·7B, black dots). In short, the general consequence of shorter nongenetic similarity duration at the population level is increased heterogeneity, and this cell-cell heterogeneity can only be reduced through lower noise in enzyme concentration levels.

2.3 Discussion

With each additional component to the toy models of signaling pathways in this work, we found new properties and features of the behavior. Simple stochastic production and binomial partition of one molecule, represented by our enzyme M, is sufficient to create a heterogeneity of concentration values across a population (Fig 2·1B). It is insufficient, however, to generate nongenetic similarity (Fig 2·2B,D). But if that one molecule produces another, even stochastically at a constant rate, as with our signal molecule A in the Production Only circuit (Fig 2·1A), nongenetic similarity can exist in the system (Fig 2·2C,D) under conditions of signal amplification (Fig 2·3A, 2·4) and this similarity lasts for a fixed time, set by cell divisions (Fig 2·5). Lastly, the addition of a degrading enzyme, as in our Production and Degradation circuit (Fig 2·6A), is sufficient to tune down the duration of this similarity to the point of nonexistence (Fig 2·6C). At a population level, the implications of this decreasing duration of nongenetic similarity was typically to increase noise in the system (Fig 2·7B). However, this relationship with population noise depended on which parameters were changed to lower the nongenetic similarity duration (Fig 2·7B).

Another interpretation of what the degrading enzyme B does in lowering the duration of nongenetic similarity is enable the population to exist at a variety of noise levels. While low noise is often desirable in engineered biological systems, natural systems use noise as a feature in ways that could be useful in biological circuit design [78]. When the degrading enzyme is less active, resulting in longer duration of nongenetic similarity, the system is constrained as to which levels of noise it can have (Fig 2·7B). In that sense, these results align the conclusions of previous experimental work [56], in concluding that nongenetic inheritance serves to constrain population heterogeneity. With this conclusion in mind, we next want to revisit the measurements from experimental work on nongenetic similarity duration.

While the toy models in this work recreate the existence of nongenetic similarity that then decays over time, the timescales in our simulations of two or three generations do not match the six or eight generations measured experimentally [56]. One possible explanation for the discrepancy is the timescale of gene expression. The toy models here bundle the processes of transcription, translation, and protein folding into a single protein producing event. In reality, these processes can play out over the order of tens of minutes. In exponentially growing bacteria, these tens of minutes could account for one or two generations worth of time. However, a simple time delay function is insufficient to fully explain these differences. Another factor is that experimentally-measured behaviors for nongenetic inheritance, whether cell growth and division time [56] or cell motility [55], are complex behaviors that involve interactions of tens if not hundreds of molecular species. As we had seen in our models, where adding components widened variances amongst a population and enabled nongenetic similarity to exist at all (Fig 2·2), perhaps the involvement of these large numbers of players would increase the duration of nongenetic similarity all the way to the measured six or eight generations. Such expansion of these models is one possible future direction for this work.

The most complex model in this work, the Production and Degradation circuit, had three components with two interactions. While this was sufficient to produce a range of nongenetic similarity durations and relationships to population noise, it still represents a simple pathway. The addition of more components, say turning A into a transcription factor that then activates production of protein C, may be sufficient to further widen the variance and produce longer durations of nongenetic similarity. Along with more components, more types of interactions and regulation could change behavior. Negative feedback is a common motif used to reduce noise in biological circuits [79, 80, 81]. What would happen if M negatively regulated itself as to decrease noise and how would that change the duration of nongenetic similarity as well as its relationship to population noise? One could also test feedback using the signal molecule A or add in positive feedback, which is commonly associated with increasing noise [82]. Further exploration of a variety of circuit architectures using toy models enables the ideation of new circuit designs to achieve cell behavior as well as could provide new ways to understand the biological design of signaling circuits.

2.4 Methods

Code writing

All code was written in Python 3.9.12 using Spyder 5.3.3 for an IDE and compiled using IPython 7.31.1. Packages used included NumPy 1.22.4 [83], SciPy 1.8.1 [84], and Matplotlib 3.5.2 [85].

Random number generation

Random number generation was done using NumPy's random generator function with seed 1000.

Simulation hardware

Code for shorter simulations and analysis was run locally on a Dell Latitude 5420 running Windows 10 Enterprise 21H1 and using an Intel 11th Gen Core i7 processor (3.0 GHz). Code involving longer simulations was run on the Janelia Compute Cluster, which runs on Oracle Linux 8.3 and uses IBM Spectrum LSF 10.1 for job management. Cores were composed of either Intel SkyLake (2.7GHz Platinum 8168) or Intel Cascade Lake (3.0 GHz Gold 6284R) processors.

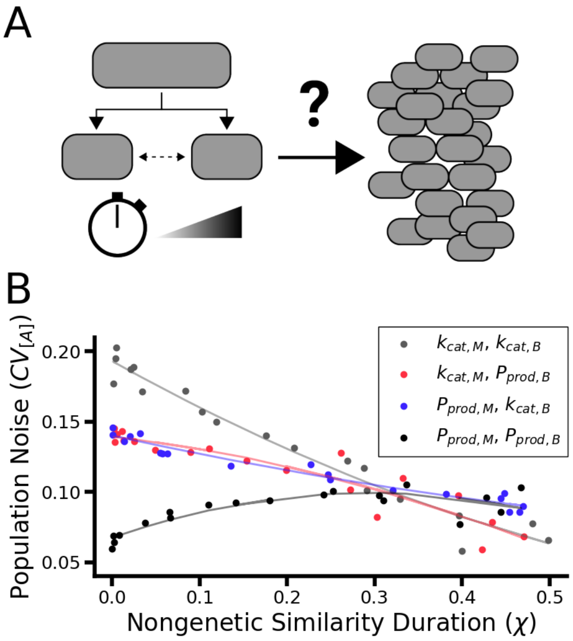

Figure 2·7: Short nongenetic similarity duration corresponds to wide population noise parameter space — (A) Schematic of experiment motivation. Given our metric of memory in the signal molecule is tunable, we wondered how these different durations would affect the behavior of a population. **(B)** Population heterogeneity (as measured by coefficient of variation of concentration of signal molecule A, $CV_{[A]}$) as a function of nongenetic similarity decay root (χ). Different combinations of variables swept over this range resulted in different relationships between noise and the χ value. In three of the four cases (ones where at least one of the parameters being altered was the k_{cat} value: gray, red, and blue dots), the population noise increases as the nongenetic similarity duration decreases. The exception is the case where both variables being altered are the P_{prod} variables (black dots), where the noise decreases at low nongenetic similarity duration values. Lines are quadratic fits intended as guides for the eye. n=1000 cells for each parameter combination.

Chapter 3

Biophysical investigation of cyclic di-GMP as a temporal filter for biofilm formation in *Bacillus subtilils*

3.1 Introduction

3.1.1 Bacterial biofilms as a model system for multigenerational behavior

In considering natural systems in which to investigate signaling pathway architecture underlying multigenerational processes, it is hard to imagine a more prolific example than bacterial biofilm formation. This is because it is estimated that over half of the cells on Earth are bacterial cells living in biofilms [86]. This chapter will begin with a review of what bacterial biofilms are, the general steps of their formation, and how signaling pathways underlying biofilm formation commonly use cyclic nucleotides as signal molecules. I will then focus on what is known about *B. subtilis* biofilms in particular, and why the relatively unknown role of cyclic nucleotides in *B. subtilis* biofilm formation represents an opportunity to understand signaling pathway behavior and cell fate in a natural multigenerational process. Lastly, I will give a brief overview of tools used for experimental measurements of intracellular c-di-GMP before diving into the results of my joint theory-experiment approach.

3.1.2 Biofilm formation is a multigenerational process

Bacterial biofilms are multicellular communities of bacteria adhered to a surface and protected by an extracellular matrix [87]. The predominance of this lifestyle is due to the protections biofilms offer from the external environment, such as resistance to antibiotics [88].

Though it can differ by species, biofilms are generally formed when a motile cell adheres to a surface. As that cell divides, subpopulations of its progeny will start to produce biofilm matrix. The biofilm will then mature into a three-dimensional structure [89, 90]. Ultimately, the biofilm will degrade a portion of its matrix and release some cells into the surrounding environment [91]. As biofilms can start from a single cell and grow to hundreds of thousands [92], the process of building a biofilm often takes days and many cell generations. This long-term process also requires a large metabolic investment [93], and the decision to stop expressing flagellar genes can take three generations to reverse [94]. Therefore, bacteria are highly incentivized to undergo biofilm formation only under certain conditions. One of the strategies they have in which to process environmental conditions is through the intracellular signaling. A common class of intracellular signaling pathways related to biofilm formation is the family of cyclic nucleotides.

3.1.3 The role of cyclic nucleotides in bacterial biofilm formation

The first discovery of a cyclic nucleotide molecule playing a role in intracellular signaling in bacteria was that of cyclic diguanylate (cyclic di-GMP, c-di-GMP, or cdG) in 1987 [95]. Since this initial discovery, its role in signaling and behavior has been widely studied [96, 97, 98, 50, 99, 64, 100], along with the roles of several subsequently discovered cyclic nucleotides [101, 102, 103]. In particular, cyclic di-GMP signaling has been implicated broadly for its role in bacterial biofilm formation [104]. Direct evidence for the role of cyclic di-GMP in biofilm formation has been shown in microbes as diverse as *Acinetobacter baumannii* [105], *Shewanella putrefaciens* [106], *Pseudomonas aeruginosa* [107], the pathogenic *Escherichia coli* O104:H4 [108], *Vibrio cholerae* [109], and *Salmonella enterica* serovar Typhimurium [110].

However, despite all of these studies in a diversity of microbes, the role of cyclic di-GMP signaling in gram-positive bacteria remains relatively understudied [111]. This is true even in the model gram-positive organism, the soil microbe *Bacillus subtilis*. This gap in knowledge is notable, as *B. subtilis* has served as a model organism for cell fate and decision-making for decades.

3.1.4 ***Bacillus subtilis*** as a model organism for cellular decision-making

Initially drawing interest for its sproulation process, *Bacillus subtilis* has long been studied, even adored [112], as a model organism for cell fate decisions [113]. As discussed in **Chapter 1**, the functional role of noise in biological circuits was mapped out by examining the competence circuit in *B. subtilis* [34] and expanded on in looking at its response to stress [114] as well as its entry into sporulation [48, 35]. In fact, much of our understanding in how cellular noise works with feedback loops to create bet hedging decisions came from studies in *B. subtilis* [115].

Additionally, the multitude of phenotypes *B. subtilis* takes on during biofilm formation has been an area of intense study [116, 117, 118]. Of particular note is the "mutually repressing repressors" motif that leads to a binary cell fate decision between motility and matrix production [119]. Even slight mutations on one of these regulators, SinR, causes changes in the frequency of different phenotypes [120]. This switch has been shown to be memoryless [121] and has been recreated synthetically *in vivo* [122].

A critical component underlying all of these cell fate decisions, from competence to biofilm formation to sporulation, is what is referred to as the "central phosphorelay" in *B. subtilis*: a chain of kinases whose activity terminates in the phosphorylation of the central regulator Spo0A [123]. Spo0A alone is known to regulate over 120 genes in the *B. subtilis* genome [124] and has pulsatile activity tied to the replication of the genome during a cell cycle [125]. Other regulators that affect the phosphorelay have been shown to be critical in biofilm formation [126]. In addition to its role in sporulation, biofilm formation, and competence, Spo0A has also been shown to have a role in c-di-GMP regulation in *B. subtilis*, where active Spo0A represses trancsription of the sole cdG-degrading PDE enzyme in the *B. subtilis* genome [127]. Yet despite the link between the c-di-GMP pathway and a key regulator in *B. subtilis* fate decisions, the role of c-di-GMP in *B. subtilis* remains relatively unknown.

3.1.5 The understudied role of cyclic di-GMP in *Bacillus subtilis* biofilm formation

To the extent it has been studied in *B. subtilis*, c-di-GMP is largely associated with swarm motility. The first comprehensive study of c-di-GMP in *B. subtilis* showed high concentrations led to a inhibition of general motility [128] before other evidence exhibited an inhibition of swarm motility in particular [129]. The mechanism of this motility inhibition is likely c-di-GMP's role in binding to the DgrA protein, which acts as a clutch on the flagellar stator [130]. As motility halting is a typical step in biofilm formation, and swarm motility often precedes biofilm [131], this link between c-di-GMP and motility offers hints at its role in biofilm formation as a whole.

Other prior work has shown more direct links between c-di-GMP and *B. subtilis* biofilm formation. A few molecular-level studies have examined the putative c-di-GMP receptors in *B. subtilis* [132, 133] and shown that high levels of c-di-GMP is associated with production of a non-canonical component of the biofilm matrix. Critically, it was shown that the proteins responsible for this non-canonical matrix component localized to cell poles, where it has been shown cdG-producing DGC molecule in *B. subtilis* also localize to [65]. The first single-cell study of c-di-GMP levels in *B. subtilis* [49] showed correlation between c-di-GMP levels and cells producing biofilm matrix. However, when intracellular c-di-GMP was artificially increased using a knockout strain, this did not lead to a higher percentage of cells expressing the matrix-producing phenotype, suggesting that c-di-GMP was not a causative signal molecule in driving this phenotypic differentiation. While c-di-GMP has been implicated in aspects of *B. subtilis* biofilm formation, a full picture of its role in combination with known cell fate decisions remains unknown.

There exist hypotheses in the field as to the role of c-di-GMP signaling that remain untested due to a lack of tools to do so. One hypothesis in the field is that second messengers serve in a capacity for “sustained sensing” [134]. And, more broadly, nucleotide second messengers are often implicated in cell decision making [135]. However, these different pictures of c-di-GMP, and second messenger signaling, have not been unified in the context of other cell fate decisions. *B. subtilis* offers a system with well-defined cell fate decisions but also a relative gap in understanding of the interplay of c-di-GMP signaling with these cell fate decisions. Others have argued that a critical gap in the field is the measuring of c-di-GMP at the single cell and subcellular level, which can now be enabled with new tool development [136]; in particular, genetically-encoded fluorescent sensors are highlighted as an effective class of tools to make these measurements.

3.1.6 Existing approaches for live-cell imaging of cyclic di-GMP

There have been many strategies developed over recent years for live-cell imaging of c-di-GMP levels in individual cells, but all have limitations that prevent them from being useful for studying the pathway dynamically. Some groups have taken the strategy of promoter-based reporter systems, in which a c-di-GMP responsive promoter is cloned upstream of a fluorescent protein such that, in the presence of cyclic di-GMP, the fluorophore is produced and fluorescence values can be read as a proxy for cyclic di-GMP concentration [137]. Especially in gram-negative bacteria, cyclic di-GMP often binds to transcription factors to activate expression of biofilm-related genes, such as the *cdrA* gene in *P. aeruginosa* [138]. While these systems are likely to respond to physiologic levels of c-di-GMP, their response time is bounded by the delay between a change in the level of c-di-GMP and the change in the fluorescence level. In the case of an increase in cyclic di-GMP, the reporter only responds after gene expression of the fluorophore and cyclization reaction, which occurs on the order of tens of minutes to over an hour, depending on the fluorophore [139]. In the case of a decrease in cyclic di-GMP, if the lifetime of the fluorophore is long, the cell relies on dilution from cell division to decrease the signal. In the case of dynamics, promoter-based reporters lack sensitivity in cases of both increase and decrease of the target molecule.

Another strategy for live-cell imaging of cyclic di-GMP that lacks temporal resolution is the riboswitch-based reporter. One design created a sensor by placing a c-di-GMP binding site between a constitutive promoter and a YFP gene such that, in the presence of c-di-GMP, the riboswitch is bound and translation cannot occur, leading to a decrease in YFP fluorescence [49]. Similar to the promoter-based reporter, this sensor is temporally limited by the timing of protein dilution from cell division (now for the on dynamics, rather than the off) and by the timing of fluorescent protein maturation (now for the off dynamics, rather than the on).

Rather than relying on fluorescent protein expression and maturation timing, other sensors have been developed that use protein-based, genetically-encoded sensors. One strategy for c-di-GMP sensor development came from using the FRET-based class of reporters [140]. Here, the c-di-GMP binding domain PilZ was placed between two fluorophores such that, in the presence of c-di-GMP the fluorophores were pulled apart and the FRET between them deceased. The ratio of the FRET between fluorophores to the fluorescence of the donor CFP could then be used as a proxy for c-di-GMP concentration. While this FRET sensor was not temporally bound by gene expression in the way the other sensors were, its K_D value exhibited a high temperature dependence and had small signal to noise ratio due to a dim CFP. Another genetically-encoded sensor is a split GFP sensor, in which the GFP protein was broken into three pieces, one with barrels 1-9, one with one barrel and the FimX c-di-GMP binding domain, and the last with the final barrel and a PilZ c-di-GMP binding domain [141]. The three pieces of the sensor are constitutively expressed and, in the presence of cyclic di-GMP, the PilZ and FimX domains are brought together. With those bound, binding of the remainder of the GFP allows the fluorophore to fully bind and mature, producing fluorescence. These sensors have the necessary features to yield high spatiotemporal resolution measurements of c-di-GMP concentrations in live cells.

3.1.7 A joint theory-experiment approach to c-di-GMP dynamics in *B. subtilis*

Given *B. subtilis* biofilm formation is a multigenerational process with many cell fate processes mapped out but lacks comprehensive study of the dynamics of the c-di-GMP pathway, I found this a compelling system in which to study the multigenerational behavior of a known signaling pathway. I began my approach by developing a kinetic model to simulate c-di-GMP regulation in *B. subtilis* before moving on to experimental implementation of genetically-encoded biosensors to generate data to compare to the predictions of the model.

3.2 Results

3.2.1 Construction and parameterization of a model to simulate the c-di-GMP pathway in *Bacillus subtilis*

To build a predictive model for c-di-GMP dynamics in *Bacillus subtilis*, I first needed to identify all of the molecular agents involved and parameterize their kinetic parameters. c-di-GMP is formed by three different diguanylate cyclases in *B. subtilis*: DgcK, DgcP, and DgcW. Previous single molecular work has quantified the distribution of counts of both DgcK and DgcP [65]. Using these distributions along with the typical bacterial mRNA transcript burst frequency of 10 transcripts per burst [142, 143], I could tune a burst frequency parameter to match these distributions (Fig 3·1). Based on these values, I could then estimate the copy number of DgcW based on a study of global transcriptome levels in *B. subtilis* [144].

With the copy numbers estimated from known values in the literature, I then considered the production of c-di-GMP. As this is an enzymatic reaction, one can think about the production rate in terms of Michaelis-Menten kinetics [75], as in Equation 3.1:

$$R_{prod,cdG} = \frac{k_{cat,DGC}(DGC)(GTP)^2}{K_{M,DGC} + (GTP)^2} \tag{3.1}$$

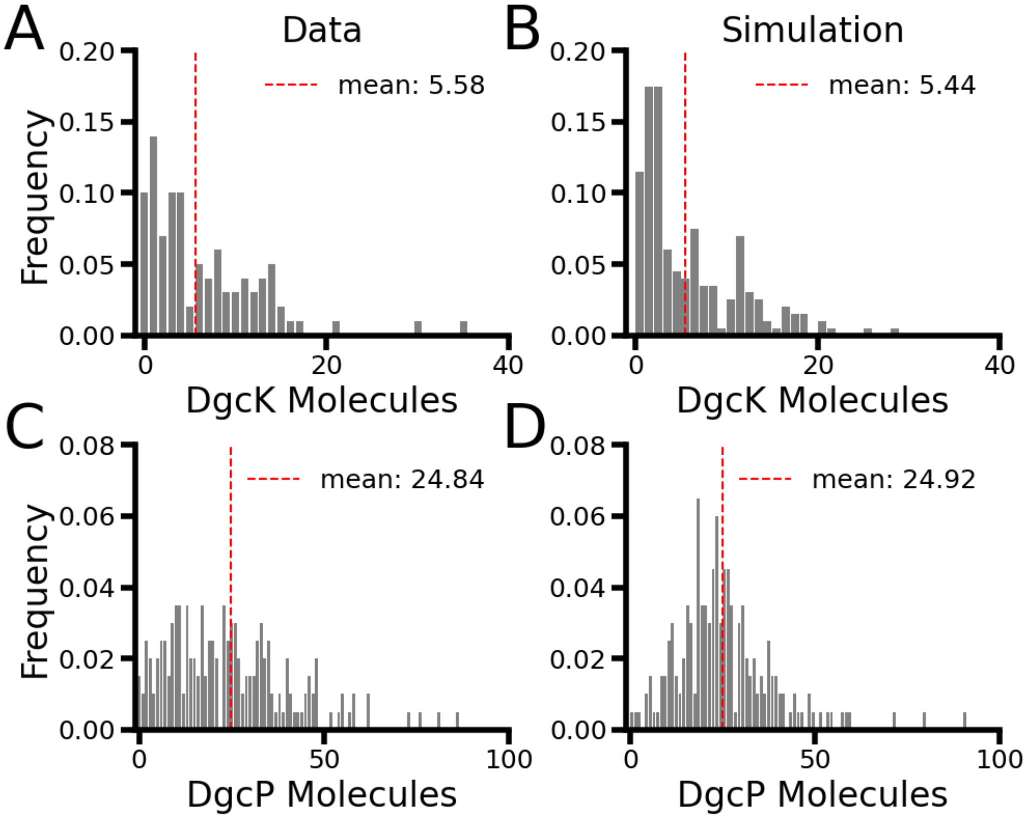

Figure 3·1: DGC production simulation tuned to single molecule data — **(A)** Data from [65] on counts of DgcK in single cells. **(B)** Counts of DgcK in 1000 simulated cells. With a burst size of 10, the burst frequency was tuned to match the experimental data. **(C)** Data from [65] on counts of DgcP in single cells. **(D)** Counts of DgcP in 1000 simulated cells. With a burst size of 10, the burst frequency was tuned to match the experimental data. For simulated data, 1 count of DgcK and DgcPs were taken from each cell at a randomized time point over a three hour simulated time period.

where $k_{cat,DGC}$ is the maximal reaction rate per enzyme, DGC is the number of DGCs in the cell, GTP is the number of GTP, and $K_{M,DGC}$ is the half-saturation constant of the DGC. Typical k_{cat} for a DGC is 1 min^{-1} [64] and K_M values are on the order of $2\mu M$ [145]. Here, we can make a simplifying assumption. Given a typical K_M value of 2μM and a typical GTP concentration on the order of 5 mM [71], the DGCs would be operating at saturation, simplifying the production equation to Equation 3.2:

$$R_{prod,cdG} = k_{cat,DGC}(DGC) \tag{3.2}$$

For the c-di-GMP degrading phosphodiesterase, PdeH, I similarly considered the copy numbers and kinetic parameters. I ballparked the copy numbers using the same transcriptomic profile we used for DgcW [144]. The typical k_{cat} value of a PDE is on the order of 1 s^{-1} [145, 146]. Note that this is much faster than the typical k_{cat} of the DGC enzymes. I was not able to find literature estimates of the K_M value, so that has been left as a free parameter in the system I will explore later. I can then describe the rate of degradation of cdG as follows:

$$R_{deg,cdG} = -\frac{k_{cat,PDE}(PDE)(cdG)}{K_{M,PDE} + (cdG)} \tag{3.3}$$

where $k_{cat,PDE}$ is the maximal reaction rate, PDE is the number of PDE molecules in a cell, cdG is the number of cdG molcules, and $K_{M,PDE}$ is the half-saturation constant. For the initial parameterization of the burst frequency of PdeH and the $K_{M,PDE}$ value, I made use of the fact it is estimated that previous reports indicated a threefold difference in cyclic di-GMP concentration between a strain with PdeH and without [49]. I simulated 1000 cells of the $\Delta spo0A$ strain and the $\Delta pdeH$ strain and compared concentrations of cyclic di-GMP until a threefold difference was achieved

(Fig 3·2). Note that the fluorescence values of the $\Delta pdeH$ strain in the experimental data run close to 0 counts, suggesting the possibility the sensor may be pushed up against its dynamic range, leading to a narrow peak in values (Fig 3·2A).

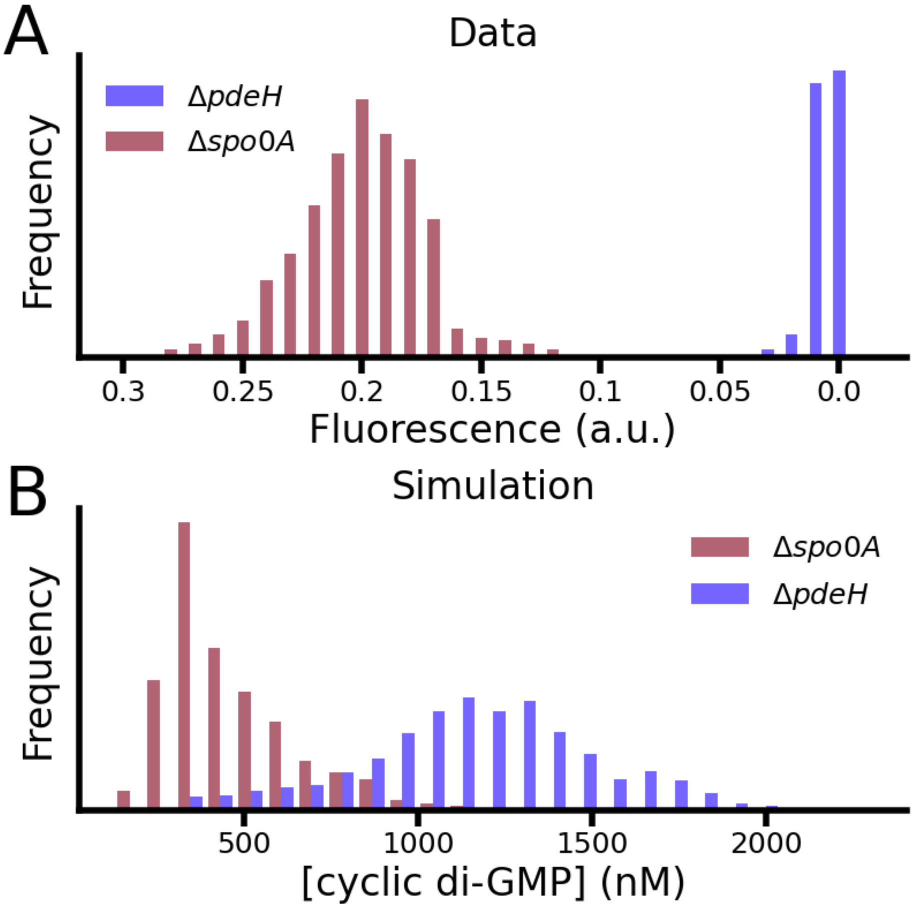

Figure 3·2: Parameters for cdG-degrading enzyme tuned to match experimental data — A. Replot of data from [49] showing difference in reported cdG concentrations using riboswitch-based fluorescent reporter. The two knockout strains correspond to a low ($\Delta spo0A$) and high ($\Delta pdeH$) subpopulation of cdG concentration. **(B)** Distribution of c-di-GMP concentrations in simulated cells matching knockout strains from experimental data. The production rate and K_M value of the PdeH enzyme was tuned to match the threefold difference in c-di-GMP noted in the experimental literature. n=1000 cells for each strain.

With all the kinetics fitted, I could then fully simulate the dynamics of cdG using a Gillespie algorithm. Stochastic simulation was chosen to account for behavior stemming from low copy numbers in the enzyme counts. The full Gillespie rate equation for the model is as follows:

$$R_{total} = R_{prod,DGC} + R_{prod,PDE} + R_{prod,cdG} + R_{deg,cdG} \tag{3.4}$$

The length of the time step between events is then given by Equation 3.5:

$$\tau = (\frac{1}{R_{total}}) * \ln \frac{1}{r_1} \tag{3.5}$$

where r_1 is a random number pulled from a uniform random distribution spanning 0 to 1.

To simulate cell growth I preset a time parameter for a fixed cell cycle time of T_{cc} = 30min, as has been observed experimentally [147, 148]. At each event time determined from the Gillespie simulation, I then calculated the new cell volume using Equation 3.6:

$$V_{new} = V_{old} + \frac{\tau}{T_{cc}} \tag{3.6}$$

The model then assumes linear growth of the cell body over this cell cycle time, doubling from a volume of 0.8 to 1.6 μm^3 [149]. When a cell divides, the intracellular contents are divided up between daughter cells using a binomial random variable and the volume is reset to 0.8 μm^3. This division approach was chosen because previous single-molecule studies have indicated that, in well-mixed bacterial systems, the inheritance of single molecules follow binomial probability [39, 40]. This setup provided all of the key pieces in order to make predictions of the dynamics of cyclic di-GMP in *B subtilis*.

With the kinetics of the signal pathway fitted to data and the Gillespie algorithm used to simulate stochastic events, the model could then simulate any given numbers of cells for any given number of generations (Fig 3·3).

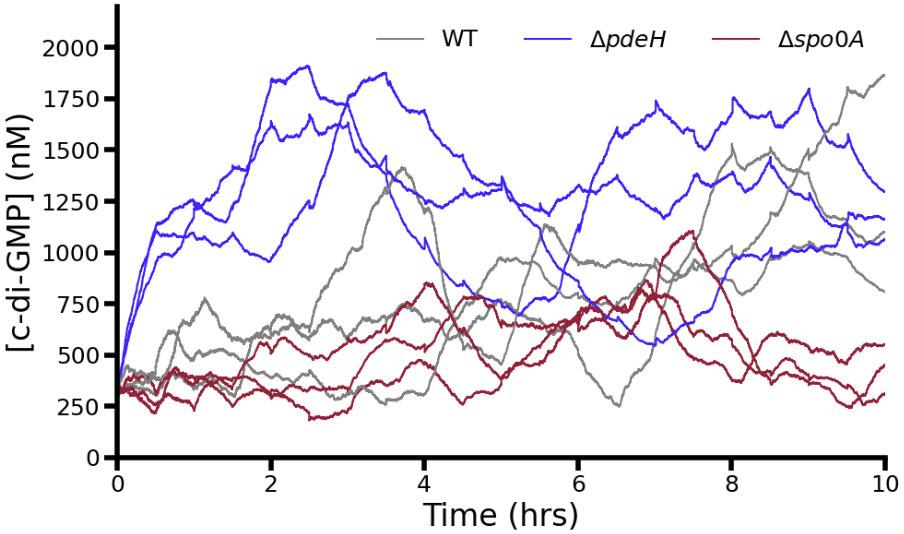

Figure 3·3: Sample time traces for three simulated strains — With kinetic parameters fitted to data, the Gillespie algorithm stochastically simulates the dynamics of cyclic di-GMP. Here, three cells of three different strains were simulated to exemplify the stochasticity. The wild-type cells (WT, grey line) start with Spo0A repression off, but stochastically switch it on, leading to c-di-GMP accumulation and ending the simulation at high concentrations. The Δ*spo*0*A* cells (maroon) are unable to repress the *pdeH* operon, maintaining expression of the degrading PdeH enzyme, keeping c-di-GMP levels low throughout the simulation. Lastly, the Δ*pdeH* strain (blue) start with a small amount of c-di-GMP, which quickly increases in the absence of c-di-GMP degradation.

The picture that the parameterized model leaves is one of the Spo0A repression on the *pdeH* operon creating a bimodal distribution in c-di-GMP concentrations. When the repression is off, as in the Δ*spo*0*A* strain, the PdeH enzyme is expressed and degrades c-di-GMP, keeping the cellular levels low. Upon repression, the PdeH enzyme is diluted out, causing c-di-GMP levels to rise.

3.2.2 Quantitative prediction of c-di-GMP accumulation time in *Bacillus subtilis*

With the parameters of the model that had data fitted, I then examined the output of the model and its predictions. As previous studies had examined the distribution of c-di-GMP values at a static point in time [49], I was mostly focused on the dynamical predictions of the model. Given the hypothesis that the bimodal distribution of c-di-GMP was stemming from the repression of the Spo0A regulator, I wanted to see what the predicted time scale of that switch would be: how long would it take, once Spo0A repression was active, to accumulate to the high concentrations of c-di-GMP.

To do this, I ran a version of the model in which the wild-type cells would stochastically switch on Spo0A repression and, once it was turned on, not turn it off, mimicking a one-way stress response, as might be seen in biofilm or sporulation-inducing media. I then took the concentration of c-di-GMP curves for each of 1000 cells and plotted them as a function of time since Spo0A repression was turned on (Fig 3·4, black lines). To get a sense of the characteristic accumulation time, I then took an average of all of these accumulation curves (Fig 3·4, magenta line). The model predicts the accumulation to take around 3 hours, corresponding to six cell generations in the model. This likely stems from the need to dilute out the fast-acting PdeH enzymes and time for the slow-producing DGCs to catch up. Thus the asymmetry in the kinetics of the pathway enable the long-term accumulation.

This prediction of a 3 hour accumulation time following activation of Spo0A repression is contingent upon the model parameters, two of which were left as free parameters with order-of-magnitude estimations (the production rate of PdeH and the half saturation constant of PdeH). I next wanted to assess the effect of these free parameters on the accumulation dynamics of c-di-GMP. To do this, I turned to a

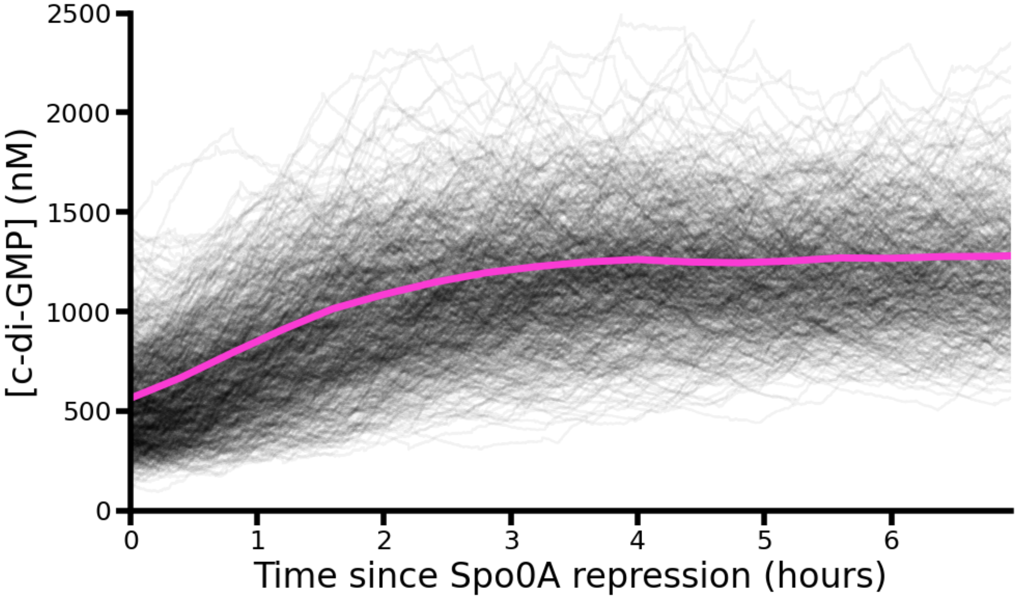

Figure 3·4: cyclic di-GMP accumulates in *B. subtilis* over six generations following Spo0A Activation — To quantify the duration of cyclic di-GMP accumulation, 1000 cells were simulated for a 10 hour period with irreversible Spo0A activation. The time step that Spo0A was activated was identified and the c-di-GMP concentration curve for that cell was plotted from that time point as t=0. This was done for all 1000 cells (black lines in trace). To quantify the characteristic time, an average accumulation curve was computed over all the time traces and shown (magenta). The accumulation is predicted to take 3 hours, corresponding to six cell generations in the simulation.

deterministic version of the model, which uses continuous equations rather than the discrete, stochastic ones. The model simulates the change in cyclic di-GMP concentration and PdeH concentration, starting with a set copy number and then turning off production, as Spo0A repression would, and plotting the accumulation curve. The change in c-di-GMP concentration is modeled with Equation 3.7 and the change in PdeH is modeled with Equation 3.8:

$$\frac{d(cdG)}{dt} = k_{cat,DGC}(DGC) - \frac{k_{cat,PDE}(PDE)(cdG)}{K_{M,PDE} + (cdG)} - \frac{ln(2)}{T_{cc}}(cdG) \tag{3.7}$$

$$\frac{d(PDE)}{dt} = -\frac{ln(2)}{T_{cc}}(PDE) \tag{3.8}$$

where cdG is the amount of c-di-GMP in the cell, DGC is the total amount of DGC molecules, PDE is the total amount of PDE molecules, and T_{cc} is the cell cycle time.

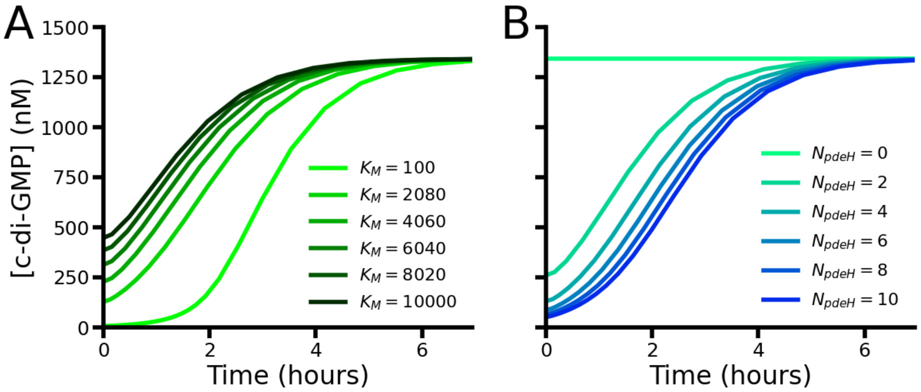

Figure 3·5: Effects of free model parameters on c-di-GMP accumulation time — (A) Effect of different values of $K_{M,pdeH}$ on cyclic di-GMP accumulation time. The deterministic version of the model was run with a sweep of values for the half saturation constant of the PdeH enzyme. Different values affect both the set point at low cyclic di-GMP (Spo0A repression off) and the timing of accumulation. **(B)** Effect of different steady-state copy numbers of PdeH on c-di-GMP accumulation time. While altering the value of the low set point for c-di-GMP, the copy number mostly does not affect timing of accumulation. Together, these free variables can be tuned to fit a range of possible accumulation times.

I first swept the half saturation constant ($K_{M,PDE}$) over two orders of magnitude, from 100 nM to 10,000 nM. As the $K_{M,PDE}$ value increased, the low set point for concentration of c-di-GMP decreased and the accumulation time increased (Fig 3·5A). This indicates how the half saturation constant affects both the fold change between the low and high states as well as the accumulation time. I then varied the steady-state copy number of PdeH from 0-10. As the copy number increased, the set point for low c-di-GMP decreased, but the accumulation time did not significantly change (Fig 3·5B). This indicates that, by changing both parameters in tandem, the model allows for some explanation of a range of accumulation times, as could be fit to experimental data. But, the predictions fell within a 2-5 hour accumulation time (Fig 3·5A), and so the idea of c-di-GMP accumulation as a timer mechanism remains plausible over a wide parameter range. To verify this hypothesis from the model, I then turned to experimental methods to measure this accumulation in vivo.

3.2.3 Imaging single-cell biofilm activation in *Bacillus subtilis*

In order to experimentally measure the predictions from the model, I needed a method to visualize biofilm activation in single *B. subtilis* cells over many cell generations. To do this, I adapted the strategy of agarose pad imaging of bacterial microcolonies [150] with the biofilm-inducing media MSgg [151]. Agarose pad imaging involves sandwiching bacterial cells between a glass imaging dish and an agarose pad composed of bacterial media. The bacteria are provided nutrients from the pad and are forced to grow in a two-dimensional monolayer. This enables imaging of microcolonies grown from single cells.

To adapt this strategy for biofilm activation in *B. subtilis*, the pads were made with the biofilm-inducing media MSgg, which is widely used in studies of *B. subtilis* biofilm formation [152]. Cells were grown up in LB to reach exponential phase before being resuspended in MSgg in liquid culture before being placed on the pads. To see if the cells were indeed activating the biofilm phenotype, I verified this method with a well-defined reporter for *B. subtilis* biofilm formation. This reporter uses a promoter fusion of the *tapA* operon, responsible for the protein component of the *B. subtilis* biofilm matrix, to a fluorescent protein. Thus, when the cell begins to produce matrix, the fluorescent protein is also produces, providing a fluorescent indicator the phenotype is active (Fig 3·6A).

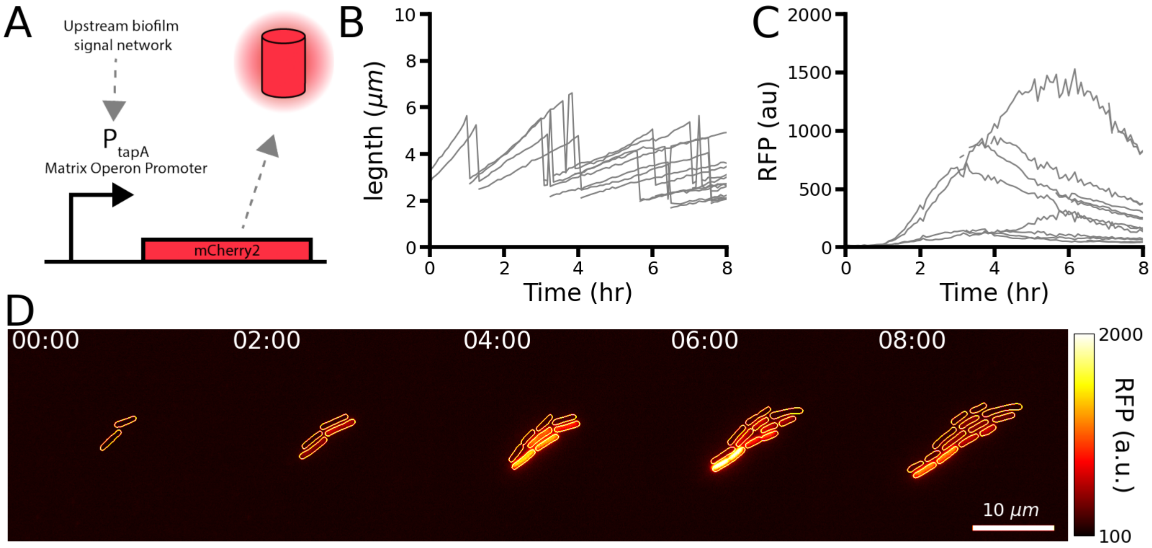

Figure 3·6: Example microcolony of biofilm matrix production at the single-cell level — (A) Schematic of biofilm phenotype reporter. Upstream biofilm signals activate matrix production through the *tapA* operon. A fusion of the *tapA* promoter with an mCherry2 fluorescent protein provides a fluorescent readout when a cell is actively producing biofilm matrix. **(B)** Length traces of single cells in a sample microcolony. As a cell grows, its length increases until the point of division, upon which two daughter cells are produced. The two daughter cells then repeat the cycle, creating the sawtooth pattern. **(C)** Single cell mean RFP traces during microcolony growth. As this sample microcolony grew, a few cells activated biofilm matrix production, as indicated by the rise in RFP signal in a few of the traces. The phenotype is transient, as the fluorescence begins to decrease after peaking. **(D)** Filmstrip of RFP channel during microcolony growth. The raw RFP channel is seen along with cell segmentation calls (yellow outlines). Cells that initially turn on the reporter go on to divide and pass the signal to daughter cells.

The properties of single cells were analyzed by segmenting and tracking single cells and lineages using the DeLTA algorithm [153]. DeLTA uses phase contrast images to track individual cell behavior over time. Length plots of single cells over time give a picture of microcolony growth dynamics (Fig 3·6B). The single-cell RFP traces indicate a few of the cells in this microcolony turned on biofilm matrix production (Fig 3·6C). This activity is evident in the images as well (Fig 3·6D). These traces provide a quantitative, spatiotemporal picture of biofilm phenotype activation in MSgg agarose pad microcolonies.

3.2.4 Biofilm phenotype activation is stochastic and difficult to predict

While I had some success in seeing biofilm activation at the single-cell level in this assay, the results proved difficult to make repeatable. Broadly, I saw three behaviors of single cells in this assay: activation (as in a few of the cells in Fig 3·6), deactivation, in which cells started with high expression of the reporter that then faded over time, and no activation in a microcolony at all. Figure 3·7 exemplifies these behaviors at the single cell level. Most of the microcolonies (I quantify this proportion in the next section) had no activation of the phenotype during imaging (Fig 3·7A,B). The microcolony grew and divided, but no RFP signal was seen during imaging. In some cases, cells that started in the phenotype turned off the reporter (Fig 3·7C,D). As the reporter is a cytosolic mCherry2, and the cells got dimmer with each division, it appears the phenotype was already switched off and I am measuring the dilution of the signal (Fig 3·7D). Lastly, I saw activation in a handful of traces (Fig 3·7E,F), where a single cell will activate the biofilm phenotype and quickly increase its fluorescence (Fig 3·7F).

To get a sense of how common these different behaviors were, I first traced the lineage of each cell to get a full timecourse plot of RFP values for each cell's full

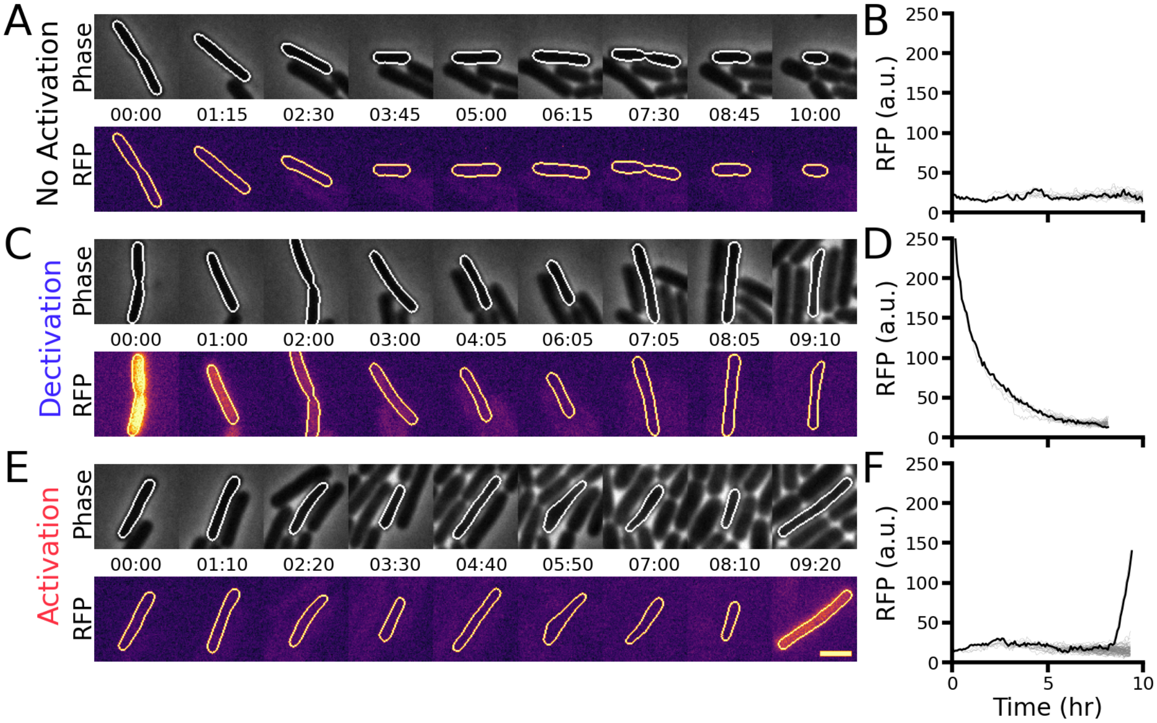

Figure 3·7: Example single-cell behaviors from biofilm phenotype reporter — **(A)** Film strip of single cell in a microcolony in which no activation or deactivation occurred. Segmentation calls are made from the phase channel (top row) and applied to the RFP channel (bottom row). **(B)** Single-cell mean RFP traces for microcolony depicted in **(A)** (grey lines) and the highlighted cell (black line). **(C)** Film strip of phase channel (top row) and RFP channel (bottom row) featuring a single cell in a microcolony where biofilm phenotype deactivation occurred. **(D)** Single-cell mean RFP traces for microcolony depicted in **(C)** (grey lines) with highlighted cell (black line). **(E)** Film strip of phase channel (top row) and RFP channel (bottom row) highlighting a single cell that activated biofilm phenotype during imaging. **(F)** Single-cell mean RFP traces for microcolony depicted in **(E)** (grey lines) with highlighted cell (black line). Scale bar = $2\mu m$.

lineage. By placing a simple threshold value, I sorted the cells into either Activating (starting below threshold and, at some point, increasing above it, Fig 3·8A, red lines), Deactivating (starting above threshold and decreasing below it, Fig 3·8A, blue lines), or having No Activation (staying below threshold the entire timecourse, Fig 3·8A, grey lines). To then quantify the frequencies of these behaviors, I looked at this from the single-cell (Fig 3·8B) and microcolony level (Fig 3·8C). Over $\approx$ 7500 cells and 85 microcolonies, the proportions were roughly similar, with around 90% of both cells and microcolonies having no activation, and about 5% of cells and microcolonies either activating or deactivating.

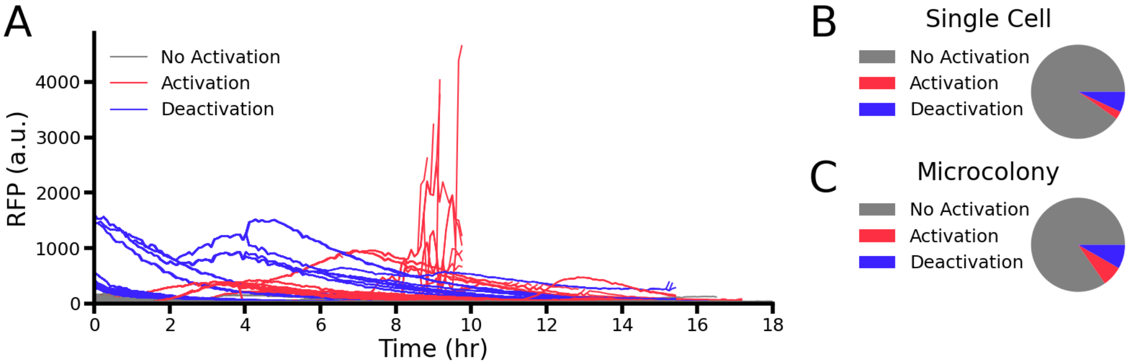

Figure 3·8: Biofilm activation and deactivation is infrequent on MSgg agarose pads — (A) Single-cell RFP timecourse traces of all cells in P_{tapA} reporter experiments. Cells are color-coded depending on whether they activated the biofilm phenotype reporter (red), deactivated it (blue), or had no activation at all (grey). **(B)** Frequency of activation, deactivation, and no activation in single cell traces. For the majority of single-cells examined ($> 90\%$), neither activation nor deactivation occurred ($n_{none} = 6483, n_{act} = 199, n_{deact} = 501$). **(C)** Frequency of activation, deactivation, and no activation at the microcolony level. If a microcolony had at least one cell that activated the phenotype, it was considered active. If it had at least one cell that deactivated, it was considered deactive. If neither occurred, it was categorized with “no activation” ($n_{none} = 72, n_{act} = 6, n_{deact} = 7$). Most microcolonies examined ($\approx 85\%$) had no activation during imaging.

In pursuit of a reliable methodology for measuring biofilm activation under MSgg agarose pads, I then went back into the experimental data to ask if there were specific conditions prior to the cells being placed on the scope that would be predictive of whether they activated, deactivated, or did not activate the biofilm phenotype. To do this, I looked at the microcolony level to see if the behavior of the microcolony was a function of different steps of the pre-imaging growth protocol. I found little correlation between microcolony biofilm behavior and the OD_{600} of the cells from the initial LB grow-up step (Fig 3·9A), the OD_{600} following resuspension in MSgg (Fig 3·9B), the OD_{600} of the MSgg culture directly prior to placement on the agarose pads (Fig 3·9C), nor the fold-change in growth in the MSgg culture (Fig 3·9D).

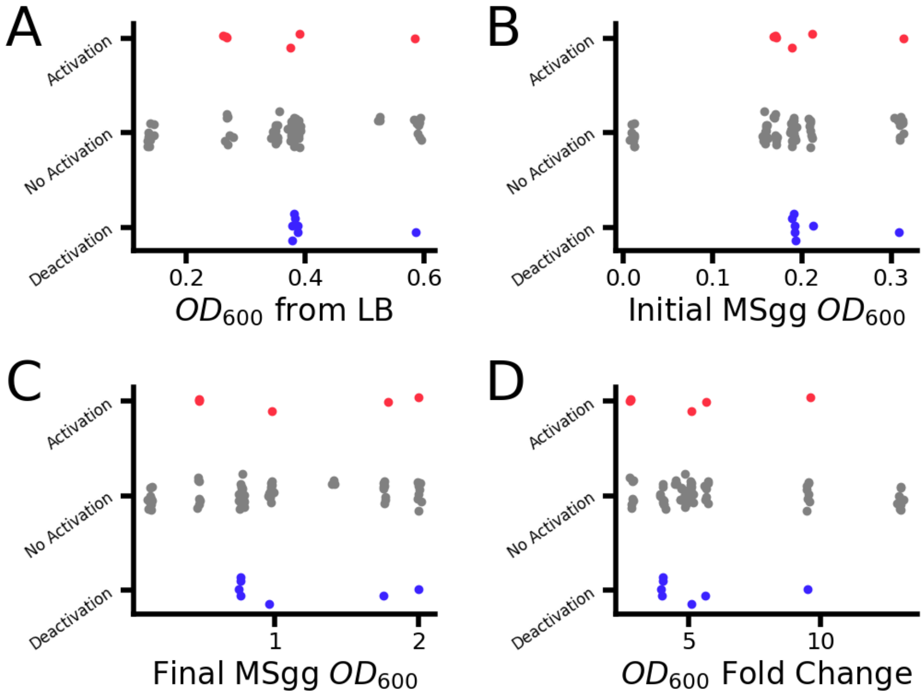

Figure 3·9: Pregrowth conditions are not predictive of microcolony biofilm phenotype behavior — Microcolony behavior was plotted as a function of different stages of the pre-imaging conditions: the OD_{600} in LB culture prior to MSgg resuspension **(A)**, the OD_{600} following MSgg resuspend **(B)**, the OD_{600} in MSgg prior to placing on the pads **(C)**, and the fold-change in OD_{600} during MSgg growth **(D)**. No one measurement appeared to be a predictive parameter. A slight random value was applied to each data point to offset overlapping values.

From these data, I concluded that, while this imaging strategy showed promise in visualizing biofilm activation at the single-cell level, the protocol requires further optimization to make this a reliable method.

3.2.5 Adaptation of a split GFP sensor for c-di-GMP narrowly distinguishes control strains

In order to measure intracellular concentrations of c-di-GMP, I turned to genetically-encoded, fluorescent biosensors. These sensors, in theory, offer the ability to noninvasively measure c-di-GMP in live cells, making them ideal for the kind of quantitative, temporal analysis required to check the predictions of the model. As discussed in the introduction of this chapter, there have been a number of attempts to create effective fluorescent biosensors for c-di-GMP, each with their own shortcomings. In this section, I will show work validating a split GFP-based sensor in *B. subtilis* but also its shortcomings, as well as work on a novel sensor designed to overcome the shortcomings of the published sensors.

I first employed a split GFP-based sensor for c-di-GMP initially created for reporting a synthetic asymmetric division system in *E. coli* [141]. The sensor breaks an eGFP into three parts and fuses two with the c-di-GMP binding domains PilZ and FimX (Fig 3·10A, left). In the presence of c-di-GMP, the PilZ and FimX bind, bringing the two barrels of the eGFP together. Association of the remaining barrels of the protein then bind, enabling the cylcization reaction that can produce fluorescence (Fig 3·10A, right). To adapt the sensor for expression in *B. subtilis*, I designed an operon that would express the three components under the same, high-expressing, constitutive promoter (P_{veg}). I then transformed this sensor construct into three strains: wild-type cells, a ΔDGC knockout strain to serve as a negative control, and a ΔPDE knockout strain to serve as a positive control (see Table 3.2).

When tested in static conditions, the sensor exhibited the expected response.

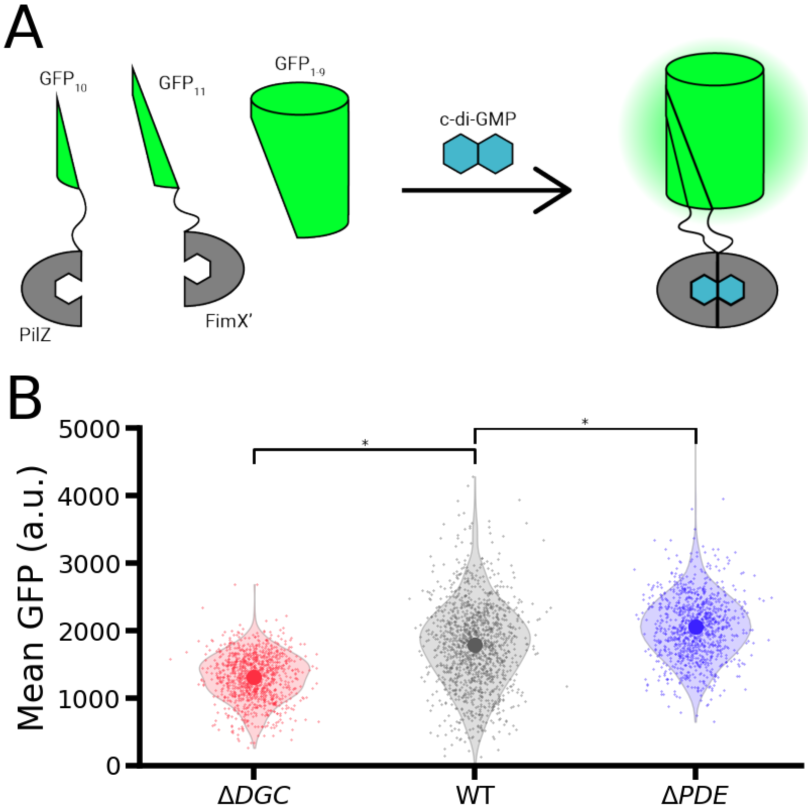

Figure 3·10: Split GFP sensor distinguishes knockout strains in static imaging condition — (A) Schematic of split GFP sensor. The sensor is expressed in three transcriptional units: barrel 10 of GFP fused to a PilZ domain, barrel 11 fused to a FimX domain, and barrels 1-9. In the presence of cyclic di-GMP, the PilZ domain binds followed by the FimX domain. The sensor then associates with the remaining barrels 1-9 fragment and the GFP undergoes cyclization, producing green fluorescence. **(B)** Single-cell mean sensor readout in cyclic di-GMP pathway knockout strains. Cells of each strain were grown to $OD_{600} = 1$ and pipetted onto agar pads for imaging. Single cells were segmented and mean GFP values over the cell area were calculated, following a background subtraction (small dots). Violin plots indicate distribution of values and large dots indicate means. * indicates $p < 10^{-3}$. For the ΔDGC strain, n=1064; for WT, n=1312; for ΔPDE, n=1024.

Colonies of each of the strains were inoculated in LB media and grown up until the culture OD_{600} was about 1.0. These cultures were then diluted tenfold and pipetted onto agarose pads made of the biofilm-inducing media MSgg (see Methods). These pads were then flipped onto a glass imaging dish, pressing the cells between the glass and the pad. The cells were then imaged in phase and GFP channels, segmented, and the mean GFP value for each cell was calculated. In analyzing the distributions, the expected trend was observed, with the ΔDGC strain exhibiting the lowest mean fluorescence value, the wild-type in the middle, and the ΔPDE strain with the highest mean value (Fig 3·10B). Also as expected, the wild-type cell distribution spanned the width of both the ΔDGC and ΔPDE strains (Fig 3·10B, grey violin). This test suggested the sensor was sensitive to the c-di-GMP range in *B. subtilis*, as it could discriminate between the lower and upper ranges of c-di-GMP concentrations. With this confirmed in a static imaging experiment, I next went on to observe the dynamics of the sensor signal in these strains.

To test the sensor in these strains over time, I imaged growing microcolonies from single cells expressing the sensor (see Methods, Sample preparation for timelapse experiments). Each microcolony was cropped out from the timelapse an segmented and tracked using the DeLTA pipeline. The negative control strain had low, but increasing levels of GFP fluorescence during imaging (Fig 3·11A, bottom). This is born out in the mean GFP traces for single cells (Fig 3·11B, red lines) as well as the microcolony mean (Fig 3·11B, black line). The increase in signal may be due to a lack of irreversibility in the sensor, as once the GFP components cyclize, there is no way for the molecule to come undone. Any off-target activity in this strain could then accumulate completed sensor over time, increasing signal. The wild-type strain also had a general increase in GFP over the timelapse (Fig 3·11C, bottom and D, black line) but also exhibited significant heterogeneity in the final few hours of imaging (Fig 3·11C, last three frames of bottom row). This amount of heterogeneity was expected, due to the range of GFP values seen in the static imaging (Fig 3·10). Lastly, the ΔPDE strain similarly exhibited a general increase in fluorescence, with the exception of a noticeable decline in hours 11-12 (Fig 3·11E, bottom row and F, black line). The final frames exhibited a similar level of heterogeneity to the wild-type strain. With the individual cell traces and the microcolony traces quantified, I could then ask if the sensor could distinguish the strains in these timelapse experiments.

Mean traces for each of the microcolonies in a strain (Fig 3·11, thin lines) were further averaged together and plotted over time (Fig 3·11G, thick lines). The means for the control strains are fully distinct during the timelapse, though they narrow in the dip in mean GFP values for the ΔPDE strain in hours 10-12. These data further validate the results from the static experiment that the sensor dynamic range can match that of these knockouts. However, the change is narrow, less than a 50% fold change between the two. This narrow readout is further complicated with the wild-type strain, whose mean value over all of the microcolonies closely follows that of the ΔPDE strain (Fig 3·11, thick blue and grey lines). This was not all that surprising, given how close the means of the WT and ΔPDE strains were in the static imaging experiment (Fig 3·10B). This combination of a narrow output range, which would make quantifying accumulation from a low to high state difficult, along with the increase baseline of fluorescence, likely due to the irreversible assembly of the sensor, suggest this sensor would be difficult to utilize in measuring the accumulation time of c-di-GMP, making it unfit for answering the question posed by the model. Having used an existing tool, I next turned to development of a novel sensor that had the possibility of addressing the shortcomings of the split GFP sensor.

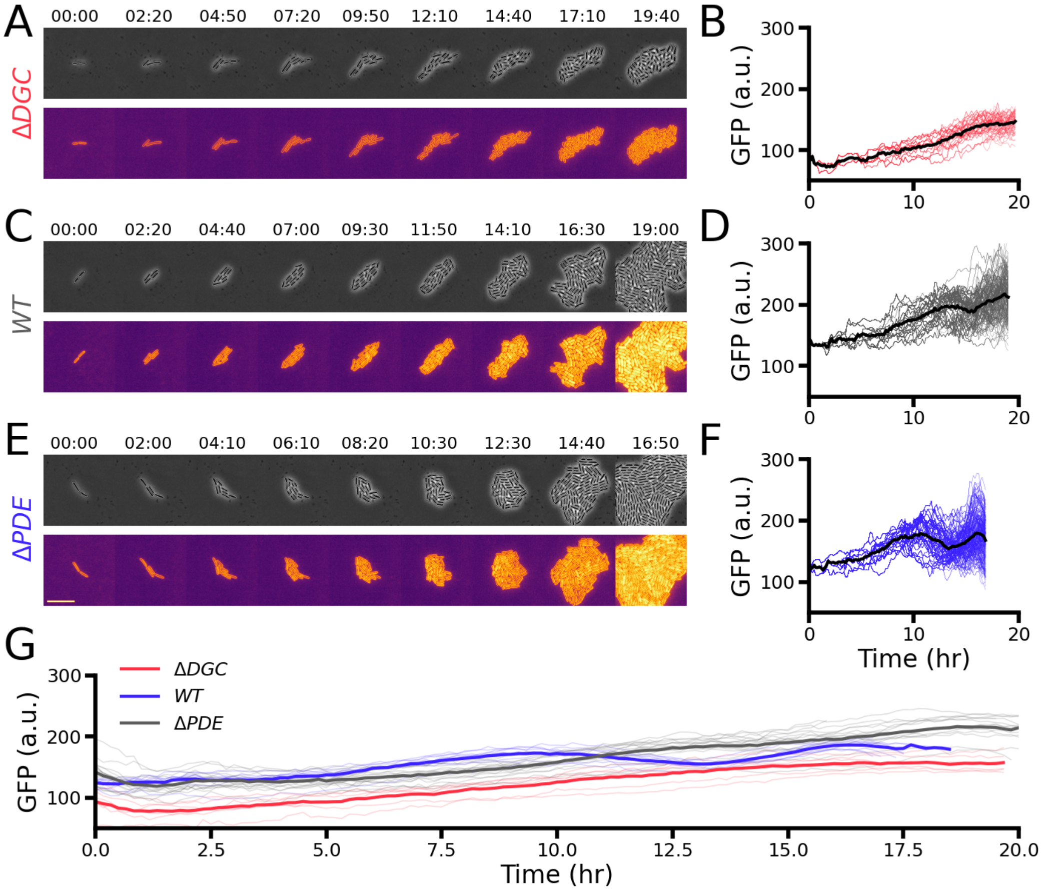

Figure 3·11: Split GFP sensor output shows narrow range in timelapse context — Sample microcolony timelapses (top: phase, bottom: GFP) of strains expressing the split GFP sensor for **(A)** ΔDGC, **(C)** WT, and **(E)** ΔPDE. For each colony, the single cell mean GFP traces are shown (colored lines in **(B)** ΔDGC, **(D)** WT, and **(F)** ΔPDE), as well as the mean of the single cell traces for the whole microcolony (black lines in **(B)**, **(D)**, and **(F)**). All traces show a general increase of signal, suggesting lack of reversibility in the sensor. n=65 cells in ΔDGC trace, n=210 cells in WT trace, n=260 in cells ΔPDE trace. **(G)** Time traces of mean microcolony GFP values (thin colored lines), along with experiment means (thick colored lines) for ΔDGC (red), WT (grey), and ΔPDE (blue) strains. While negative control (ΔDGC) and positive control (ΔPDE) strains show good separation throughout imaging, WT cells show little difference from ΔPDE strain. n=6 microcolonies for ΔDGC trace, n=16 microcolonies for WT trace, n=9 microcolonies for ΔPDE trace.

3.2.6 A novel circularly permuted GFP sensor is entangled with cell growth dynamics

With concerns about the reversibility of the split GFP-based sensor, I turned to a collaboration to develop a novel sensor for intracellular c-di-GMP based on the family of circularly permuted biosensors (reviewed in [154], first developed by [155]). The sensor was designed with the same c-di-GMP binding domains from the split GFP sensor and attached to a sfGFP. In the presence of c-di-GMP, the binding domains cause a conformational change in the sfGFP, allowing it to cyclize and produce fluorescence (Fig 3·12A). *In vitro* characterization of the sensor validated that its binding kinetics operated with μM sensitivity, as is expected *in vivo* (Fig 3·12B). The sensor was then transformed into the same three strains used in testing the split GFP sensor: wild-type, negative control ΔDGC, and positive control ΔPDE. These strains underwent similar testing to the split GFP sensor.

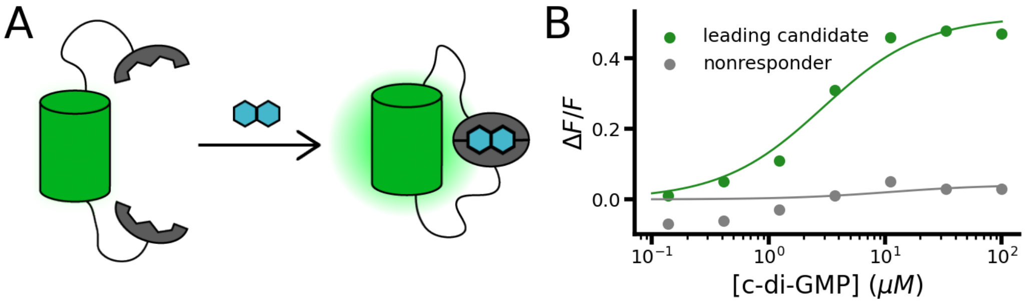

Figure 3·12: A novel circularly permuted GFP sensor for c-di-GMP exhibits signal modulation over physiologic range in vitro — (A) Schematic of a circularly permuted GFP sensor for cyclic di-GMP. The FimX and PilZ cdG-binding domains are fused to a GFP such that, in absence of cdG, the fluorophore is pulled apart such that it cannot cyclize and produce fluorescence. In the presence of cdG, the binding domains cause the GFP to undergo a conformational change that enables it to cyclize and fluoresce. **(B)** In vitro binding data for purified sensor. Isolated sensor was titrated with known quantities of cdG and change in fluorescence was measured. Many sensor configurations did not respond (grey dots) while a leading candidate saw a 50% change in fluorescence at saturating values (green dots).

When performing timelapse testing of the cpGFP-based sensor, a peculiar pattern was observed in the sensor signal in the context of the cell growth rates. In the timelapse experiments, the single cells divided two or three times before having a slowdown in growth (Fig 3·13A, upper row, time stamps 6:45-11:15). During this slowdown, the GFP signal from the cells appeared to increase (Fig 3·13A, lower row, time stamps 6:45-11:15), before decreasing after resuming growth. I hypothesized the increase in signal was not due to an increase in c-di-GMP concentration, but rather a decrease in the growth rate and thus an accumulation of sensor in the cells. To quantify this, took the length traces of single cells (Fig 3·13B, left axis) and took the timepoint-by-timepoint difference to get an estimate of the growth rate (Fig 3·13B, right axis). Indeed, this metric captured a slowdown in growth during the middle hours of this experiment, followed by a resumption. To see if this growth slowdown corresponded with the increase in fluorescence, I plotted the growth rate for each cell in the experiment (Fig 3·13C, left axis) along with the inverse of its mean GFP value (Fig 3·13C, right axis). The decrease and increase in the growth rate nearly perfectly corresponds with the decrease and increase in inverse fluorescence (meaning increase then decrease in absolute fluorescence). From these data, it was clear the growth rates of the cells were intervening with the sensor signal, making it impossible to determine if the sensor was detecting differences between the strains. While the architecture and *in vitro* performance of the sensor remain promising, future design iterations will be needed in order to determine if it is functioning in the context of *B. subtilis.*

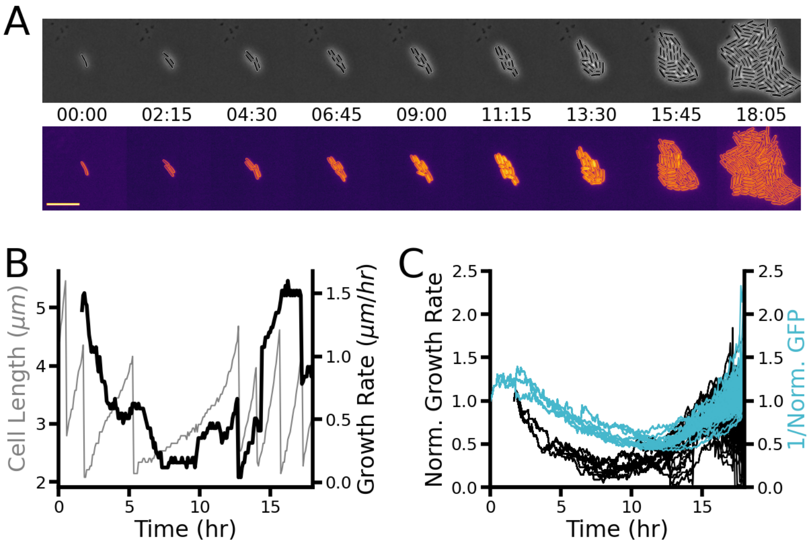

Figure 3·13: cpGFP sensor signal is dominated by growth rate impacting sensor concentration — **(A)** Timelapse of a microcolony expressing the circularly permuted GFP sensor. Segmentation calls from the phase imaging (top) are used to calculate the mean GFP value on a per cell basis (bottom). Scale bar = $10\mu m$. Time stamps are hh:mm. **(B)** Example single-cell trace for calculating growth rate metric. Smoothed derivatives of the cell length over time trace (left axis, grey) were taken to get a metric for instantaneous growth rate at each time point (right axis, black). The point-to-point difference was taken between each point of the length trace. Highly negative points originating from cell division were smoothed by taking the average of the two surrounding values. Lastly, a moving average of 20 frames was applied to get the growth rate metric. **(C)** Overlay of growth rate and inverse GFP fluorescence traces. Normalized growth rates for each cell in the microcolony (imaged in **(A)**) are shown (left axis, black) along with the normalized inverse of the mean GFP for each cell (right axis, cyan). The near overlap indicates a strong correlation between growth rate slowing down and the GFP signal increasing, suggesting the possibility of sensor accumulation driving increase in signal as opposed to c-di-GMP accumulation. n=115 growth rate and GFP traces.

3.3 Discussion

In setting out to examine the multigenerational dynamics of the c-di-GMP circuit in *B. subtilis*, I employed two approaches. In the first, I used simulations of the kinetics of the circuit, based on measured values and fitted to published studies, to make a prediction of the time for c-di-GMP between its low and high subpopulation [49] (Fig 3·4). In the second, I used a number of tools in order to take experimental measurements to compare to the predictions from the model. I adapted the bacterial agarose pad imaging strategy to examine *B. subtilis* biofilm activation on the single cell level in growing microcolonies (Fig 3·7) and then used two different sensors to attempt to measure intracellular c-di-GMP concentrations. While the first sensor, a split GFP-based design, did show signal differences between control knockout strains (Fig 3·10), the limited output range as well as an increasing baseline observed during timelapse imaging (Fig 3·11), hypothesized to be due to the inability of the sensor to turn off once cyclized, rendered it ineffective at taking accurate measurements of c-di-GMP accumulation. I then turned to a novel circularly permuted design, which has off dynamics and does not depend on dilution from cell division to decrease signal. This sensor, however, exhibited a strong dependency on growth dynamics (Fig 3·13) and any signal from intracellular c-di-GMP, if it was sensitive to it, was likely being subsumed by signal changes due to sensor accumulation during periods of slow cell growth.

Future work includes a number of new strategies for improving on the use of the existing biosensors. To normalize out the effects of cell growth in the circularly permuted GFP sensor, a normalizing fluorophore can be fused to the sensor. An effective strategy would be to fuse the HaloTag system [156] to bind with a bright dye [157, 158] to normalize the signal. This design improvement would enable validation of the sensor's dynamic range matching the range of c-di-GMP concentrations in *B. subtilis*. Another strategy would be to use a previously-published FRET sensor [140] for intracellular c-di-GMP concentration. Improved single-cell imaging and analysis would make the sensor easier to implement in my experimental pipeline. Another technical shortcoming of the experimental design is the limitation of imaging cells in a monolayer. As the cells begin to stack, the phase images render it impossible for the segmentation and tracking algorithm to distinguish cells in different layers. One strategy would be to simply confine analysis to the monolayer edge of growing microcolonies, as has been done elsewhere [159]. This would enable longer term imaging, as the image stacks would not have to be cropped immediately prior to cell stacking in the microcolonies. These improvements in the experimental approach have the potential to yield high-quality data for c-di-GMP dynamics in single cells, enabling full comparison to the model predictions.

While perfect matching between theory and experiment would validate the predictions of the model as it stands, discrepancies would lead to formulation of new hypotheses to understand the system. One possible cause for discrepancy could be accounted for by tweaking the free parameters of the model (as illustrated in Fig 3·5). The two parameters could be used to explain a range of difference in timing. One phenomena they could not explain, however, is a decrease in c-di-GMP concentration. The model assumes a one-way transition from low Spo0A activity to high Spo0A activity, but experimental work has shown that Spo0A activity occurs with a baseline increase punctuated with pulses synchronized to the cell cycle [160]. In fact, this particular architecture of Spo0A pulsing speaks to a larger design for multigenerational behavior.

The slow accumulation of active Spo0A through pulsing is part of a robust *B. subtilis* differentiation network. The nature of the Spo0A activity pulses arises from the placement of its kinase and phosphatase on opposite ends of the genome. As the genome begins to be replicated, extra copies of the kinase operon exist, pushing up phosphorylated levels of Spo0A. As the genome replication finishes, the copy number of the phosphatase operon catch up, lowering phosphorylated Spo0A levels [125]. The long term accumulation of active Spo0A allows for *B. subtilis* to decide between competing fate decisions [48] and the pulsing behavior has been shown to be an effective design for delaying cell fate decisions [46]. This behavior of an increase through pulsing has been termed polyphasic feedback, and has been shown to be the most robust activation network architecture to cell cycle timing [161]. As Spo0A activity increasing is already a robust, multigenerational process, it is an interesting open question of how its dynamics combine with repression of the *pdeH* operon to drive c-di-GMP accumulation, here predicted to be its own multigenerational process. Indeed, multigenerational accumulation of a signal is a cellular design with its own associated term: cellular timer.

Cellular timers are a well-characterized class of multigenerational cell signal architecture. In general, the definition is any circuit that enables the long-term accumulation of critical signal molecules is a cellular timer [162]. The structure of cellular timers, in addition to the other strategies listed above, enable their operation to be robust in the face of cellular noise [163]. Spo0A achieves cellular timer behavior through a carefully-tuned polyphasic feedback network. The model results here for c-di-GMP, on the other hand, predict long accumulation coming from dilution of the degrading PDE and slow production from the DGCs. What happens then, when one cellular timer is driving the activity of another cellular timer? Also interesting to ask is how the effects of cell division impact both the Spo0A activity and the c-di-GMP components of this polyphasic feedback circuit. Oscillating c-di-GMP concentrations are not unheard of in the bacterial context, and have been shown to be important for driving *Caulobacter crescentus* cell cycle [164] and *Pseudomonas aeruginosa* surface attachment [165]. Further exploration of c-di-GMP dynamics in *B. subtilis* offers the possibility for insights for processes coupled to polyphasic feedback circuits and perhaps new designs for cellular timers, furthering our understanding of bacterial differentiation in the biofilm lifestyle.

3.4 Methods and Materials

Code writing

All code was written in Python 3.9.12 using Spyder 5.3.3 for an IDE and compiled using IPython 7.31.1. Packages used included NumPy 1.22.4 [83], SciPy 1.8.1 [84], and Matplotlib 3.5.2 [85].

Random number generation

Random number generation was done using NumPy's random generator function with seed 1000.

Simulation hardware

Code for shorter simulations and analysis was run locally on a Dell Latitude 5420 running Windows 10 Enterprise 21H1 and using an Intel 11th Gen Core i7 processor (3.0 GHz).

Bacillus subtilis transformation

All transformations were done in the transformable 3610 *comI*^{Q12L} strain, which has been shown to have identical characteristics as the wild-type 3610, but is more naturally transformable [166]. The desired strain for transformation was streaked out on LB plates and grown overnight at 37°C. The following day, a colony was inoculated in 2mL of LM Media (LB + 3 mM $MgSO_4$) and grown for 3 hours (37°C, shaking 250rpm). Once the colony reached an OD_{600} between 0.8 and 1.2, 100 μL was pipetted into 2mL of MD media (9.25 mL of 1.1x PC Buffer, 400 μL of 50 % glucose, 30 μL of 1M $MgSO_4$, 250 μL of 100mg/mL K-aspartate, and 50 μL of 2.2mg/mL ferric ammonium citrate). Glucose solution was made day of, all other solutions were made from stocks. K-aspartate and ferric ammonium citrate were stored in the fridge, PC Buffer and $MgSO_4$ were stored at room temperature. 1.1x PC Buffer was made from diluting 165 mL of a 10x stock solution (107g K_2HPO_4, 60g KH_2PO_4, 10g trisodium citrate dihydrate, pH'ed to 7.5 with KOH in 1L of diH_20) in 1335mL diH_20 and autoclaving. After 3 hours of growth in the MD medium (37°C, shaking 250rpm), 200 μL of cells are placed in a new 15mL culture tube with DNA (1ng of plasmid DNA or genome extract, but no more than $20\mu L$) and incubated for 60min (37°C, shaking 250rpm). 200 mL of this culture was then plated on agar plates with antibiotic (for spectinomycin selection, final concentration of $100\mu g/mL$) and incubated overnight at 37°C. The following day, colonies are restreaked again on selective agar plates and incubated overnight at 37°C. The next day, colonies are picked from the restreaked plate and grown in LB for 3 hours to $OD_{600} \approx 1$ and then stored in glycerol stocks made from a 1mL LB culture, 1mL 50 % glycerol mixture and stored at -80°C.

Generation of *Bacillus subtilis* knockout strains

To generate knockouts in the competent $comI^{Q12L}$ strain, strains from the Koo knockout collection [167] were ordered from the Bacillus Genetic Stock Center (Table 3.1).

The genomes were extracted using a Qiagen DNA Mini Kit following the protocol modification for gram-positive bacteria. Extracts of knockout strain genomes were then transformed into the transformable 3610 $comI^{Q12L}$ strain of *B. subtilis* using

Strain Name	**Genotype**
BKE09120	ΔdgcK::erm
BKK09120	ΔdgcK::kan
BKE29650	ΔdgcP::erm
BKK29650	ΔdgcP::kan
BKE13420	ΔdgcW::erm
BKK13420	ΔdgcW::kan
BKE31740	ΔpdeH::erm
BKK31740	ΔpdeH::kan
BKE24220	Δspo0A::erm
BKK24220	Δspo0A::kan

Table 3.1: Knockout strains from *Bacillus* Genetic Stock Center

standard *B. subtilis* transformation protocol (see above). In the case of the triple DGC knockout, the *dgcP* gene was knocked out using the genome extraction from a triple DGC knockout strain from other work [129] while the *dgcK* and *dgcW* genes were knocked out using genome extractions from the Koo collection.

Bacillus subtilis **strains used in this study**

Strain Name	**Genotype**
WT pB186	3610 $comI^{Q12L}$ $\Delta amyE$::P_{tapA}-R0-mCherry2 $spec^R$
WT pB147	3610 $comI^{Q12L}$ $\Delta amyE$::P_{veg}-R0-GFP_{10}-Pilz-R0-FimX-GFP_{11}-R0-GFP_{1-9}
ΔDGC pB147	3610 $comI^{Q12L}$ $\Delta dgcW$::erm^R $\Delta dgcK$::kan^R Δdpc::tet^R $\Delta amyE$::P_{veg}-R0-GFP_{10}-Pilz-R0-FimX-GFP_{11}-R0-GFP_{1-9}
ΔPDE pB147	3610 $comI^{Q12L}$ $\Delta pdeH$::erm^R $\Delta amyE$::P_{veg}-R0-GFP_{10}-Pilz-R0-FimX-GFP_{11}-R0-GFP_{1-9}
WT pB190	3610 $comI^{Q12L}$ $\Delta amyE$::P_{veg}-R0-FimX-cpGFP-PilZ $spec^R$
ΔDGC pB190	3610 $comI^{Q12L}$ $\Delta dgcW$::erm^R $\Delta dgcK$::kan^R Δdpc::tet^R $\Delta amyE$::P_{veg}-R0-FimX-cpGFP-PilZ $spec^R$
ΔPDE pB190	3610 $comI^{Q12L}$ $\Delta pdeH$::erm^R $\Delta amyE$::P_{veg}-R0-FimX-cpGFP-PilZ $spec^R$

Table 3.2: *Bacillus strains* generated for this study

Biofilm media preparation

1x MSgg solution was created as follows. 2.4μL of MSgg LMP salt/ion mix was created with 1000μL of 50% glycerol, 1000μL of 50 % glutamate monosodium salt, 100μL of 0.7M $CaCl_2$, 100μL of 2M $MgCl_2$, 100μL of 0.1M $FeCl_3 - 6H_2O$, 100μL of 0.05M $MnCl_2$, 1μL of 0.1M $ZnCl_2$, and 2μL of 0.1M thiamine HCl. The 50 % glutamate and 0.1M $FeCl_3 - 6H_2O$ solutions were made day of, all other salt mixes were stored at room temperature. This mixture was then passed through a 0.22 μm syringe filter. 1x MSgg was then made in 25mL batches with 21.6mL of ddH_20, 2.5mL of 1M MOPS buffer (pH 7.0 with KOH), 250μL of 0.5M potassium phosphate buffer (2.72g K_2HPO_4 and 1.275g of KH_2PO_4 in 50mL at pH 7.0), and 601μL of MSgg salt mix.

Sample preparation for timelapse experiments

The day prior to starting overnight timelapse experiment, cells were streaked from glcyerol stocks on selective plates and grown overnight at 37°C. In the morning, a colony was inoculated in 2mL LB media and incubated (37°C, shaking 250rpm) until mid-exponential ($OD_{600} \approx 0.8$). Cells were then spun down (1min, 2100rcf), LB poured off, resuspended in 1mL MSgg (made day of, see above), spun down again (1min, 2100rcf), and resuspend in 2mL of MSgg. Cells were further incubated (37°C, shaking 250rpm) to enculturate to the media (2hr, between 1 and 2 doublings). MSgg agarose pads were made by adding 0.15g of agarose powder to 10mL of MSgg media, microwaving the solution until the agarose dissolved (15-20s), and then pipetting the solution onto a standard 1" x 1" coverslip. A second coverslip was then placed on top. Pads were allowed to dry for at least 45 min before being sliced into quarters. After cells had enculturated to the MSgg, the OD_{600} value was taken and the culture was diluted to $OD_{600} = 10^{-4}$ using 1x PBS. $5\mu L$ of this diluted culture was then dropped on the 1/4 agarose pad and allowed to air dry for 10-15min. The dried pad was then inverted onto a glass imaging dish (WillCo) using a 1mL pipet tip. The dish was then stuffed with damp Kimwipes and sealed using parafilm and placed on the scope at temperature (30°C) for one hour prior to starting image acquisition.

Time-lapse Microscopy

All timelapse images were taken on a Nikon Ti2 Eclipse (Nikon) using a CFI Plan Apo DM Lambda 100X Oil objective (MRD31905, Nikon) equipped with an ORCA-Fusion BT CMOS camera (Hamamatsu Photonics) and illuminated with a SPECTRA Light Engine (Lumencor, Inc.). Temperature was maintained at 30°C using an oko Bold Line cage incubator (Okolab s.r.l.). For phase imaging, exposure was 20ms with 50% light power. For RFP imaging, the 594mm source from the SPECTRA was used with an MXR00254 dichroic and a Texas Red HYQ emission filter (608-683nm) with 80% laser power and a 200ms exposure. For GFP imaging, the 488nm light source from the Spectra was used with the same MXR00254 dichroic mirror and a FITC emission filter (515-555nm) with 80% laser power and a 200ms exposure. Unless otherwise stated, images were taken every 5 minutes during overnight timelapses. Timelapse experiments were acquired using Nikon Elements software (version 5.41.02, Nikon).

Image Analysis

Raw .nd2 files were cropped around microcolonies and trimmed just before the microcolony was disrupted by stacking of cells or swarming of motile cells. The cropped .nd2 files were then analyzed using the DeLTA 2.0 pipeline [153] on the Janelia Compute Cluster, which runs on Oracle Linux 8.3 and uses IBM Spectrum LSF 10.1 for job management. Cores were composed of either Intel SkyLake (2.7GHz Platinum

8168) or Intel Cascade Lake (3.0 GHz Gold 6284R) processors. GPUs were either GeForce RTX 2080Tis, Quadro RTX 8000s, Tesla T4 PCIe cards, Tesla V100 32GB PCIe cards, Tesla V100 SXM2s, or Tesla A100 SXM4s. Further analysis was done in Python 3.9.12 using Spyder 5.3.3 for an IDE and compiled using IPython 7.31.1. Packages used included NumPy 1.22.4 [83], SciPy 1.8.1 [84], and Matplotlib 3.5.2 [85].

Chapter 4

Conclusions

4.1 Summary

In **Chapter 2**, I used simulations of toy signaling circuits to develop some key rules around the impact of cell division on signaling states. I did this by examining whether nongenetic similarity existed in the concentrations of the different components of the toy signal pathway. In the Production Only pathway with a single enzyme producing a signal molecule, I found that nongenetic similarity existed only in the concentration of the signal molecule. That is, sister cells were more similar to each other than random cells in their concentrations of signal molecule, and not of the enzyme. I found this was due to the super-Poisson distribution of the signal molecule, stemming from the doubly stochastic process dictating its formation. Through a collaboration to analytically derive formulas for these variances (**Appendix A**), I found the key parameter separating the variance of the signal molecule in sisters from random pairs was the number of reactions per enzyme per cell cycle.

I then found that the duration of this nongenetic similarity was the same over all kinetic parameter combinations in the Production Only circuit, and followed a decay curve by one half each generation. This duration could be tuned down with the addition of a degrading enzyme to the circuit, due to the faster turnover of the signal molecule in the presence of a degrading enzyme. Lastly, I assessed the impact of these circuit architectures on population noise, finding that, at short nongenetic durations, the noise in the enzyme concentrations dictated noise levels in the signal molecule while at high durations of nongenetic similarity, the population had a similar level of noise regardless of the parameter combinations underlying the circuit. From this work, I showed that doubly stochastic processes can create nongenetic similarity that remains robust to noise introduced from cell division and that duration is tuned down as the turnover rate of the signal molecule increases.

After doing a theoretical examination of possible circuit architectures, I then examined a natural signaling circuit in the c-di-GMP pathway in *Bacillus subtilis.* I built a model simulating the production and degradation of the cyclic nucleotide with parameters based on values from literature or fitted to published data. This model predicted that, upon Spo0A repression of the *pdeH* promoter, accumulation of cyclic di-GMP to its high state would occur over six generations. This slow accumulation was due to a combination of diluting the highly active PDEs and the slow reaction rates of the DGCs.

To test the model predictions, I turned to developing a method for visualizing c-di-GMP dynamics in single cells. I first adapted the technique of agarose pad imaging for bacteria by creating pads with the biofilm-inducing media MSgg. I validated that this approach could image biofilm activation by showing activation of the matrix production phenotype using a matrix phenotype reporter strain. By segmenting single-cells using the DeLTA pipeline, I could track the proportions of cells that turned on, turned off, and did not activate the biofilm phenotype in growing microcolonies. Wit the approach validated, I then turned to evaluating tools for measuring intracellular c-di-GMP.

I began with implementing a split GFP-based sensor previously only used for *E. coli.* Static imaging of single cells showed significant differences between wild-type cells, a ΔDGC negative control strain, and a ΔPDE positive control strain. I then examined the readout of this sensor over time. I found the structure of the sensor resulted in a lack of signal decrease and concluded the sensor would not be adequate for validating the predictions of the model. Through another collaboration, I implemented a novel, circularly permuted GFP sensor for c-di-GMP. During validation of the sensor, I found that cell growth rate had a major impact on sensor readout and could not determine if the sensor was appropriately sensing intracellular c-di-GMP concentrations.

Lastly, in **Appendix B**, I performed back-of-the-envelope style calculations to evaluate if a hypothesis that ventral stress fiber formation during fibroblast to myofibroblast transition could be responsible for the creation of cytosolic pockets that could pool cytosolic fluorophores. I found that the energy required to bend the membrane and to contract the fibers fell well within the energy budget of a typical fibroblast cell over the time period the phenomenon was observed, validating the hypothesis.

4.2 Future Directions

In the era of the -omics revolution, our ability to identify biological components has accelerated while our understanding of these systems have lagged. New technologies capable of identifying all components of signaling circuits across an entire cell continue to develop and improve [168, 169]. At the same time, we have come into an appreciation that the identification of components in biological systems is insufficient to understanding how the system operates, as these static maps lack a sense of dynamical properties. While these dynamical properties are encoded into the architecture of biological circuits, their functions are often only elucidated through temporal approaches such as dynamic modeling and timelapse imaging. Thus, to advance our understanding of how biological systems operate, it is critical to continue to identify networks but then to study those networks from a temporal lens.

The study of bacterial biofilm development exemplifies the shortcomings of current approaches and the promise of new ones. Bacterial biofilms use complex, overlapping signal pathways to determine cell fate in a developing, multicellular system [117]. In *B. subtilis*, these developmental phenotypes have been shown to be tightly spatially and temporally regulated [170]. And while we know a lot of about how signaling circuits underlie cell fate decisions in individual cells [47, 46, 125], we lack a picture of how this operates on a multicellular level. Advances in 3D, timelapse imaging with the ability to track thousands of single cells [92] offers a potential solution to understand how single cell pathways and differentiation programs interact and whole biofilm developmental programs emerge. Such studies would advance our knowledge of systems as they operate, but would tell us less about unifying features or motifs in signal pathway design.

The use of theory, such as toy model simulations, is useful for identifying the simple rules governing signal pathway behavior in complex systems, and could be especially applied, as shown here, to the context of multigenerational processes, such as biofilm formation. The toy models in this work used assumptions about gene expression, but layering of more thermodynamic models could enhance their applicability to natural systems [171]. Additionally, the investigation of other circuit architectures, such as feed-forward loops [172], could make further predictions about the behavior of signal pathways in multigenerational contexts and the impact of circuit architecture on nongenetic inheritance and population-level noise. These models serve as a tool not only for identifying rules and motifs but also for exploring designs beyond the scope of what evolution has generated in biological systems.

4.3 Outlook

Furthering our understanding of biological systems as they have evolved to operate is critical to new advances in human health, but represents only a piece of potential future directions. New knowledge in the operation of signaling pathways has the potential to advance our ability to treat complex diseases such as cancer [173] and Alzheimer's [174]. But the work from the theory side also explores designs that have not been evolved by biological systems, opening up new opportunities in synthetic biological systems. Synthetic biologists have made advances in programming novel signaling pathways [175] and engineering complex behavior into synthetic multicellular systems [176]. We have previously suggested the engineering of biological noise for synthetic applications in the multicellular context as well [78]. These types of approaches can be used to engineer signal pathways to design smart, living drugs [177]. Strategies for engineering biology such as these illuminates horizons of technology beyond what has already evolved in the natural world.

In a way, this brings us full circle back to the idea introduced in **Chapter 1** of designing around the physical constraints of biological life. The example of simple gene activation illustrated how, by evolving new signal network architectures, bacterial systems could speed up or slow down the timing of gene expression. The constraint of the inherent noisiness of biochemical reactions was designed around by using the noisiness of circuits to drive phenotypic differentiation in a bet hedging strategy. Constraints on the noise introduced upon cell division have been designed around through the evolution of biological timer mechanisms. But evolution itself is a process defined by the physical constraints of genetic drift and natural selection; it is only capable of generating new solutions within the criteria of passing genetic information to future generations. The advent of synthetic biology then, the design of novel biological systems not selected for by nature, perhaps represents the ultimate design around physical constraints as we, human beings, ourselves are a product of the constraints of evolutionary design. How profound that, in the human engineering of biological life, evolution has yielded a biological system that can design around the constraints of its own process of selection.

Appendix A

Supplemental Information for Chapter 2

A.1 Acknowledgements

Analytical derivations were done by James Fitzgerald (Group Leader, Computation and Theory, Janelia Research Campus, HHMI).

A.2 Derivations of variance variables as functions of model parameters

Part 1: Effect of binomial partition on variance

Let X denote some molecular species (either enzyme M or signal molecule A). The initial number of molecules is a binomial partition of the number present before cell division. We thus write

$$P(X_0|X_T) = Bin(X_T, 1/2) \tag{A.1}$$

where X_T is the amount of molecule X in the mother cell and X_0 is the amount in a daughter cell. By direct calculation:

$$\langle X_0^2 \rangle_{X_0} = \langle X_0^2 \rangle_{X_0,X_T} \tag{A.2}$$

$$\langle X_0^2 \rangle_{X_0} = \langle \langle X_0^2 \rangle_{X_0|X_T} \rangle_{X_T} \tag{A.3}$$

$$\langle X_0^2 \rangle_{X_0} = \left\langle \frac{X_T}{4} + \left(\frac{X_T}{2} \right)^2 \right\rangle_{X_T} \tag{A.4}$$

$$\langle X_0^2 \rangle_{X_0} = \frac{1}{4}(\langle X_T \rangle_{X_T} + \langle X_T^2 \rangle_{X_T}) \tag{A.5}$$

note that:

$$\langle X_0 \rangle = \frac{1}{2}\langle X_T \rangle \tag{A.6}$$

Thus,

$$Var(X_0) = \langle X_0^2 \rangle_{X_0} - \langle X_0 \rangle_{X_0}^2 \tag{A.7}$$

$$Var(X_0) = \frac{1}{4}\left(\langle X_T^2 \rangle + \langle X_T \rangle\right) - \left(\frac{1}{2}\langle X_T \rangle \right)^2 \tag{A.8}$$

$$Var(X_0) = \frac{1}{4}(\langle X_T^2 \rangle - \langle X_T \rangle^2 + \langle X_T \rangle) \tag{A.9}$$

$$Var(X_0) = \frac{1}{4}(Var(X_T) + \langle X_T \rangle) \tag{A.10}$$

Part 2: Effect of Poisson Signal Accumulation

During cell growth, new molecules are added with some stochastic rate. By definition:

$$X_T = X_0 + \Delta X_T \tag{A.11}$$

The variance addition formula says

$$Var(X_T) = Var(X_0) + Var(\Delta X_T) + 2Cov(X_0, \Delta X_T) \tag{A.12}$$

Note that at steady state, we expect that both equations will hold, as the statistics of X_T are unchanging across generations.

Part 3: Statistics for the Variance of Enzyme M

For the enzyme M,

$$Cov(M_0, \Delta M_T) = 0 \tag{A.13}$$

Moreover, ΔM_T is generated by a Poisson process with rate $P_{prod,M}$. Therefore:

$$Var(\Delta M_T) = \langle \Delta M_T \rangle = P_{prod,M} T_{cc} \tag{A.14}$$

This implies (from Part 2) that,

$$Var(M_T) = Var(M_0) + P_{prod,M} T_{cc} \tag{A.15}$$

similarly (by Part 1),

$$\langle M_T \rangle = \langle M_0 \rangle + \langle \Delta M_T \rangle \tag{A.16}$$

$$\langle M_T \rangle = \langle M_0 \rangle + P_{prod,M} T_{cc} \tag{A.17}$$

$$\langle M_T \rangle = \frac{1}{2} \langle M_T \rangle + P_{prod,M} T_{cc} \tag{A.18}$$

$$\langle M_T \rangle = 2 P_{prod,M} T_{cc} \tag{A.19}$$

$$\langle M_0 \rangle = P_{prod,M} T_{cc} \tag{A.20}$$

Therefore (by Part 1),

$$Var(M_0) = \frac{1}{4}(Var(M_T) + 2P_{prod,M} T_{cc}) \tag{A.21}$$

$$Var(M_0) = \frac{1}{4}(Var(M_0) + P_{prod,M} T_{cc} + 2P_{prod,M} T_{cc}) \tag{A.22}$$

$$3Var(M_0) = 3P_{prod,M} T_{cc} \tag{A.23}$$

$$Var(M_0) = P_{prod,M} T_{cc} \tag{A.24}$$

$$Var(M_T) = 2P_{prod,M} T_{cc} \tag{A.25}$$

From these, we can calculate the variance of differences between our sister and random pairs. For the sister pairs

$$\Delta M_{sis} = M_{sis,1} - M_{sis,2} = M_0 - (M_T - M_0) = 2M_0 - M_T \tag{A.26}$$

Given $\langle \Delta M_{sis} \rangle = 0$, we get the variance as

$$Var(\Delta M_{sis}) = \langle \Delta M_{sis}^2 \rangle = 4\langle M_0^2 \rangle - 4\langle M_0 M_T \rangle + \langle M_T^2 \rangle \tag{A.27}$$

We first consider that

$$\langle M_0 M_T \rangle = \langle \langle M_0 \rangle_{M_0|M_T} M_T \rangle_{M_T} = \frac{1}{2} \langle M_T^2 \rangle_{M_T} \tag{A.28}$$

Therefore

$$Var(\Delta M_{sis}) = 4\langle M_0^2 \rangle - \langle M_T^2 \rangle \tag{A.29}$$

Plugging in our derived formulas, we find

$$Var(\Delta M_{sis}) = 4((P_{prod,M}T_{cc})^2 + P_{prod,M}T_{cc}) - ((2P_{prod,M}T_{cc})^2 + 2P_{prod,M}T_{cc}) \tag{A.30}$$

$$Var(\Delta M_{sis}) = 2P_{prod,M}T_{cc} \tag{A.31}$$

It follows that the variance of the two random cells would be two times the variance of each individual cell:

$$Var(\Delta M_{rnd}) = 2Var(M_0) \tag{A.32}$$

$$Var(\Delta M_{rnd}) = 2P_{prod,M}T_{cc} \tag{A.33}$$

Note that the variance of difference formulas for sister and random pairs are equivalent when considering M.

Part 4: Persistence of Poisson Statistics

Suppose

$$P(X_T) = Poiss(\bar{X}_T) \tag{A.34}$$

$$P(X_0|X_T) = Bin(X_T, 1/2) \tag{A.35}$$

Then X_0 ~Poiss($\bar{X}_T/2$). We proceed by direct calculation

$$P(X_0) = \Sigma_{X_T} P(X_0, X_T) \tag{A.36}$$

$$P(X_0) = \Sigma_{X_T} P(X_T) P(X_0|X_T) \tag{A.37}$$

$$P(X_0) = \Sigma_{X_T=X_0}^{\infty} \frac{(\bar{X}_T)^{X_T} e^{-\bar{X}_T}}{X_T!} \frac{X_T!(1/2)^{X_T}}{(X_T - X_0)!X_0!} \tag{A.38}$$

$$P(X_0) = \frac{e^{-\bar{X}_T}}{X_0!} \Sigma_{X_T=X_0}^{\infty} \frac{(\bar{X}_T/2)^{X_T}}{(X_T - X_0)!} \tag{A.39}$$

$$P(X_0) = \frac{e^{-\bar{X}_T}}{X_0!} \Sigma_{k=0}^{\infty} \frac{(\bar{X}_T/2)^{X_0+k}}{k!} \tag{A.40}$$

$$P(X_0) = \frac{e^{-\bar{X}_T}}{X_0!} (\frac{\bar{X}_T}{2})^{X_0} e^{\bar{X}_T/2} \tag{A.41}$$

$$P(X_0) = \frac{(\bar{X}_T/2)^{X_0} e^{-\bar{X}_T/2}}{X_0!} \tag{A.42}$$

$$P(X_0) = Poiss(\bar{X}_T/2) \tag{A.43}$$

Thus a Poisson random variable is still Poisson after it is partitioned binomially. As a corollary:

$$P(M_t) = Poiss(P_{prod,M} T_{cc} + P_{prod,M} t) \tag{A.44}$$

for all time points.

Part 5: Calculating the mean of the signal molecule A

We write

$$A_t = A_0 + \Delta A_t \tag{A.45}$$

where ΔA_t is generated from a double stochastic Poisson process with rate $k_{cat,M} M_t$. Thus

$$\langle A_t \rangle = \langle A_0 \rangle + \langle \Delta A_t \rangle \tag{A.46}$$

$$\langle A_t \rangle = \langle A_0 \rangle + k_{cat,M} \int_0^t dt' M_{t'} \tag{A.47}$$

$$\langle A_t \rangle = \langle A_0 \rangle + k_{cat,M} \int_0^t dt' (P_{prod,M} T_{cc} + P_{prod,M} t') \tag{A.48}$$

$$\langle A_t \rangle = \langle A_0 \rangle + k_{cat,M} P_{prod,M} T_{cc} t + \frac{1}{2} k_{cat,M} P_{prod,M} t^2 \tag{A.49}$$

setting $t = T_{cc}$ and noting that $\langle A_0 \rangle = \frac{1}{2} \langle A_T \rangle$, we get

$$\langle A_T \rangle = \left\langle \frac{A_T}{2} \right\rangle + \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.50}$$

$$\langle A_T \rangle = 3 k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.51}$$

$$\langle A_0 \rangle = \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.52}$$

$$\langle A_t \rangle = \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 + k_{cat,M} P_{prod,M} T_{cc} t + \frac{1}{2} k_{cat,M} P_{prod,M} t^2 \tag{A.53}$$

Part 6: Variance of accumulated signal molecule A

We next aim to calculate $Var(\Delta A_T)$. As a first step, we calculate

$$\langle \Delta A_T^2 \rangle = \langle \langle \Delta A_T^2 \rangle_{\Delta A_T | M_t} \rangle_{M_t} \tag{A.54}$$

Since ΔA_T is Poisson when conditioned on M_t, we have

$$\langle \Delta A_T^2 \rangle = \left\langle k_{cat,M} \int\limits_0^{T_{cc}} dt M_t + \left(k_{cat,M} \int\limits_0^{T_{cc}} dt M_t \right)^2 \right\rangle_{M_t} \tag{A.55}$$

$$\langle \Delta A_T^2 \rangle = k_{cat,M} \int\limits_0^{T_{cc}} dt \langle M_t \rangle_{M_t} + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{T_{cc}} dt' \langle M_t M_{t'} \rangle_{M_t} \tag{A.56}$$

Relabeling integrated variables such that $t \leq t'$, we can write

$$M_{t'} = M_t + (\Delta M)_{(t,t')} \tag{A.57}$$

where $(\Delta M)_{(t,t')}$ is the number of molecules of enzymes M added between t and t'. Therefore,

$$\langle \Delta A_T^2 \rangle = k_{cat,M} \int\limits_0^{T_{cc}} dt \langle M_t \rangle + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{t} dt' \langle M_t M_{t'} \rangle + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_t^{T_{cc}} dt' \langle M_t M_{t'} \rangle \tag{A.58}$$

$$\langle \Delta A_T^2 \rangle = k_{cat,M} \int\limits_0^{T_{cc}} dt \langle M_t \rangle + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{t} dt' \langle (M_{t'} + \Delta M_{(t',t)}) M_{t'} \rangle + \\ k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_t^{T_{cc}} dt' \langle M_t (M_t + \Delta M_{(t,t')}) \rangle \quad (A.59)$$

$$\langle \Delta A_T^2 \rangle = k_{cat,M} \int\limits_0^{T_{cc}} dt (P_{prod,M} T_{cc} + P_{prod,M} t) \\ + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{t} dt' (\langle M_{t'}^2 \rangle + \langle \Delta M_{(t',t)} \rangle \langle M_{t'} \rangle) \\ + k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_t^{T_{cc}} dt' (\langle M_t^2 \rangle + \langle M_t \rangle \langle \Delta M_{(t,t')} \rangle) \quad (A.60)$$

The first integral is

$$I_1 = k_{cat,M} \left(P_{prod,M} T_{cc}^2 + \frac{1}{2} P_{prod,M} T_{cc}^2 \right) \quad (A.61)$$

$$I_1 = \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \quad (A.62)$$

The second integral is

$$I_2 = k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{t} dt' (P_{prod,M} t' + P_{prod,M} T_{cc} + P_{prod,M}^2 (T_{cc} + t')^2 \\ + P_{prod,M} (t - t') P_{prod,M} (T_{cc} + t')) \quad (A.63)$$

$$I_2 = k_{cat,M}^2 \int\limits_0^{T_{cc}} dt \int\limits_0^{t} dt'(P_{prod,M}T_{cc} + P_{prod,M}^2 T_{cc}^2 + P_{prod,M}^2 t T_{cc} +$$
$$t'(P_{prod,M} + 2P_{prod,M}^2 T_{cc} + P_{prod,M}^2 t - P_{prod,M}^2 T_{cc}) + (t')^2(P_{prod,M}^2 - P_{prod,M}^2)) \quad \text{(A.64)}$$

$$I_2 = k_{cat,M}^2 \int\limits_0^{T_{cc}} dt(P_{prod,M}T_{cc}t + P_{prod,M}^2 T_{cc}^2 t + P_{prod,M}^2 T_{cc} t^2 + \frac{1}{2}P_{prod,M}t^2 +$$
$$P_{prod,M}^2 T_{cc} t^2 + \frac{1}{2}P_{prod,M}^2 t^3 - \frac{1}{2}P_{prod,M}^2 T_{cc} t^2) \quad \text{(A.65)}$$

$$I_2 = k_{cat,M}^2 \int\limits_0^{T_{cc}} dt(t(P_{prod,M}T_{cc} + P_{prod,M}^2 T_{cc}^2) +$$
$$t^2(\frac{3}{2}P_{prod,M}^2 T_{cc} + \frac{1}{2}P_{prod,M}) + t^3\frac{1}{2}P_{prod,M}^2) \quad \text{(A.66)}$$

$$I_2 = k_{cat,M}^2 \left(\frac{1}{2}P_{prod,M}T_{cc}^3 + \frac{1}{2}P_{prod,M}^2 T_{cc}^4 + \frac{1}{2}P_{prod,M}^2 T_{cc}^4 + \frac{1}{6}P_{prod,M}T_{cc}^3 + \frac{1}{8}P_{prod,M}^2 T_{cc}^4\right) \quad \text{(A.67)}$$

$$I_2 = k_{cat,M}^2 \left(\frac{2}{3}P_{prod,M}T_{cc}^3 + \frac{9}{8}P_{prod,M}^2 T_{cc}^4\right) \quad \text{(A.68)}$$

The third integral is

$$I_3 = k_{ca,M}^2 \int_0^{T_{cc}} dt \int_t^{T_{cc}} dt'(P_{prod,M}T_{Cc} + P_{prod,M}t$$
$$+ P_{prod,M}^2(T_{cc}+t)^2 + P_{prod,M}(T_{cc}+t)P_{prod,M}(t'-t) \quad \text{(A.69)}$$

$$I_3 = k_{ca,M}^2 \int_0^{T_{cc}} dt \int_t^{T_{cc}} dt'(P_{prod,M}(T_{cc}+t)$$
$$+ P_{prod,M}^2(T_{cc}+t)^2 - P_{prod,M}^2(T_{cc}+t)t + P_{prod,M}^2(T_{cc}+t)t') \quad \text{(A.70)}$$

$$I_3 = k_{ca,M}^2 \int_0^{T_{cc}} dt((P_{prod,M}(T_{cc}+t)(T_{cc}-t) + P_{prod,M}^2(T_{cc}+t)T_{cc}(T_{cc}-t)+$$
$$\frac{1}{2}P_{prod,M}^2(T_{cc}+t)T_{cc}^2 - \frac{1}{2}P_{prod,M}^2(T_{cc}+t)t^2) \quad \text{(A.71)}$$

$$I_3 = k_{ca,M}^2 \int_0^{T_{cc}} dt(P_{prod,M}T_{cc}^2 + \frac{3}{2}P_{prod,M}^2T_{cc}^3 + t(\frac{1}{2}P_{prod,M}^2T_{cc}^2)+$$
$$t^2(-Pprod,M - P_{prod,M}^2T_{cc} - \frac{1}{2}P_{prod,M}^2T_{cc}) + t^3(-\frac{1}{2}P_{prod,M}^2)) \quad \text{(A.72)}$$

$$I_3 = k_{ca,M}^2 (P_{prod,M} T_{cc}^3 \frac{3}{2} P_{prod,M}^2 T_{cc}^4 + \frac{1}{4} P_{prod,M}^2 T_{cc}^4 - \\ \frac{1}{3} P_{prod,M} T_{cc}^3 - \frac{1}{2} P_{prod,M}^2 T_{cc}^4 - \frac{1}{8} P_{prod,M}^2 T_{cc}^4) \quad (A.73)$$

$$I_3 = k_{cat,M}^2 \left(\frac{2}{3} P_{prod,M} T_{cc}^3 + \frac{9}{8} P_{prod,M}^2 T_{cc}^4 \right) \quad (A.74)$$

Note that, as expected, $I_2 = I_3$. Putting these terms together

$$\langle \Delta A_T^2 \rangle = \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{4}{3} k_{cat,M}^2 P_{prod,M} T_{cc}^3 + \frac{9}{4} k_{cat,M}^2 P_{prod,M}^2 T_{cc}^4 \quad (A.75)$$

$$Var(\Delta A_T) = \langle \Delta A_T^2 \rangle - \left(\frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \right)^2 \quad (A.76)$$

$$Var(\Delta A_T) = \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{4}{3} k_{cat,M}^2 P_{prod,M} T_{cc}^3 \quad (A.77)$$

Part 7: Signal molecule covariance statistics

The final quantity to compute is $Cov(A_0, \Delta A_T)$. This is nonzero because both are correlated with the amount of enzyme in the mother cell. However, they become independent when conditioned on M_0:

$$P(A_0, \Delta A_T) = \Sigma_{M_0} P(A_0, \Delta A_T, M_0) \tag{A.78}$$

$$P(A_0, \Delta A_T) = \Sigma_{M_0} P(M_0) P(A_0, \Delta A_T | M_0) \tag{A.79}$$

$$P(A_0, \Delta A_T) = \Sigma_{M_0} P(M_0) P(A_0 | M_0) P(\Delta A_T | M_0) \tag{A.80}$$

This implies that

$$Cov(A_0, \Delta A_T) = \langle \langle \delta A_0 \rangle_{A_0|M_0} \langle \delta \Delta A_T \rangle_{\Delta A_T | M_0} \rangle_{M_0} \tag{A.81}$$

note that

$$P(\Delta A_T | M_t) = Poiss \left(k_{cat,M} \int\limits_0^{T_{cc}} dt M_t \right) \tag{A.82}$$

$$P(\Delta A_T | M_t) = Poiss \left(k_{cat,M} \int\limits_0^{T_{cc}} dt (M_0 + \Delta M_t) \right) \tag{A.83}$$

$$P(\Delta A_T | M_0) = \Sigma_{M_t > 0} P(\Delta A_T, M_{t>0} | M_0) \tag{A.84}$$

$$P(\Delta A_T | M_0) = \Sigma_{M_t > 0} P(M_{t>0} | M_0) \tag{A.85}$$

$$P(\Delta A_T | M_t) = \Sigma_{\Delta M_t} P(\Delta M_t) P(\Delta A_T | M_0, \Delta M_t) \tag{A.86}$$

we then get

$$\langle \delta \Delta A_T \rangle_{\Delta A_T|M_0} = k_{cat,M} \int_0^{T_{cc}} dt \left((M_0 + \langle \Delta M_t \rangle) - \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \right) \tag{A.87}$$

$$\langle \delta \Delta A_T \rangle_{\Delta A_T|M_0} = k_{cat,M} M_0 T_{cc} + k_{cat,M} \int_0^{T_{cc}} dt \left(P_{prod,M} t - \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \right) \tag{A.88}$$

$$\langle \delta \Delta A_T \rangle_{\Delta A_T|M_0} = k_{cat,M} M_0 T_{cc} + \frac{1}{2} k_{cat,M} P_{prod,M} T_{cc}^2 - \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.89}$$

$$\langle \delta \Delta A_T \rangle_{\Delta A_T|M_0} = k_{cat,M} M_0 T_{cc} - k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.90}$$

To compute $\langle A_0 \rangle_{A_0|M_\infty}$, we note linear relationship between the A and M variables, which translates into a linear relationship between the means

$$\langle A_0 \rangle_{A_0|M_0} = \frac{1}{2} \langle A_T \rangle_{A_T|M_0} \tag{A.91}$$

$$\langle M_T \rangle_{M_T|M_0} = 2M_0 \tag{A.92}$$

$$\langle M_t \rangle_{M_t|M_0} = \langle M_T - \Delta M_{(t,T)} \rangle_{M_T,\Delta M|M_0} \tag{A.93}$$

$$\langle M_t \rangle_{M_t|M_0} = 2M_0 - P_{prod,M}(T_{cc} - t) \tag{A.94}$$

$$\langle A_t \rangle_{A_t|M_0} = k_{cat,M} \int\limits_0^t dt' \langle M_{t'} \rangle_{M_t|M_0} \tag{A.95}$$

$$\langle A_t \rangle_{A_t|M_0} = k_{cat,M} \int\limits_0^t dt' (2M_0 - P_{prod,M} T_{cc} + P_{prod,M} t') \tag{A.96}$$

$$\langle A_t \rangle_{A_t|M_0} = 2k_{cat,M} M_0 t - k_{cat,M} P_{prod,M} T_{cc} t + \frac{1}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.97}$$

$$\langle A_T \rangle_{A_T|M_0} = 2k_{cat,M} M_0 T_{cc} - \frac{1}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.98}$$

$$\langle A_0 \rangle_{A_0|M_0} = k_{cat,M} M_0 T_{cc} - \frac{1}{4} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.99}$$

$$\langle \delta A_0 \rangle_{A_0|M_0} = \langle A_0 \rangle_{A_0|M} - \frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.100}$$

$$\langle \delta A_0 \rangle_{A_0|M_0} = k_{cat,M} M_0 T_{cc} - \frac{7}{4} k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.101}$$

Putting the pieces together

$$\begin{aligned} Cov(A_0, \Delta A_T) = \Bigg\langle \bigg(& k_{cat,M} M_0 T_{cc} - \frac{7}{4} k_{cat,M} P_{prod,M} T_{cc}^2 \bigg) \\ & \Big(k_{cat,M} M_0 T_{cc} - k_{cat,M} P_{prod,M} T_{cc}^2 \Big) \Bigg\rangle_{M_0} \end{aligned} \tag{A.102}$$

$$Cov(A_0, \Delta A_T) = \frac{7}{4}k_{cat,M}^2 P_{prod,M}^2 T_{cc}^4 - k_{cat,M}^2 P_{prod,M} T_{cc}^3 \langle M_0 \rangle - \frac{7}{4}k_{cat,M}^2 P_{prod,M} T_{cc}^3 \langle M_0 \rangle + k_{cat,M}^2 T_{cc}^2 \langle M_0^2 \rangle \quad (A.103)$$

$$Cov(A_0, \Delta A_T) = k_{cat,M}^2 T_{cc}^2 (P_{prod,M} T_{cc} + P_{prod,M}^2 T_{cc}^2) - k_{cat,M}^2 P_{prod,M}^2 T_{cc}^4 \quad (A.104)$$

$$Cov(A_0, \Delta A_T) = k_{cat,M}^2 P_{prod,M} T_{cc}^3 \quad (A.105)$$

Part 8: Putting it all together

From Part 1,

$$Var(A_0) = \frac{1}{4}(Var(A_T) + \langle A_T \rangle) \quad (A.106)$$

$$Var(A_0) = \frac{1}{4}(Var(A_T) + 3k_{cat,M} P_{prod,M} T_{cc}^2) \quad (A.107)$$

Substituting this into the equation from Part 2,

$$Var(A_T) = \frac{1}{4}(Var(A_T) + 3k_{cat,M} P_{prod,M} T_{cc}^2) + \frac{3}{2}k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{4}{3}k_{cat,M}^2 P_{prod,M} T_{cc}^3 + 2k_{cat,M}^2 P_{prod,M} T_{cc}^3 \quad (A.108)$$

$$\frac{3}{4}Var(A_T) = \frac{9}{4}k_{cat,M}P_{prod,M}T_{cc}^2 + \frac{10}{3}k_{cat,M}^2P_{prod,M}T_{cc}^3 \tag{A.109}$$

$$Var(A_T) = 3k_{cat,M}P_{prod,M}T_{cc}^2 + \frac{40}{9}k_{cat,M}^2P_{prod,M}T_{cc}^3 \tag{A.110}$$

This implies, by direct substitution, that

$$Var(A_0) = \frac{1}{4}(3k_{cat,M}P_{prod,M}T_{cc}^2 + \frac{40}{9}k_{cat,M}^2P_{prod,M}T_{cc}^3 + 3k_{cat,M}P_{prod,M}T_{cc}^2) \tag{A.111}$$

$$Var(A_0) = \frac{3}{2}k_{cat,M}P_{prod,M}T_{cc}^2 + \frac{10}{9}k_{cat,M}^2P_{prod,M}T_{cc}^3 \tag{A.112}$$

Note that each variance expression looks like a Poisson variance plus an excess variance that can be attributed to the doubly stochastic part.

Part 9: Sister cell variance

Finally, we calculate the variance of the difference between sister cells

$$\Delta A_{sis} = A_{sis,1} - A_{sis,2} = A_0 - (A_T - A_0) = 2A_0 - A_T \tag{A.113}$$

Clearly

$$\langle \Delta A_{sis} \rangle = 0 \tag{A.114}$$

Therefore

$$Var(\Delta A_{sis}) = \langle \Delta A_{sis}^2 \rangle = 4\langle A_0^2 \rangle - 4\langle A_0 A_T \rangle + \langle A_T^2 \rangle \tag{A.115}$$

The piece of this we haven't considered yet is

$$\langle A_0 A_T \rangle = \langle \langle A_0 \rangle_{A_0|A_T} A_T \rangle_{A_T} = \frac{1}{2} \langle A_T^2 \rangle_{A_T} \tag{A.116}$$

Therefore

$$Var(\Delta A_{sis}) = 4\langle A_0^2 \rangle - \langle A_T^2 \rangle \tag{A.117}$$

Plugging in the formulas from Part 8, we find

$$\begin{aligned} Var(\Delta A_{sis}) = 4((&\frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2)^2 + \\ &\frac{3}{2}(k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{10}{9} k_{cat,M}^2 P_{prod,M} T_{cc}^3)) \\ -((&\frac{3}{2} k_{cat,M} P_{prod,M} T_{cc}^2)^2 + \frac{3}{2}(k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{10}{9} k_{cat,M}^2 P_{prod,M} T_{cc}^3)) \end{aligned} \tag{A.118}$$

$$Var(\Delta A_{sis}) = 3 k_{cat,M} P_{prod,M} T_{cc}^2 \tag{A.119}$$

Part 10: Random pair variance

It follows that the variance of two random cells would be two times the variance of each individual cell

$$Var(\Delta A_{rnd}) = 2 Var(A_0) \tag{A.120}$$

$$Var(\Delta A_{rnd}) = 3 k_{cat,M} P_{prod,M} T_{cc}^2 + \frac{20}{9} k_{cat,M}^2 P_{prod,M} T_{cc}^3 \tag{A.121}$$

A.3 Supplemental Figures

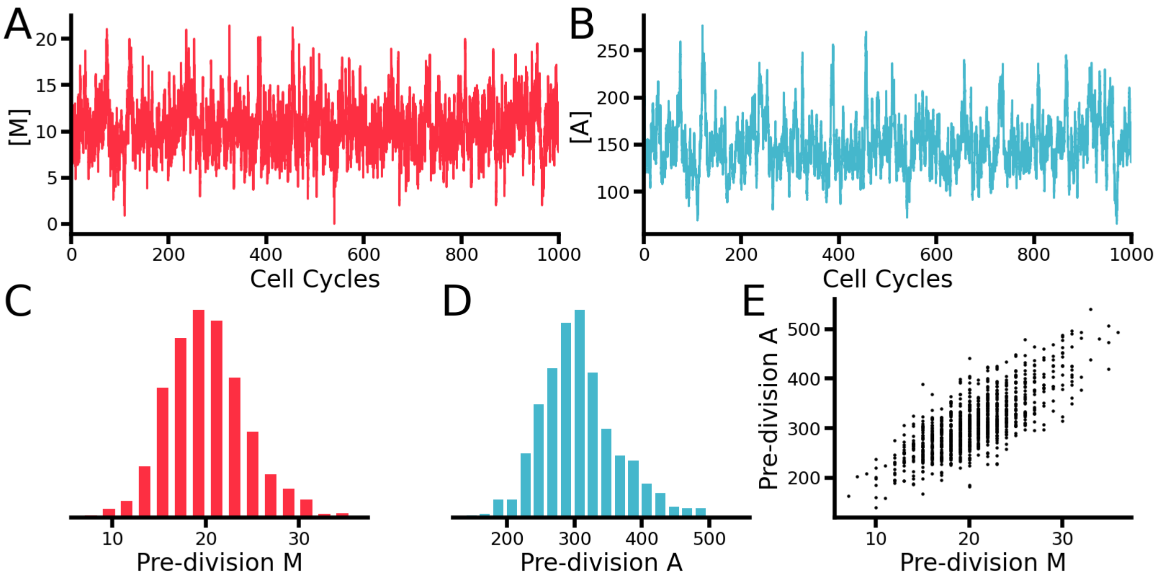

Figure A·1: Initial model time trace — Our initial parameter set, as inspired by the cyclic di-GMP circuit in *B. subtilis* ($k_{cat,M} = 10^{-2}, P_{prod,M} = 10^{-2}, T_{cc} = 1000$), was run for 1000 cell cycles. **(A)** Concentration trace of the enzyme M over the 1000 cell cycles. The value stochastically hovers around the mean ($\mu_{[M]} = P_{prod,M}T_{cc}$). **(B)** Concentration trace of the signal molecule A over the 1000 simulated cell cycles. The value of $[A]$ stochastically fluctuates around the mean ($\mu_{[A]} = 3/2k_{cat,M}P_{prod,M}T_{cc}^2$). We then identified the timepoints prior to cell division (when $V > 2$) and pulled out the counts of the enzyme M **(C)** and signal molecule A **(D)**. Plotting these values against each other for each timepoint reveals a cross correlation between enzyme count and signal molecule **(E)**.

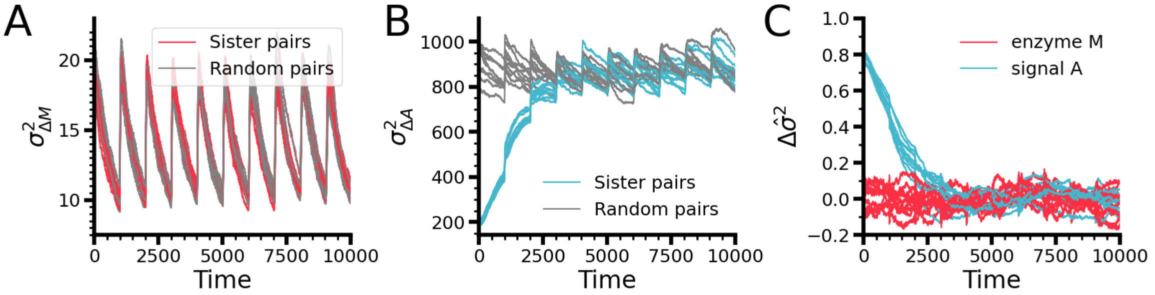

Figure A·2: Signal molecule memory result robust to random seed of model — Simulations were started with the typical random seed for Numpy's generator (seed=1000), but run successively ten times, ensuring each iteration used a different sequence of random numbers. **(A)** Variance of pairwise differences for enzyme M. Across all ten computational repeats, the results had the same behavior as Fig 2·2B, with the sawtooth pattern in the variance of both the enzyme and signal molecule pairs. **(B)** Variance of pairwise differences for the signal molecule A. Similarly, across all ten repetitions, the same behavior of the variance of the sister pairs starting at a low value before increasing to meet the constant variance of the random pairs is seen. **(C)** Normalized difference of variances of pairwise differences for enzyme M (red lines) and signal molecule A (blue lines). As in Fig 2·2D, the memory exists only in the signal molecule across all ten sequences of random numbers.

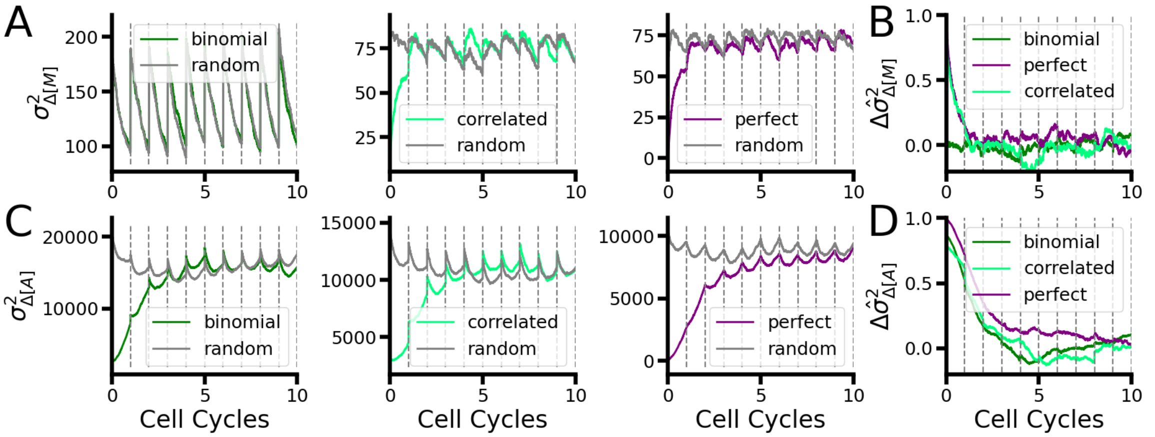

Figure A·3: Nongenetic memory in signal molecule occurs in all partition models — Along with the binomial partition model used in the main text, two other partition models were tested: a correlated binomial partition model, in which the random number pull for partitioning of the enzyme M and signal molecule A was the same, and a perfect partition model, in which exactly half of each molecule were allotted to each daughter cell. **(A)** Variance of pairwise differences in enzyme M for binomial (left), correlated binomial (center) and perfect (right) partition models. While binomial partitioning generated large variance between cells, correlated partition models and perfect partition had one generation of memory between sister cells before reverting to the level of variance of the random pairs (gray lines). **(B)** Normalized difference of variances for pairwise differences of enzyme M in the three partition models. Here, the one generation of memory in the perfect and correlated partition models is evident, while no memory is present in the binomial partitioning. **(C)** Variances of pairwise differences in the signal molecule A for the three partition models. Across all cases, the sister cells (colored lines) have a few generations of memory while the random pairs have a consistent level of variance. **(D)** Normalized difference of variances for pairwise differences in the signal molecule A across the three partition models. Despite the three different models used, the duration of nongenetic memory in the signal molecule remains similar. n=1000 pairs for each simulation.

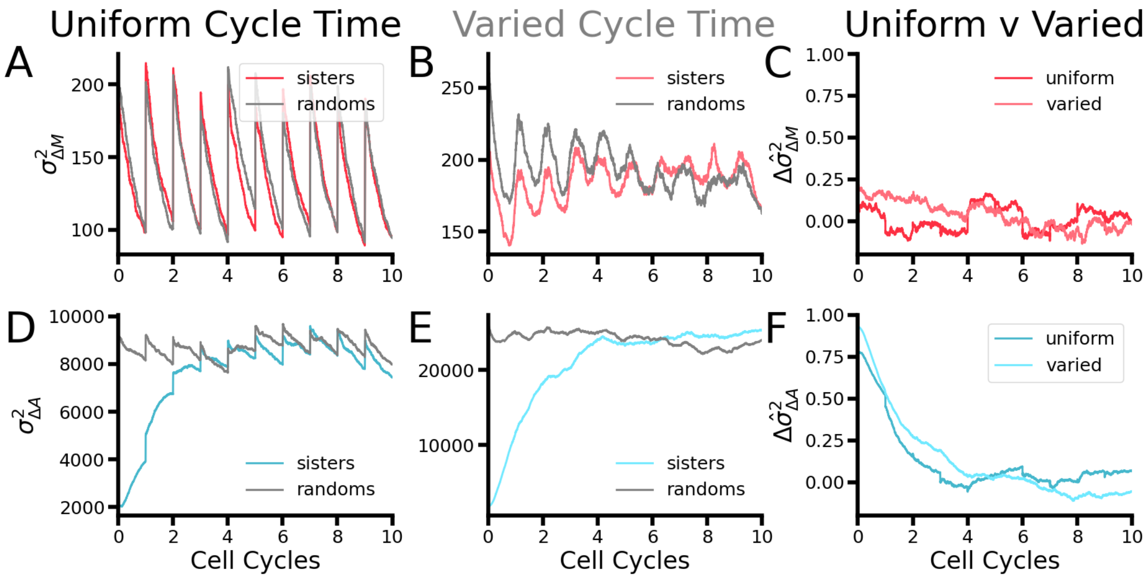

Figure A·4: Memory encoded in signal molecule concentration occurs under variable cell cycle time — To see if the fixed cell cycle time was necessary for memory to occur in the signaling molecule concentration, we created a version of the simulation in which the cell cycle time for each cycle was chosen from a normal distribution centered with a mean of T_{cc}. We then performed our same variance of pairwise difference analysis for the enzyme M and signal molecule A. While the varied cell cycle time **(B)** smoothed out the sawtooth pattern seen in the uniform cycle time **(A)**, the was similarly no memory held in the concentration of the M enzyme **(C)**. The same general trend occurred in the signal molecule A, with the varied cell cycle time **(E)** smoothing out the variance curves from the uniform cycle time **(D)** but not changing the duration of the memory **(F)**. n=1000 pairs for each trace.

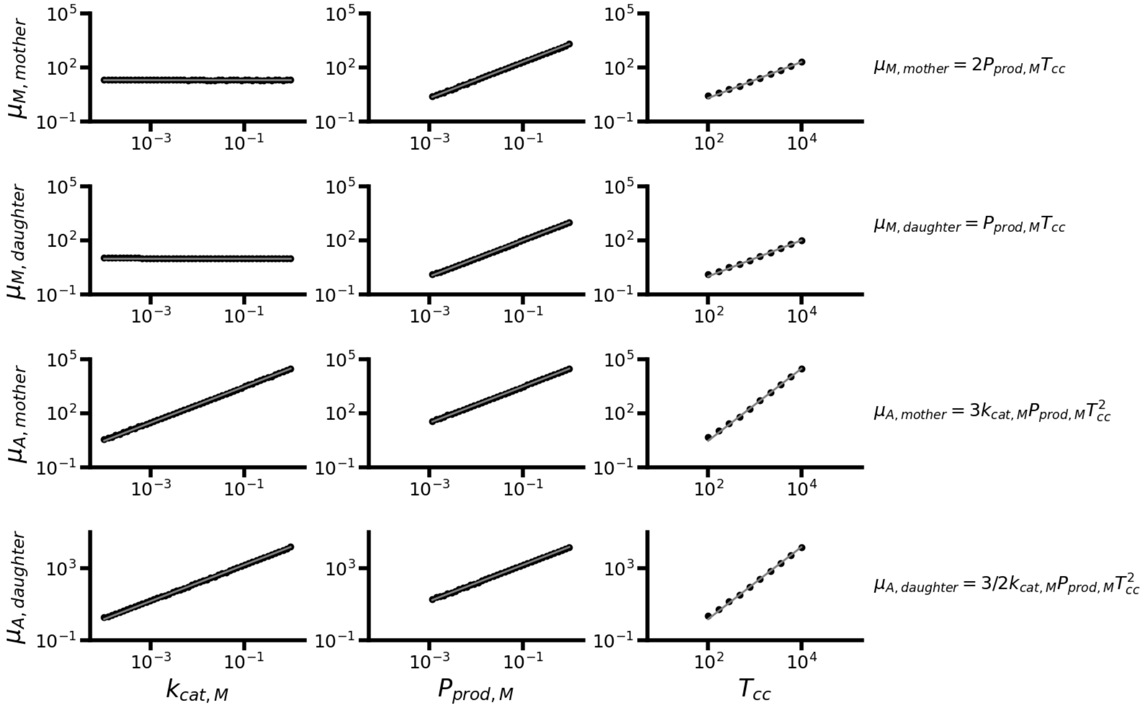

Figure A·5: Comparison of analytically-derived and numerical simulations for variable means across parameter sweeps — To check the validity of the analytically-derived equations, we ran simulations along sweeps of the parameters in the Production Only circuit, calculating the mean values for the amount of enzyme M and signal molecule A in the mother and daughter cells (dots). These values were compared to the predictions from the analytically-derived formulas (grey lines). Across all variables and parameter sweeps, the analytical models concurred with the numerical simulations of the data. n=1000 cells for each mean calculation.

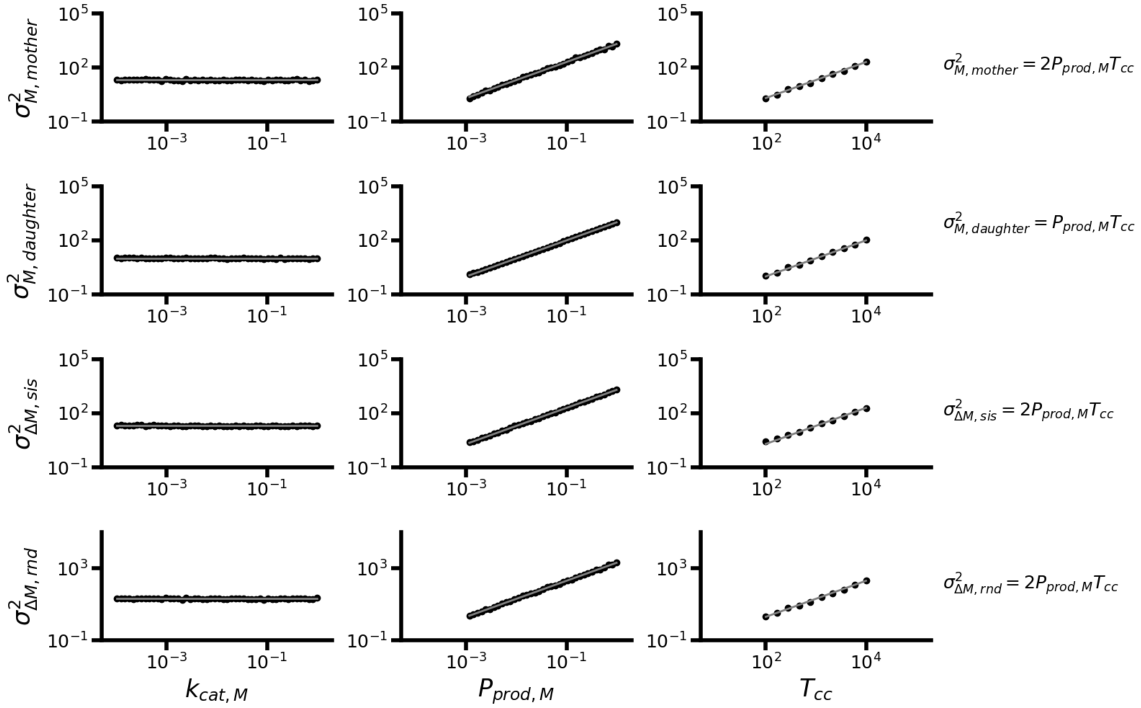

Figure A·6: Comparison of analytically-derived and numerical simulations for variance of M-related variables across parameter sweeps — To check the validity of the analytically-derived equations, we ran simulations along sweeps of the parameters in the Production Only circuit, calculating the variance of values amount of enzyme M in mother cells (top row), enzyme in daughter cells (second row), difference between sister cells (third row), and difference between random cells (bottom row). These values were compared to the predictions from the analytically-derived formulas (grey lines). Across all variables and parameter sweeps, the analytical models concurred with the numerical simulations of the data. n=1000 cells for each mean calculation.

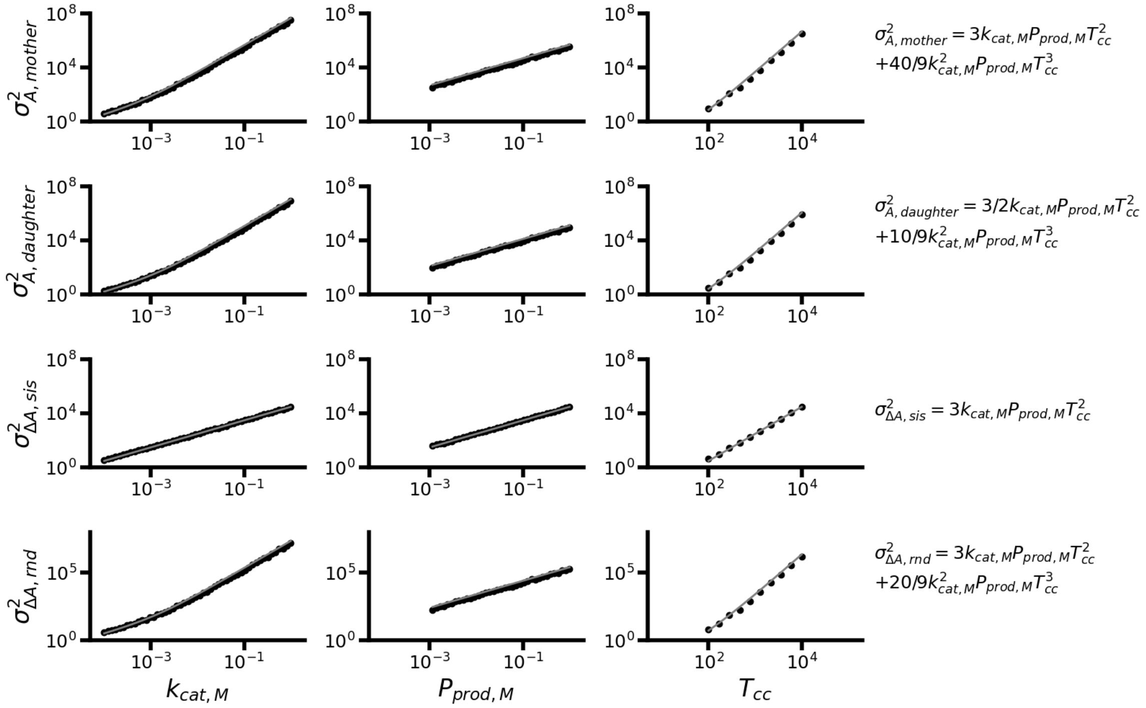

Figure A·7: Comparison of analytically-derived and numerical simulations for variance of A-related variables across parameter sweeps — To check the validity of the analytically-derived equations, we ran simulations along sweeps of the parameters in the Production Only circuit, calculating the variance of values amount of signal molecule A in mother cells (top row), enzyme in daughter cells (second row), difference between sister cells (third row), and difference between random cells (bottom row). These values were compared to the predictions from the analytically-derived formulas (grey lines). Across all variables and parameter sweeps, the analytical models concurred with the numerical simulations of the data. n=1000 cells for each mean calculation.

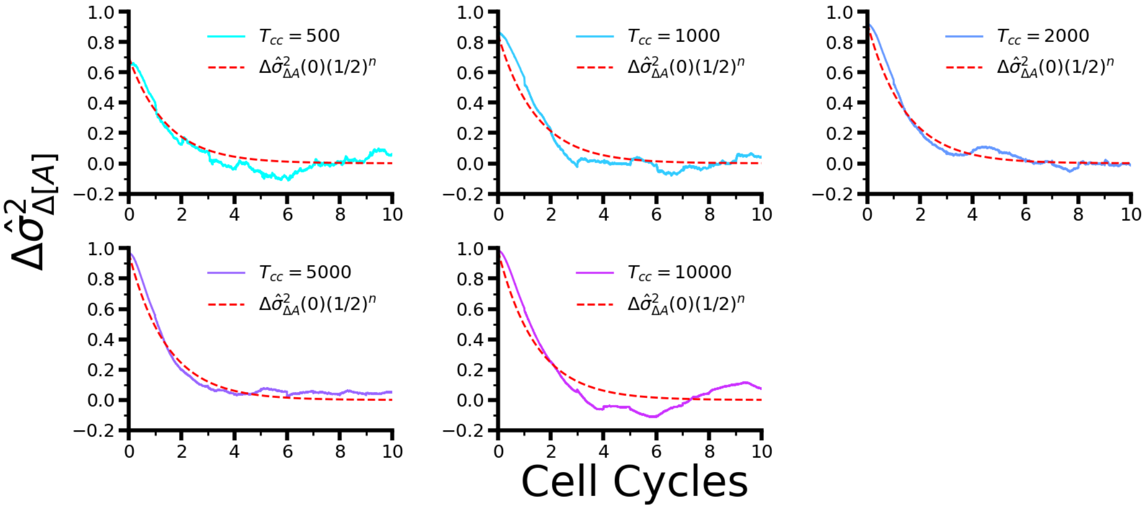

Figure A·8: Geometric decay fits for all tested cycle times Full data set for model fits seen in Fig 2·5C. For each cell cycle time (T_{cc}), 1000 pairs of sister and random cells were simulated for 10 generations and the normalized difference of variance metric was calculated for the time course (Equation 2.8) and plotted as a function of cell generations (blue-colored lines). Additionally, the conceptual model described in Fig 2·5A was applied using Equation 2.22 for each value of T_{cc} (red dashed lines). Across the entire tested range of T_{cc} values, the conceptual model fits well with the data, suggesting that geometric decay by cell cycle sets the limit on the nongenetic memory described here.

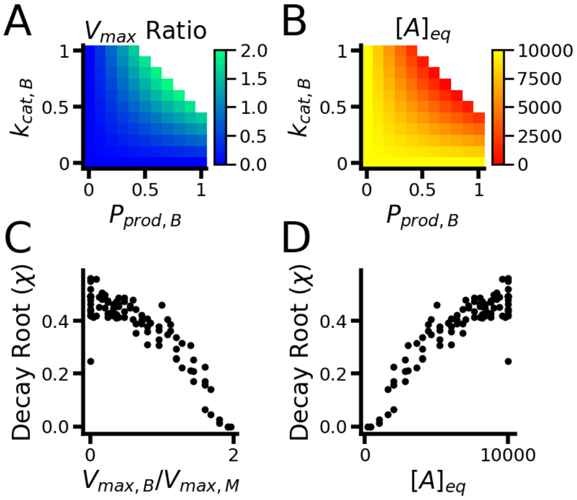

Figure A·9: Sweep of breaker enzyme parameters alone entangles nongenetic similarity duration with concentration of signal molecule — (A) Heatmap of V_{max} ratio values when breaker enzyme B parameters (maximal reaction rate $k_{cat,B}$ and production rate $P_{prod,B}$) were swept over a linear range. As either parameter increases, the overall V_{max} ratio increases until its limit at 2 (over which the breaker is too active and degrades all signal molecule in the system). **(B)** Heatmap of equilibrium concentration of signal molecule A as a function of breaker enzyme B parameter sweep. When only the parameters of breaker enzyme B are increased, the equilibrium concentration of A decreases until it reaches 0. **(C)** Nongenetic similarity duration as a function of the V_{max} ratio when only the parameters for breaker enzyme B are altered. As the V_{max} ratio increases, the value of the decay constant (χ) for the nongenetic similarity of signal molecule A decreases. **(D)** Nongenetic similarity duration as a function of the equilibrium concentration of signal molecule A. In this parameter sweep, the decrease in nongenetic memory duration (as represented by the decay root χ) was entangled with the decrease in the concentration of signal molecule A. To untangle these, a sweep was performed altering parameters in both maker enzyme M and breaker enzyme B (Fig SA·10). n=1000 pairs of sister cells and 1000 pairs of random cells for each trace fitted to a value of χ in **(C)** and **(D)**.

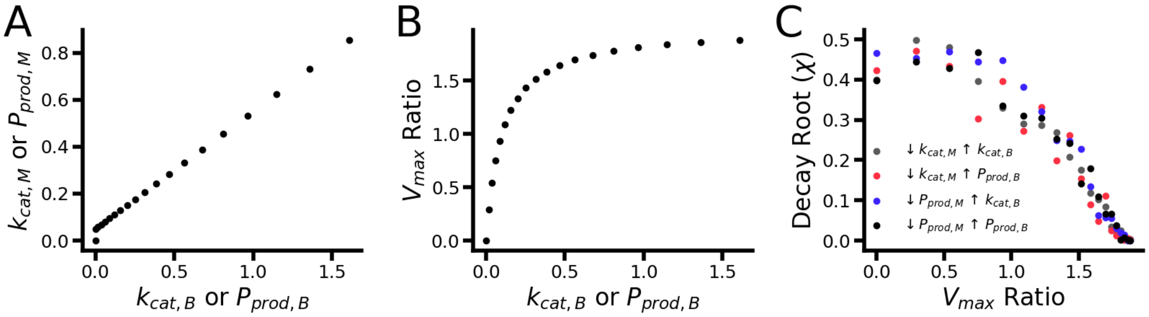

Figure A·10: Sweep of V_{max} ratio while maintaining concentration of signal molecule — (A) Function describing relationship between parameters for maker enzyme M ($k_{cat,M}$ and $P_{prod,M}$, y-axis) and parameters for breaker enzyme B ($k_{cat,B}$ and $P_{prod,B}$, x-axis). Both parameters were altered in order to maintain a constant equilibrium concentration of the signal molecule A at 5 μM. **(B)** Relationship between parameters for breaker enzyme B ($k_{cat,B}$ and $P_{prod,B}$, x-axis) and the corresponding V_{max} ratio (y-axis). Altering parameters for both the maker enzyme M and breaker enzyme B enabled a sweep over the full range of $0 < V_{max} < 2$ while maintaining a constant equilibrium concentration of signal molecule A. **(C)** Decay root of nongenetic similarity of signal molecule A as a function of the V_{max} ratio. Across all four combinations of parameters, an inverse relationship exists between the nongenetic similarity duration (as quantified by the decay root χ) and the V_{max} ratio. n = 1000 pairs of sister and 1000 pairs of random cells for each calculation of the decay root.

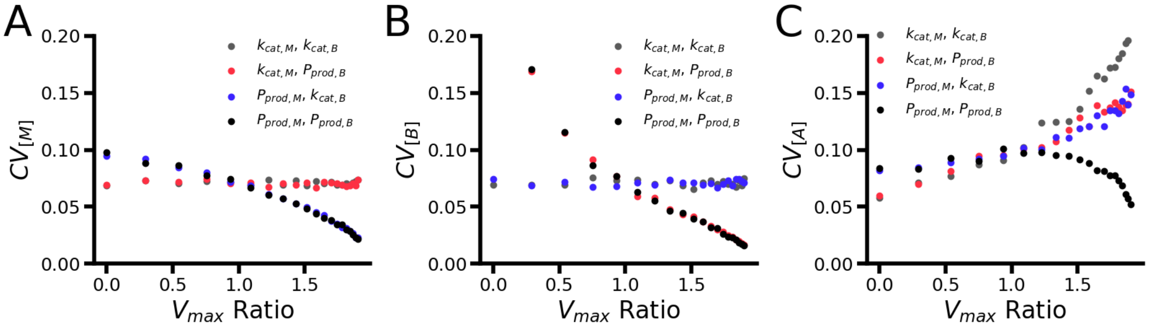

Figure A·11: CV values of enzymes and signal molecule A over V_{max} ratio sweep — **(A)** Coefficient of variation of maker enzyme M over V_{max} ratio sweep. As parameters are altered to change the V_{max} ratio while maintaining a constant M_{eq}, the $CV_{[M]}$ values changes only under the conditions when the $P_{prod,M}$ value is included in the pairs of parameters being altered (blue and black dots), as opposed to conditions where $k_{cat,M}$ is being altered (red and grey dots). The high $CV_{[M]}$ value of the $P_{prod,M}$ conditions at low V_{max} ratio values contributes to the high $CV_{[A]}$ values at low V_{max} ratio values of those conditions (blue and black dots). **(B)** Coefficient of variation of breaker enzyme B over V_{max} ratio sweep. As parameters are altered to change the V_{max} ratio while maintaining a constant M_{eq}, the $CV_{[B]}$ values change only under the conditions when the $P_{prod,B}$ value is included in the pairs of parameters being altered (red and black dots), as opposed to conditions where $k_{cat,B}$ is being altered (blue and grey dots). **(C)** Coefficient of variation of signal molecule A over V_{max} ratio sweep. In general, as the V_{max} ratio increases, the $CV_{[A]}$ value also increases. The exception to this is the scenario in which both P_{prod} parameters were altered to achieve the V_{max} ratio sweep (black dots). For this condition, at the upper range of the V_{max} ratio sweep, the $CV_{[A]}$ values decreases. This is due to the decrease in the noise contributions from each enzyme in this part of the parameter sweep (small values of $CV_{[M]}$ in **(A)**, black dots, and small values of $CV[B]$ in **(B)**, black dots). CV values were calculated from n=1000 cells for each parameter combination.

Appendix B

Quantitative Hypothesis Testing of Stress Fiber-Induced Plasma Membrane Deformation

B.1 Disclosure and Copyright Statement

This chapter adapts portions of “Ventral stress fibers induce plasma membrane deformation in human fibroblasts” by Samuel J. Ghilardi, Mark S. Aronson, and Allyson E. Sgro, 2021 Molecular Biology of the Cell. 2021; 32 (18). © The Authors

B.2 Introduction

This section describes some quantitative hypothesis testing I did in collaboration with Sam Ghilardi. During induction of myofibroblast transition, it was observed that cytosolic fluorphores illuminated long, fiber-like structures in the cells (Fig B·1).

It was found that these fluorescent structures were colocalized to ventral actin stress fibers (Fig B·2).

With these data, we generated a hypothesis that contraction of the ventral stress fibers during the fibroblast-to-myofibroblast transition created pockets that the cytosolic fluorophores could flow into, creating the fluorescent structures (Fig B·3).

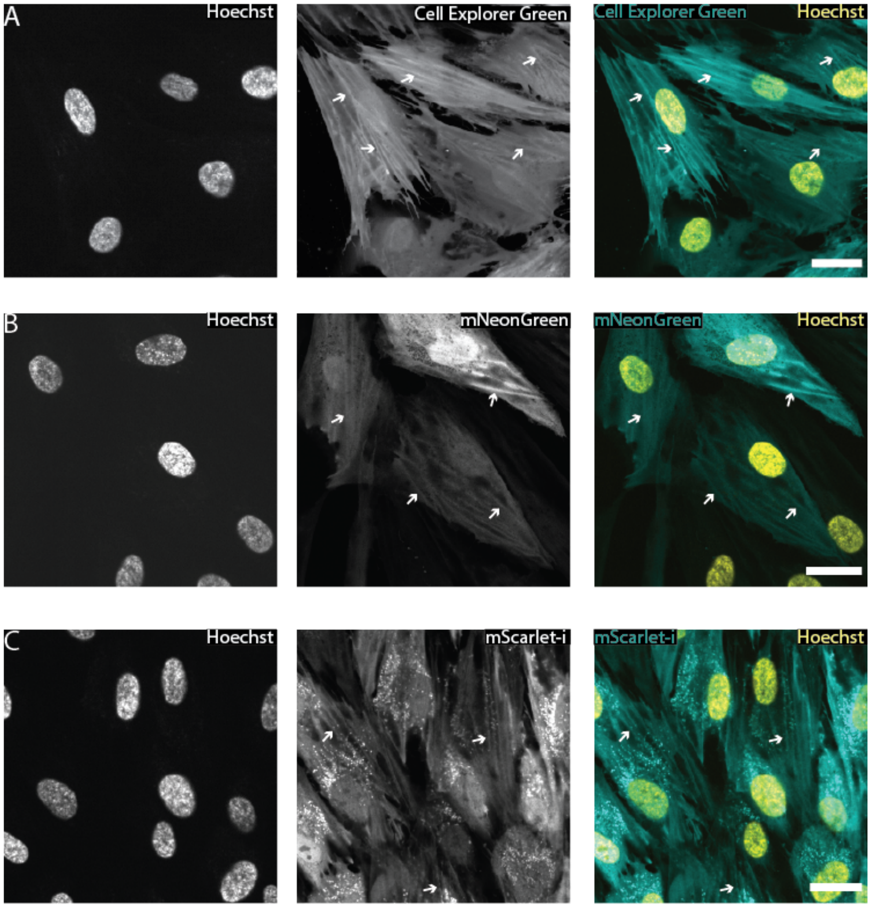
A
Hoechst
Cell Explorer Green
Cell Explorer Green
Hoechst
B
Hoechst
mNeonGreen
mNeonGreen
Hoechst
C
Hoechst
mScarlet-i
mScarlet-i
Hoechst

Figure B·1: Fluorescent Structures are Visible in Human Dermal Myofibroblasts Loaded with a Cytosolic Fluorophore After 96 hours of TGF-β1 treatment, Human Dermal Fibroblasts transition into myofibroblasts and separately develop fluorescent structures on the ventral side of the cell (examples marked by white arrows). These ridges can be observed in naive cells labeled with **(A)** cell permeable dye, or cells expressing fluorescent proteins such as **(B)** mNeonGreen or **(C)** mScarlet-i. Note that, at this magnification, fluorescent puncta can be seen in cells expressing either mScarlet-i, but not mNeonGreen. There is also some visible bleedthrough from the blue (Hoechst) channel into the green (cell explorer/mNeongreen) channel. Scale Bar = 25 μm. Each experiment was conducted in parallel in three separate wells, and a representative confocal slice from one well is shown.

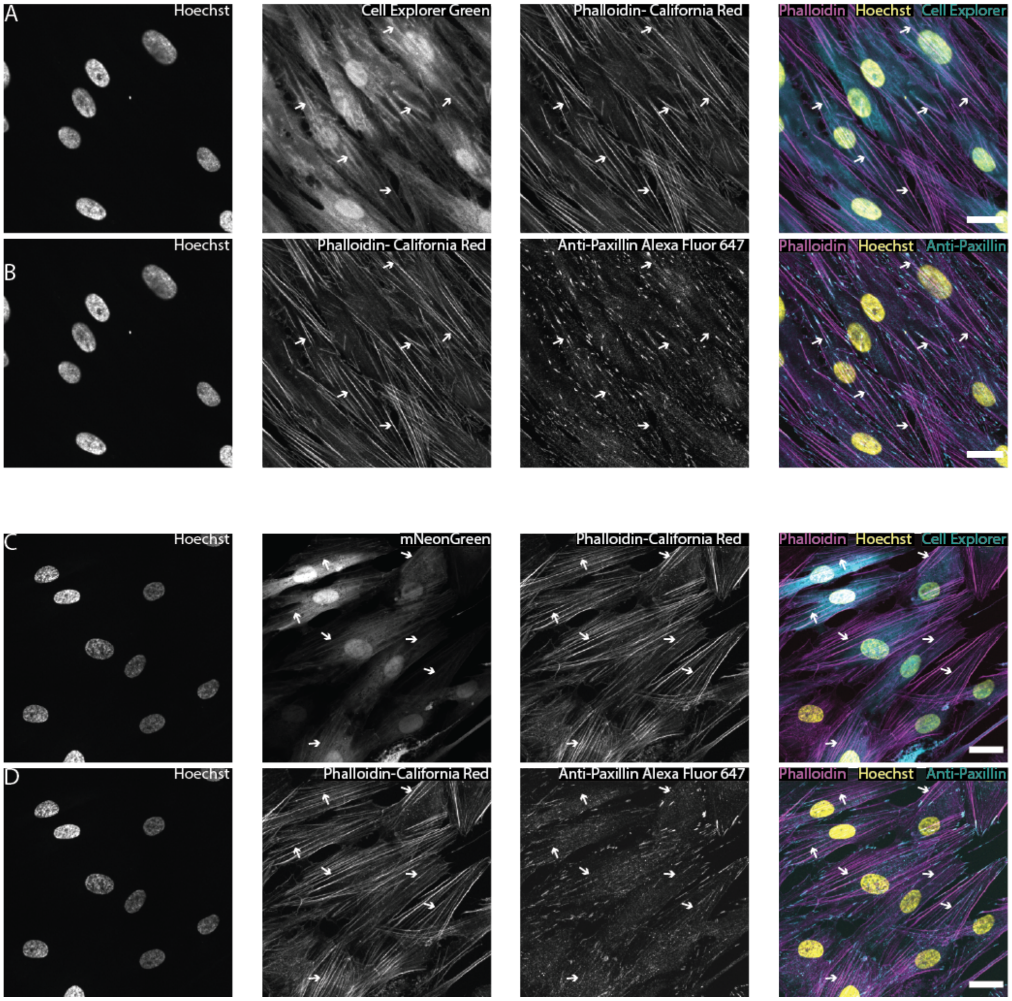
A
Hoechst
Cell Explorer Green
Phalloidin- California Red
Phalloidin Hoechst Cell Explorer
B
Hoechst
Phalloidin- California Red
Anti-Paxillin Alexa Fluor 647
Phalloidin Hoechst Anti-Paxillin
C
Hoechst
mNeonGreen
Phalloidin-California Red
Phalloidin Hoechst Cell Explorer
D
Hoechst
Phalloidin-California Red
Anti-Paxillin Alexa Fluor 647
Phalloidin Hoechst Anti-Paxillin

Figure B·2: Fluorescent Structures Colocalize with Ventral Actin Stress Fibers (A&B) Cells either stained with green Cell Explorer dye or expressing mNeonGreen **(C&D)** were fixed and stained with Hoechst (nuclei), phalloidin-California Red (actin), and an anti-phospho-paxillin primary antibody (focal adhesions) with an Alexa Fluor-647 secondary. The fluorescent structures (examples marked with white arrows) observed with either the **(A)** Cell Explorer dye or **(C)** mNeonGreen colocalize with phalloidin-stained stress fibers. The colocalized fibers have focal adhesions on both ends of the fiber **(B&D)**, identifying them as ventral stress fibers. Scale Bar = 25 μm A&B and C&D are different channels for the same field of view (note: There is some bleedthrough from the Hoechst Channel into the green channel.In addition, there is some accumulation of fluorophore in the nucleus, as they all have a molecular weight below the 40 kDa nuclear diffusion limit [178]). Each experiment was conducted in parallel in three separate wells, and a representative confocal slice from one well is shown.

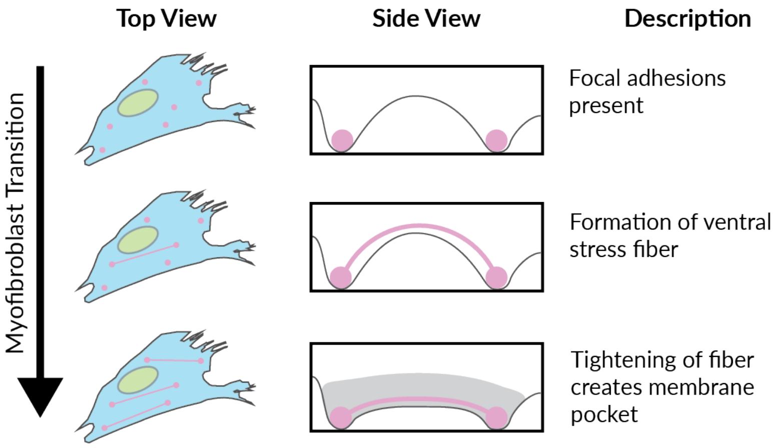
Top View
Side View
Description
Myofibroblast Transition
Focal adhesions present
Formation of ventral stress fiber
Tightening of fiber creates membrane pocket

Figure B·3: Schematic of a proposed mechanism for the development of the observed fluorescent structures. As fibroblasts transition into myofibrobolasts, ventral actin stress fibers (magenta rods) originating from focal adhesions (magenta circles) deform the plasma membrane, creating cytosolic pockets (grey) for the fluorescent dye or proteins to flow into, leading to the observed fluorescent structures.

Before experimentally testing this, I performed some simple modeling to ensure that the energy requirements for this hypothesis fell within reasonable bounds of the cellular energy budget.

B.3 Results

B.3.1 Estimation of energy required for membrane curvature

I first estimated if the forces and energy required to create these ridges are both possible and reasonable within the constraints of the cellular energy budget. To estimate the energy cost of membrane deformation by the ventral stress fiber, I started with calculating the free energy, $\mathrm{G}_{\mathrm{bend}}$, required to bend a membrane:

$$G_{bend}[h(x,y)] = \frac{K_b}{2}\int[\kappa_1(x,y)+\kappa_2(x,y)]^2 da \tag{B.1}$$

where h(x,y) is the height of the membrane at position (x,y) relative to some reference height, K_b is the membrane bending rigidity (typically on the order of 10-20 $k_B T$), κ_1 is the curvature of the membrane in the x dimension at position (x,y), κ_2 is the curvature of the membrane in the y dimension at position (x,y), and da is differential area.

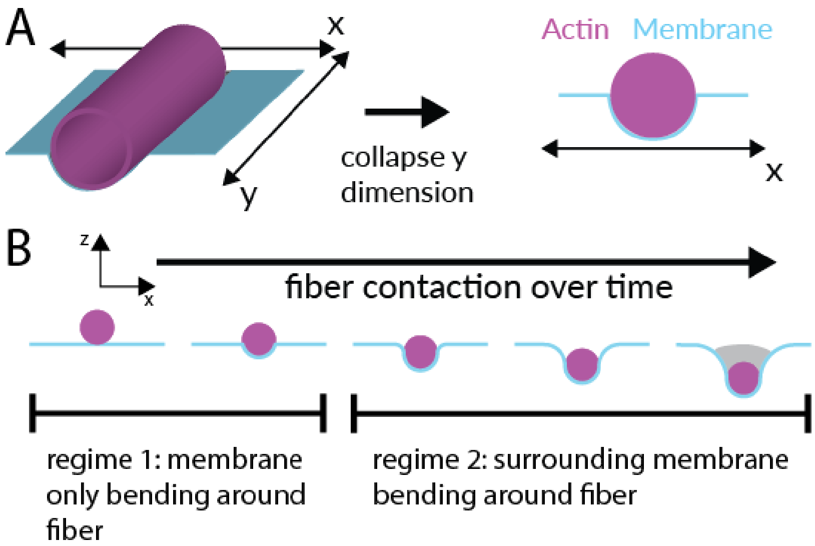
A
x
y
collapse y dimension
Actin
Membrane
x
B
z
x
fiber contaction over time
regime 1: membrane only bending around fiber
regime 2: surrounding membrane bending around fiber

Figure B·4: Model conceptualization of stress fiber-induced membrane deformation (A) The ventral stress fiber is modeled as a cylinder deforming a planar membrane. As the y dimension is uniform, the model is collapsed to one dimension. **(B)** Membrane deformation model used for membrane energy calculation. A stress fiber was modeled as lowering into a membrane, causing the membrane to curve. This fell into two regimes: one where the membrane is only deforming around the fiber and one where parts of the membrane beyond the fiber are deforming. The energy requirement for bending the membrane were calculated across both regimes.

To simplify my calculations, I assumed the fibers to be a straight cylinder indenting a planar membrane (Figure B·4A). This collapses the curvature consideration to one dimension, as all of the curvature along the y-axis will be the same at any given x (Figure B·4B). This simplified my free energy equation calculation to Equation B.2:

$$G_{bend}[h(x,y)] = \frac{K_b * L}{2} \int [\kappa_1(x)]^2 dx \tag{B.2}$$

where L is the length of the fiber cylinder. Curvature of a 1D line is calculated using Equation B.3:

$$\kappa(x) = \frac{|f''(x)|}{(1 + f'(x)^2)^{3/2}} \tag{B.3}$$

where *f(x)* is the function describing the change in height of a line. I built some simple fitting equations to model this profile of the membrane around the fiber (see Methods). The range of free energy requirements fell into two regimes: when the center point of the fiber was modeled above the plane of the membrane and when the center point of the fiber was below the plane of the membrane. In the first regime, the free energy requirements were calculated using the assumption that the membrane bent directly around the fiber (Figure B·4B, first two panels). The calculated energies fell in the range of 0-500 $k_B T$. In the second regime, along with bending around the fiber, parts of the membrane extending out past the fiber diameter were also simulated as bending (Figure B·4B, last three panels). While the exact energy values depend on how these bending equations describe the membrane bending, we found the estimates generally fell in the range of a few thousand $k_B T$. As a comparison point, the free energy of vesicle formation is about 500 $k_B T$, so the first regime of this membrane bending phenomenon is estimated to be in the same order of magnitude, while the second regime, the one that predicts the pocket where fluorescent molecules would diffuse (Figure B·4B, final panel), falls no more than one order of magnitude above this known phenomenon. After examining the overall energy requirements, we then examined if the actual energy budget required to induce this phenomenon was reasonable given the time frame and estimated energy requirement.

B.3.2 Estimation of energy required for fiber contraction

To see if the energy of the membrane bending hypothesis fell within a reasonable energy budget of a fibroblast cell, I estimated the number of ATP molecules required to contract the stress fibers to cause sufficient bending in the membrane. While we recognize that the addition of actin filaments to a ventral stress fiber is a dynamic process, we assumed a static bundle of fibers for the purposes of this calculation. I started by assuming a range of possible radii for the stress fibers. I then calculated the number of individual actin filaments in a given cross-section based on an actin filament radius of 8 nm. This allowed us to calculate the number of individual fibers of actin that needed to contract using Equation B.4:

$$N_{actinfilaments} = \frac{CSA_{VSF}}{CSA_{actinfilament}} = \frac{CSA_{VSF}}{8nm} \tag{B.4}$$

Actin polymers are contracted by myosin motors, whose step size has been estimated at 5 nm [179]. It has also been measured that a myosin motor requires 1 ATP/step [180]. Using these estimates, I explored a range of ATP requirements for a variety of fiber radii and contraction lengths. Given the assumptions and estimates made, I calculated a linear relationship between energy requirement and how much length the fiber contracts, as shown by Equation B.5:

$$N_{ATP} = \frac{L_{contraction}}{L_{stepsize}} * N_{ATP/step} * N_{actinfilaments} \tag{B.5}$$

I calculated this relationship over a range of potential radius values for the fibers (Figure B·5). Even for the largest estimate of fiber radius (250 nm), we calculated the energy requirement to be on the order of millions of ATP molecules.

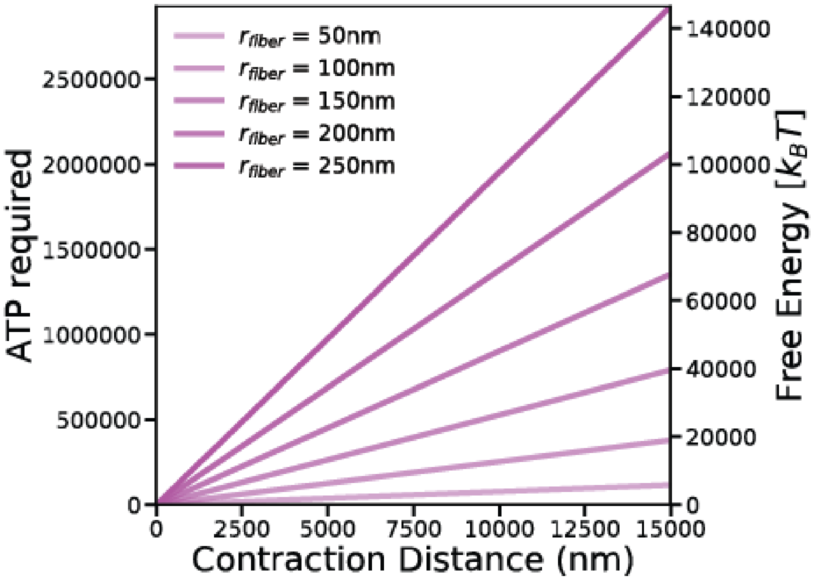

Figure B·5: Calculation results from stress fiber contraction calculation. In our proposed model, the contraction of the ventral stress fibers by myosin II motors drives the formation of the cytosolic pockets. Here, we consider the energy required for that contraction over a range of observed fiber radii (50-250 nm) and contraction distances (0-15000 nm). The calculated ATP (left y-axis) and k_BT equivalents (right y-axis) indicate the proposed model is reasonable given the timeframe of cytosolic pocket formation.

B.4 Discussion and conclusions

To understand if these values were reasonable, I compared the ATP requirement we calculated with an estimate of the total ATP budget for a fibroblast cell. One calculation estimated fibroblast ATP production at 1 billion ATP/sec/cell [181], putting a multi-day formation of these fibers comfortably within the energy budget of the cell, ensuring that energy constraints were not a reason to rule out our hypothesis that the contraction of ventral stress fibers causes the formation of these structures.

B.5 Materials and Methods

Plasmid Construction

pLenti CMV Puro Dest ERK-KTR was digested with *BsrgI* to remove the ERK-KTR gene. The appropriate primers for each gene (see primer table) were used to generate PCR fragments of the gene and add 20-25 base homology arms to each end of the PCR fragment. The digested pLenti CMV Puro Dest vector and PCR fragment were assembled using NEB HiFi Assembly mix, and the mixture was transformed into NEBstable *E. coli* and grown at 30°C overnight. Plasmids were sequence verified by Sanger Sequencing provided by Quintara Biosceince using the in-house CMV Forward (BP0002) and WPRE (BP0156) reverse primers.

Viral Production

HEK 293FT cells (passages 3-15) were plated at 90% confluency in a T-25 flask in HEK cell media (DMEM, 10% FBS, 1X Glutamax, 1x NEAA). After 24 hours, the flask was transfected with the pLenti plasmid and the packaging VSV-G and PSPAX-2 plasmids using Lipofectamine 3000, according to the manufacturer's protocol. After 12 hours, the media in the dish was discarded and replaced. Media from the flask containing viral particles was collected 24 hours later, replaced and collected again 24 hours later. The viral media was spun at 300 x g for 10 minutes to pellet any cells, and the supernatant was then passed through a .45 μm syringe filter. The resulting media was aliquoted in 500 μL tubes and stored at -80°C.

NHDF Cell Culture

Neonatal Human Dermal Fibroblasts (passages 1-8) were cultured in fibroblast media (FGM media supplemented with an FGM-2 OneShot kit) in an incubator at 37°C and

5% CO_2. Cells were passaged at 80-90% confluency, and media was changed every 48 hours. To generate stably expressing pools of cells, fibroblasts were lifted from a flask by incubating the cells in .05% trypsin and then pelleted by centrifugation at 300 x g for 5 minutes. The cells were then resuspended in fresh fibroblast media and seeded in a 24 well plate at a concentration of 20,000 cells/well. 500 μL of viral media and 500 μL of fibroblast media were then added to the well, and the dish was incubated at 37°C for 48 hours. The media in the well was then changed to fibroblast media with 1 μg/mL puromycin to select for positively transduced cells. After 48 hours, the cells were transferred to either a 6 well dish for continued passaging, or a new 24 well plate for experimentation.

Stress Fiber Induction

Fibroblasts were seeded at 20,000 cells/well in a glass bottomed 24 well dish. After 24 hours, stress fibers were induced by changing the cell media to a serum-free induction media (DMEM, 2% B-27, 10 ng/μL TGFβ-1) which was refreshed every 48 hours. Cells were used after 96 hours of induction.

Confocal Microscopy

After 96 hours of induction, the cells were stained with various live cell stains according to the manufacturer's instructions, which generally involved diluting a stock solution 1000x (10,000x for Cell Mask Actin stain) in DMEM, and incubating the cells for 15 - 30 minutes at 37°C and 5% CO_2. After staining, the cell media was changed to imaging media (Fluorobrite DMEM, 1% Glutamax, 1% OxyFluor). Cells were imaged on a Ti-2E Eclipse (Nikon Instruments) with a Dragonfly Spinning Disk confocal system (Oxford Instruments) in a 37°C and 5% CO_2 stage top incubator (OKO labs). Images were acquired on an iXon 888 Life EM-CCD camera (Oxford Instruments).

Fluorescent dyes were imaged through a 405/488/561/647 dichroic mirror using the following excitation laser/emission filter combinations: Ex.405 nm-Em.445/50, Ex. 488 nm-Em. 515/30, Ex. 561 nm-Em. 590/60, Ex. 647- Em. 698/60. All staining and drug treatments were repeated in 3 separate wells, and a representative confocal slice was selected from each treatment for display in a figure. Staining images were acquired in a 5x5 grid with 10% overlap and stitched using the Fiji Grid/Collection Stitching plugin [182]. To display the small structures clearly, a 1024x1024 region of the stitched image is displayed in each figure.

Immunofluorescence

After 96 hours of induction, cells were fixed in 4% Paraformaldehyde in PBS and permeabilized using 0.1% Triton-x. Non-specific interactions were blocked using 10% normal goat serum in PBS. The cells were then incubated overnight with the primary antibody (anti-paxillin 1:50 and anti-ERM 1:100), washed in PBS, and then followed by incubation with the appropriate Alexa Fluor 647 secondary antibody (1:200) in 10% normal goat serum for 1 hour at room temperature. In experiments where phalloidin or membrane stain was used, it was added after the secondary at 1:1000 in PBS and incubated for 30 minutes at room temperature. Cells were washed 3X in PBS and then imaged using the same parameters as previously described using a Plan Apochromatic 100x silicone oil immersion objective (Nikon), but at room temperature with no CO_2.

Membrane Bending Energy Calculations

All model calculations were carried out using a Python script in Spyder (version 4.1.5) using the NumPy [83], Matplotlib [85], and SciPy [84] packages. To calculate membrane bending, the profile of the membrane wrapping around the fiber was defined as follows: until the midpoint of the fiber is level with the membrane, the membrane wraps directly around the fiber using a square root function. As the midpoint of the fiber dips below the level of the membrane, parts of the membrane to either side begin to bend. This is modeled by fitting hyperbolic curves that start at the y value at the midpoint of the fiber and have a length half the distance from the fiber midpoint to the resting level of the membrane. After the profile of the membrane has been defined, the curvature at each point was calculated using the derivative function from the SciPy package. The bending energy equation is applied for each point of curvature and summed over the whole stretch of membrane modeled to get the total bending energy. Code available online at https://github.com/sgrolab/ventralsfpaper.

Bibliography

[1] C.-H. Heldin, B. Lu, R. Evans, and J. S. Gutkind, "Signals and Receptors," *Cold Spring Harbor Perspectives in Biology*, vol. 8, p. a005900, Apr. 2016.

[2] A. Ortega, I. B. Zhulin, and T. Krell, "Sensory Repertoire of Bacterial Chemoreceptors," *Microbiology and Molecular Biology Reviews*, vol. 81, pp. e00033–17, Oct. 2017.

[3] D. G. Knyazev, R. Kuttner, A.-N. Bondar, M. Zimmerman, C. Siligan, and P. Pohl, "Voltage Sensing in Bacterial Protein Translocation," *Biomolecules*, vol. 10, p. 78, Jan. 2020.

[4] C. D. Cox, N. Bavi, and B. Martinac, "Bacterial Mechanosensors," *Annual Review of Physiology*, vol. 80, no. 1, pp. 71–93, 2018.

[5] H. C. Berg and E. M. Purcell, "Physics of chemoreception," *Biophysical Journal*, vol. 20, pp. 193–219, Nov. 1977.

[6] S. Mukherjee and B. L. Bassler, "Bacterial quorum sensing in complex and dynamically changing environments," *Nature Reviews Microbiology*, vol. 17, pp. 371–382, June 2019.

[7] K. Lewis, "Programmed Death in Bacteria," *Microbiology and Molecular Biology Reviews*, vol. 64, pp. 503–514, Sept. 2000.

[8] E. U. Azeloglu and R. Iyengar, "Signaling Networks: Information Flow, Computation, and Decision Making," *Cold Spring Harbor Perspectives in Biology*, vol. 7, p. a005934, Apr. 2015.

[9] N. Blüthgen, "Signaling output: it's all about timing and feedbacks," *Molecular Systems Biology*, vol. 11, p. 843, Nov. 2015.

[10] J. Das, M. Ho, J. Zikherman, C. Govern, M. Yang, A. Weiss, A. K. Chakraborty, and J. P. Roose, "Digital Signaling and Hysteresis Characterize Ras Activation in Lymphoid Cells," *Cell*, vol. 136, pp. 337–351, Jan. 2009.

[11] S. Jeknić, T. Kudo, and M. W. Covert, "Techniques for Studying Decoding of Single Cell Dynamics," *Frontiers in Immunology*, vol. 10, 2019.

[12] Q. Ni, S. Mehta, and J. Zhang, "Live-cell imaging of cell signaling using genetically encoded fluorescent reporters," *The FEBS Journal*, vol. 285, no. 2, pp. 203–219, 2018.

[13] D. Muzzey and A. van Oudenaarden, "Quantitative Time-Lapse Fluorescence Microscopy in Single Cells," *Annual Review of Cell and Developmental Biology*, vol. 25, no. 1, pp. 301–327, 2009.

[14] H.-J. Cheng, C.-H. Hsu, C.-L. Hung, and C.-Y. Lin, "A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies," *Biomedical Journal*, vol. 45, pp. 465–471, June 2022.

[15] S. Bandara and T. Meyer, "Design of Experiments to Investigate Dynamic Cell Signaling Models," in *Computational Modeling of Signaling Networks* (X. Liu and M. D. Betterton, eds.), Methods in Molecular Biology, pp. 109–118, Totowa, NJ: Humana Press, 2012.

[16] K. A. Janes and D. A. Lauffenburger, "Models of signalling networks – what cell biologists can gain from them and give to them," *Journal of Cell Science*, vol. 126, pp. 1913–1921, May 2013.

[17] X. Zhu, E. R. Hager, C. Huyan, and A. E. Sgro, "Leveraging the model-experiment loop: Examples from cellular slime mold chemotaxis," *Experimental Cell Research*, vol. 418, p. 113218, Sept. 2022.

[18] K. Nath and A. L. Koch, "Protein Degradation in Escherichia coli: I. MEASUREMENT OF RAPIDLY AND SLOWLY DECAYING COMPONENTS," *Journal of Biological Chemistry*, vol. 245, pp. 2889–2900, June 1970.

[19] C. P. Zschiedrich, V. Keidel, and H. Szurmant, "Molecular Mechanisms of Two-Component Signal Transduction," *Journal of Molecular Biology*, vol. 428, pp. 3752–3775, Sept. 2016.

[20] N. Rosenfeld and U. Alon, "Response Delays and the Structure of Transcription Networks," *Journal of Molecular Biology*, vol. 329, pp. 645–654, June 2003.

[21] T. G. Minchington, S. Griffiths-Jones, and N. Papalopulu, "Dynamical gene regulatory networks are tuned by transcriptional autoregulation with microRNA feedback," *Scientific Reports*, vol. 10, p. 12960, July 2020.

[22] M. A. Savageau, "Comparison of classical and autogenous systems of regulation in inducible operons," *Nature*, vol. 252, pp. 546–549, Dec. 1974.

[23] N. Rosenfeld, M. B. Elowitz, and U. Alon, "Negative Autoregulation Speeds the Response Times of Transcription Networks," *Journal of Molecular Biology*, vol. 323, pp. 785–793, Nov. 2002.

[24] Y. T. Maeda and M. Sano, "Regulatory Dynamics of Synthetic Gene Networks with Positive Feedback," *Journal of Molecular Biology*, vol. 359, pp. 1107–1124, June 2006.

[25] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature Reviews Neuroscience*, vol. 9, pp. 292–303, Apr. 2008.

[26] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, Aug. 2002.

[27] J. M. Raser, "Control of Stochasticity in Eukaryotic Gene Expression," *Science*, vol. 304, pp. 1811–1814, June 2004.

[28] A. Raj and A. van Oudenaarden, "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences," *Cell*, vol. 135, pp. 216–226, Oct. 2008.

[29] M. L. Simpson, C. D. Cox, M. S. Allen, J. M. McCollum, R. D. Dar, D. K. Karig, and J. F. Cooke, "Noise in biological circuits," *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, vol. 1, pp. 214–225, Mar. 2009.

[30] A. Eldar and M. B. Elowitz, "Functional roles for noise in genetic circuits," *Nature*, vol. 467, pp. 167–173, Sept. 2010.

[31] G. Balázsi, A. van Oudenaarden, and J. J. Collins, "Cellular Decision-Making and Biological Noise: From Microbes to Mammals," *Cell*, vol. 144, pp. 910–925, Mar. 2011.

[32] O. Symmons and A. Raj, "What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism," *Molecular Cell*, vol. 62, pp. 788–802, June 2016.

[33] N. Eling, M. D. Morgan, and J. C. Marioni, "Challenges in measuring and understanding biological noise," *Nature Reviews Genetics*, vol. 20, pp. 536–548, Sept. 2019.

[34] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, "An excitable gene regulatory circuit induces transient cellular differentiation," *Nature*, vol. 440, pp. 545–550, Mar. 2006.

[35] J. R. Russell, M. T. Cabeen, P. A. Wiggins, J. Paulsson, and R. Losick, "Noise in a phosphorelay drives stochastic entry into sporulation in Bacillus subtilis," *The EMBO journal*, vol. 36, pp. 2856–2869, Oct. 2017.

[36] G. Chalancon, C. N. J. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, and M. M. Babu, "Interplay between gene expression noise and regulatory network architecture," *Trends in Genetics*, vol. 28, pp. 221–232, May 2012.

[37] A. Becskei and L. Serrano, “Engineering stability in gene networks by autoregulation,” *Nature*, vol. 405, pp. 590–593, June 2000.

[38] G. Shinar, R. Milo, M. R. Martínez, and U. Alon, “Input–output robustness in simple bacterial signaling systems,” *Proceedings of the National Academy of Sciences*, vol. 104, pp. 19931–19935, Dec. 2007.

[39] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, “Real-Time Kinetics of Gene Activity in Individual Bacteria,” *Cell*, vol. 123, pp. 1025–1036, Dec. 2005.

[40] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, “Gene Regulation at the Single-Cell Level,” *Science*, vol. 307, pp. 1962–1965, Mar. 2005.

[41] D. Huh and J. Paulsson, “Non-genetic heterogeneity from stochastic partitioning at cell division,” *Nature Genetics*, vol. 43, pp. 95–100, Feb. 2011.

[42] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, “Intercellular Variability in Protein Levels from Stochastic Expression and Noisy Cell Cycle Processes,” *PLOS Computational Biology*, vol. 12, p. e1004972, Aug. 2016.

[43] D. Huh and J. Paulsson, “Random partitioning of molecules at cell division,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 15004–15009, Sept. 2011.

[44] T. M. Norman, N. D. Lord, J. Paulsson, and R. Losick, “Stochastic Switching of Cell Fate in Microbes,” *Annual Review of Microbiology*, vol. 69, no. 1, pp. 381–403, 2015.

[45] W. L. Nicholson, N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow, “Resistance of Bacillus Endospores to Extreme Terrestrial and Extraterrestrial Environments,” *Microbiology and Molecular Biology Reviews*, vol. 64, pp. 548–572, Sept. 2000.

[46] J. H. Levine, M. E. Fontes, J. Dworkin, and M. B. Elowitz, “Pulsed Feedback Defers Cellular Differentiation,” *PLOS Biology*, vol. 10, p. e1001252, Jan. 2012.

[47] A. Kuchina, L. Espinar, J. Garcia-Ojalvo, and G. M. Süel, “Reversible and Noisy Progression towards a Commitment Point Enables Adaptable and Reliable Cellular Decision-Making,” *PLOS Computational Biology*, vol. 7, p. e1002273, Nov. 2011.

[48] A. Kuchina, L. Espinar, T. Çağatay, A. O. Balbin, F. Zhang, A. Alvarado, J. Garcia-Ojalvo, and G. M. Süel, “Temporal competition between differentiation programs determines cell fate choice,” *Molecular Systems Biology*, vol. 7, p. 557, Jan. 2011.

[49] C. A. Weiss, J. A. Hoberg, K. Liu, B. P. Tu, and W. C. Winkler, “Single-Cell Microscopy Reveals That Levels of Cyclic di-GMP Vary among Bacillus subtilis Subpopulations,” *Journal of Bacteriology*, vol. 201, pp. e00247–19, July 2019.

[50] H. Sondermann, N. J. Shikuma, and F. H. Yildiz, “You’ve come a long way: c-di-GMP signaling,” *Current Opinion in Microbiology*, vol. 15, pp. 140–146, Apr. 2012.

[51] J. L. Spudich and D. E. Koshland, “Non-genetic individuality: chance in the single cell,” *Nature*, vol. 262, pp. 467–471, Aug. 1976.

[52] I. El Meouche and M. J. Dunlop, “Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation,” *Science*, vol. 362, pp. 686–690, Nov. 2018.

[53] I. Adrian-Kalchhauser, S. E. Sultan, L. N. Shama, H. Spence-Jones, S. Tiso, C. I. Keller Valsecchi, and F. J. Weissing, “Understanding ’Non-genetic’ Inheritance: Insights from Molecular-Evolutionary Crosstalk,” *Trends in Ecology & Evolution*, vol. 35, pp. 1078–1089, Dec. 2020.

[54] C. S. Gokhale, S. Giaimo, and P. Remigi, “Memory shapes microbial populations,” *PLOS Computational Biology*, vol. 17, p. e1009431, Oct. 2021.

[55] M. Pleška, D. Jordan, Z. Frentz, B. Xue, and S. Leibler, “Nongenetic individuality, changeability, and inheritance in bacterial behavior,” *Proceedings of the National Academy of Sciences*, vol. 118, p. e2023322118, Mar. 2021.

[56] H. Vashistha, M. Kohram, and H. Salman, “Non-genetic inheritance restraint of cell-to-cell variation,” *eLife*, vol. 10, p. e64779, Feb. 2021.

[57] S. K. Govers, A. Adam, H. Blockeel, and A. Aertsen, “Rapid phenotypic individualization of bacterial sister cells,” *Scientific Reports*, vol. 7, p. 8473, Aug. 2017.

[58] E. E. Kuchen, N. B. Becker, N. Claudino, and T. Höfer, “Hidden long-range memories of growth and cycle speed correlate cell cycles in lineage trees,” *eLife*, vol. 9, p. e51002, Jan. 2020.

[59] A. J. Waite, N. W. Frankel, Y. S. Dufour, J. F. Johnston, J. Long, and T. Emonet, “Non-genetic diversity modulates population performance,” *Molecular Systems Biology*, vol. 12, p. 895, Dec. 2016.

[60] H. H. Mattingly and T. Emonet, “Collective behavior and nongenetic inheritance allow bacterial populations to adapt to changing environments,” *Proceedings of the National Academy of Sciences*, vol. 119, p. e2117377119, June 2022.

[61] D. R. Rigney, “Stochastic model of constitutive protein levels in growing and dividing bacterial cells,” *Journal of Theoretical Biology*, vol. 76, pp. 453–480, Feb. 1979.

[62] O. G. Berg, “A model for the statistical fluctuations of protein numbers in a microbial population,” *Journal of Theoretical Biology*, vol. 71, pp. 587–603, Apr. 1978.

[63] D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” *The Journal of Physical Chemistry*, vol. 81, pp. 2340–2361, Dec. 1977.

[64] T. Schirmer, “C-di-GMP Synthesis: Structural Aspects of Evolution, Catalysis and Regulation,” *Journal of Molecular Biology*, vol. 428, pp. 3683–3701, Sept. 2016.

[65] S. Kunz, A. Tribensky, W. Steinchen, L. Oviedo-Bocanegra, P. Bedrunka, and P. L. Graumann, “Cyclic di-GMP Signaling in Bacillus subtilis Is Governed by Direct Interactions of Diguanylate Cyclases and Cognate Receptors,” *mBio*, vol. 11, pp. e03122–19, Mar. 2020.

[66] S. Cooper and C. E. Helmstetter, “Chromosome replication and the division cycle of Escherichia coli Br,” *Journal of Molecular Biology*, vol. 31, pp. 519–540, Feb. 1968.

[67] D. R. Cox, “Some Statistical Methods Connected with Series of Events,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 17, no. 2, pp. 129–157, 1955.

[68] K. P. Koutsoumanis and A. Lianou, “Stochasticity in Colonial Growth Dynamics of Individual Bacterial Cells,” *Applied and Environmental Microbiology*, vol. 79, pp. 2294–2301, Apr. 2013.

[69] D. Davidi, E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, and R. Milo, “Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements,” *Proceedings of the National Academy of Sciences*, vol. 113, pp. 3401–3406, Mar. 2016.

[70] B. Soufi, K. Krug, A. Harst, and B. Macek, “Characterization of the E. coli proteome and its modifications during growth and ethanol stress,” *Frontiers in Microbiology*, vol. 6, 2015.

[71] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, “Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli,” *Nature Chemical Biology*, vol. 5, pp. 593–599, Aug. 2009.

[72] R. G. Eagon, “Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes,” *Journal of Bacteriology*, vol. 83, pp. 736–737, Apr. 1962.

[73] X. Dai, Z. Shen, Y. Wang, and M. Zhu, “Sinorhizobium meliloti, a Slow-Growing Bacterium, Exhibits Growth Rate Dependence of Cell Size under Nutrient Limitation,” *mSphere*, vol. 3, pp. e00567–18, Nov. 2018.

[74] T. M. Thomson, K. R. Benjamin, A. Bush, T. Love, D. Pincus, O. Resnekov, R. C. Yu, A. Gordon, A. Colman-Lerner, D. Endy, and R. Brent, “Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 20265–20270, Dec. 2011.

[75] K. A. Johnson and R. S. Goody, “The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper,” *Biochemistry*, vol. 50, pp. 8264–8269, Oct. 2011. Publisher: American Chemical Society.

[76] A. Amir, O. Kobiler, A. Rokney, A. B. Oppenheim, and J. Stavans, “Noise in timing and precision of gene activities in a genetic cascade,” *Molecular Systems Biology*, vol. 3, p. 71, Jan. 2007.

[77] E. A. Arriaga, “Determining biological noise via single cell analysis,” *Analytical and Bioanalytical Chemistry*, vol. 393, pp. 73–80, Jan. 2009.

[78] M. S. Aronson, C. Ricci-Tam, X. Zhu, and A. E. Sgro, “Exploiting noise to engineer adaptability in synthetic multicellular systems,” *Current Opinion in Biomedical Engineering*, vol. 16, pp. 52–60, Dec. 2020.

[79] C. L. Kelly, A. W. Harris, H. Steel, E. J. Hancock, J. T. Heap, and A. Papachristodoulou, “Synthetic negative feedback circuits using engineered small RNAs,” *Nucleic Acids Research*, vol. 46, pp. 9875–9889, Oct. 2018.

[80] M. M. Hansen, W. Y. Wen, E. Ingerman, B. S. Razooky, C. E. Thompson, R. D. Dar, C. W. Chin, M. L. Simpson, and L. S. Weinberger, “A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization,” *Cell*, vol. 173, pp. 1609–1621.e15, June 2018.

[81] K. S. Farquhar, D. A. Charlebois, M. Szenk, J. Cohen, D. Nevozhay, and G. Balázsi, “Role of network-mediated stochasticity in mammalian drug resistance,” *Nature Communications*, vol. 10, pp. 1–14, June 2019.

[82] C. Tan, F. Reza, and L. You, “Noise-Limited Frequency Signal Transmission in Gene Circuits,” *Biophysical Journal*, vol. 93, pp. 3753–3761, Dec. 2007.

[83] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.

[84] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, Mar. 2020.

[85] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, pp. 90–95, May 2007.

[86] H.-C. Flemming and S. Wuertz, "Bacteria and archaea on Earth and their abundance in biofilms," *Nature Reviews Microbiology*, vol. 17, pp. 247–260, Apr. 2019.

[87] H.-C. Flemming, E. D. van Hullebusch, T. R. Neu, P. H. Nielsen, T. Seviour, P. Stoodley, J. Wingender, and S. Wuertz, "The biofilm matrix: multitasking in a shared space," *Nature Reviews Microbiology*, vol. 21, pp. 70–86, Feb. 2023.

[88] L. Hall-Stoodley and P. Stoodley, "Evolving concepts in biofilm infections," *Cellular Microbiology*, vol. 11, no. 7, pp. 1034–1043, 2009.

[89] M. Asally, M. Kittisopikul, P. Rué, Y. Du, Z. Hu, T. Çağatay, A. B. Robinson, H. Lu, J. Garcia-Ojalvo, and G. M. Süel, "Localized cell death focuses mechanical forces during 3D patterning in a biofilm," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 18891–18896, Nov. 2012.

[90] R. Hartmann, P. K. Singh, P. Pearce, R. Mok, B. Song, F. Díaz-Pascual, J. Dunkel, and K. Drescher, "Emergence of three-dimensional order and structure in growing biofilms," *Nature Physics*, vol. 15, pp. 251–256, Mar. 2019.

[91] K. P. Rumbaugh and K. Sauer, "Biofilm dispersion," *Nature Reviews Microbiology*, vol. 18, pp. 571–586, Oct. 2020.

[92] B. Qin, C. Fei, A. A. Bridges, A. A. Mashruwala, H. A. Stone, N. S. Wingreen, and B. L. Bassler, "Cell position fates and collective fountain flow in bacterial biofilms revealed by light-sheet microscopy," *Science*, vol. 369, pp. 71–77, July 2020.

[93] T. Pisithkul, J. W. Schroeder, E. A. Trujillo, P. Yeesin, D. M. Stevenson, T. Chaiamarit, J. J. Coon, J. D. Wang, and D. Amador-Noguez, “Metabolic Remodeling during Biofilm Development of Bacillus subtilis,” *mBio*, vol. 10, pp. e00623–19, May 2019.

[94] S. Mukherjee and D. B. Kearns, “The Structure and Regulation of Flagella in Bacillus subtilis,” *Annual Review of Genetics*, vol. 48, no. 1, pp. 319–340, 2014.

[95] P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom, and M. Benziman, “Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid,” *Nature*, vol. 325, pp. 279–281, Jan. 1987.

[96] U. Jenal and J. Malone, “Mechanisms of Cyclic-di-GMP Signaling in Bacteria,” *Annual Review of Genetics*, vol. 40, no. 1, pp. 385–407, 2006.

[97] R. Hengge, “Principles of c-di-GMP signalling in bacteria,” *Nature Reviews Microbiology*, vol. 7, pp. 263–273, Apr. 2009.

[98] C. M. Waters, “The Meteoric Rise of the Signaling Molecule Cyclic di-GMP,” *Microbe (Washington, D.C.)*, vol. 7, no. 8, pp. 353–359, 2012.

[99] U. Römling, M. Y. Galperin, and M. Gomelsky, “Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger,” *Microbiology and Molecular Biology Reviews*, vol. 77, pp. 1–52, Mar. 2013.

[100] U. Jenal, A. Reinders, and C. Lori, “Cyclic di-GMP: second messenger extraordinaire,” *Nature Reviews Microbiology*, vol. 15, pp. 271–284, May 2017.

[101] D. Kalia, G. Merey, S. Nakayama, Y. Zheng, J. Zhou, Y. Luo, M. Guo, B. T. Roembke, and H. O. Sintim, “Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis,” *Chemical Society Reviews*, vol. 42, pp. 305–341, Dec. 2012.

[102] R. Hengge, S. Häussler, M. Pruteanu, J. Stülke, N. Tschowri, and K. Turgay, “Recent Advances and Current Trends in Nucleotide Second Messenger Signaling in Bacteria,” *Journal of Molecular Biology*, vol. 431, pp. 908–927, Mar. 2019.

[103] R. Hengge, A. Gründling, U. Jenal, R. Ryan, and F. Yildiz, “Bacterial Signal Transduction by Cyclic Di-GMP and Other Nucleotide Second Messengers,” *Journal of Bacteriology*, vol. 198, pp. 15–26, Dec. 2015.

[104] M. Valentini and A. Filloux, “Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from Pseudomonas aeruginosa and Other Bacteria*,” *Journal of Biological Chemistry*, vol. 291, pp. 12547–12555, June 2016.

[105] I. Ahmad, E. Nygren, F. Khalid, S. L. Myint, and B. E. Uhlin, "A Cyclic-di-GMP signalling network regulates biofilm formation and surface associated motility of Acinetobacter baumannii 17978," *Scientific Reports*, vol. 10, p. 1991, Feb. 2020.

[106] C. Liu, D. Sun, J. Liu, Y. Chen, X. Zhou, Y. Ru, J. Zhu, and W. Liu, "cAMP and c-di-GMP synergistically support biofilm maintenance through the direct interaction of their effectors," *Nature Communications*, vol. 13, p. 1493, Mar. 2022.

[107] D.-G. Ha and G. A. O'Toole, "c-di-GMP and its Effects on Biofilm Formation and Dispersion: a Pseudomonas Aeruginosa Review," *Microbiology Spectrum*, vol. 3, p. 3.2.27, Apr. 2015.

[108] A. M. Richter, T. L. Povolotsky, L. H. Wieler, and R. Hengge, "Cyclic-di-GMP signalling and biofilm-related properties of the Shiga toxin-producing 2011 German outbreak Escherichia coli O104:H4," *EMBO Molecular Medicine*, vol. 6, pp. 1622–1637, Dec. 2014.

[109] J. G. Conner, D. Zamorano-Sánchez, J. H. Park, H. Sondermann, and F. H. Yildiz, "The ins and outs of cyclic di-GMP signaling in Vibrio cholerae," *Current Opinion in Microbiology*, vol. 36, pp. 20–29, Apr. 2017.

[110] A. Lamprokostopoulou, C. Monteiro, M. Rhen, and U. Römling, "Cyclic di-GMP signalling controls virulence properties of Salmonella enterica serovar Typhimurium at the mucosal lining," *Environmental Microbiology*, vol. 12, no. 1, pp. 40–53, 2010.

[111] E. B. Purcell and R. Tamayo, "Cyclic diguanylate signaling in Gram-positive bacteria," *FEMS Microbiology Reviews*, vol. 40, pp. 753–773, Sept. 2016.

[112] R. Losick, "A Love Affair with Bacillus subtilis," *Journal of Biological Chemistry*, vol. 290, pp. 2529–2538, Jan. 2015.

[113] J. Errington, "Determination of cell fate in Bacillus subtilis," *Trends in Genetics*, vol. 12, pp. 31–34, Jan. 1996.

[114] J. C. W. Locke, J. W. Young, M. Fontes, M. J. H. Jimenez, and M. B. Elowitz, "Stochastic Pulse Regulation in Bacterial Stress Response," *Science*, vol. 334, pp. 366–369, Oct. 2011.

[115] N. Mirouze and D. Dubnau, "Chance and Necessity in Bacillus subtilis Development," *Microbiology Spectrum*, vol. 1, p. 1.1.02, Oct. 2013.

[116] B. Mielich-Süss and D. Lopez, "Molecular mechanisms involved in *B acillus subtilis* biofilm formation: Biofilm formation in *Bacillus subtilis*," *Environmental Microbiology*, vol. 17, pp. 555–565, Mar. 2015.

[117] H. Vlamakis, Y. Chai, P. Beauregard, R. Losick, and R. Kolter, "Sticking together: building a biofilm the Bacillus subtilis way," *Nature Reviews Microbiology*, vol. 11, pp. 157–168, Mar. 2013.

[118] S. Arnaouteli, N. C. Bamford, N. R. Stanley-Wall, and A. T. Kovács, "Bacillus subtilis biofilm formation and social interactions," *Nature Reviews Microbiology*, vol. 19, pp. 600–614, Sept. 2021.

[119] Y. Chai, R. Kolter, and R. Losick, "Paralogous Antirepressors Acting on the Master Regulator for Biofilm Formation in Bacillus subtilis," *Molecular microbiology*, vol. 74, pp. 876–887, Nov. 2009.

[120] J. Kampf, J. Gerwig, K. Kruse, R. Cleverley, M. Dormeyer, A. Grünberger, D. Kohlheyer, F. M. Commichau, R. J. Lewis, and J. Stülke, "Selective Pressure for Biofilm Formation in Bacillus subtilis: Differential Effect of Mutations in the Master Regulator SinR on Bistability," *mBio*, vol. 9, pp. e01464–18, Sept. 2018.

[121] T. M. Norman, N. D. Lord, J. Paulsson, and R. Losick, "Memory and modularity in cell-fate decision making," *Nature*, vol. 503, pp. 481–486, Nov. 2013.

[122] N. D. Lord, T. M. Norman, R. Yuan, S. Bakshi, R. Losick, and J. Paulsson, "Stochastic antagonism between two proteins governs a bacterial cell fate switch," *Science*, vol. 366, pp. 116–120, Oct. 2019.

[123] A. T. Kovács, "Bacterial differentiation via gradual activation of global regulators," *Current Genetics*, vol. 62, pp. 125–128, Feb. 2016.

[124] V. Molle, M. Fujita, S. T. Jensen, P. Eichenberger, J. E. González-Pastor, J. S. Liu, and R. Losick, "The Spo0A regulon of Bacillus subtilis," *Molecular Microbiology*, vol. 50, no. 5, pp. 1683–1701, 2003.

[125] D. Schultz, "Coordination of cell decisions and promotion of phenotypic diversity in *B. subtilis* via pulsed behavior of the phosphorelay," *BioEssays*, vol. 38, pp. 440–445, May 2016.

[126] V. Parashar, M. A. Konkol, D. B. Kearns, and M. B. Neiditch, "A Plasmid-Encoded Phosphatase Regulates Bacillus subtilis Biofilm Architecture, Sporulation, and Genetic Competence," *Journal of Bacteriology*, vol. 195, pp. 2437–2448, May 2013.

[127] M. Fujita, J. E. González-Pastor, and R. Losick, "High- and Low-Threshold Genes in the Spo0A Regulon of Bacillus subtilis," *Journal of Bacteriology*, vol. 187, pp. 1357–1368, Feb. 2005.

[128] Y. Chen, Y. Chai, J.-h. Guo, and R. Losick, “Evidence for Cyclic Di-GMP-Mediated Signaling in Bacillus subtilis,” *Journal of Bacteriology*, vol. 194, pp. 5080–5090, Sept. 2012.

[129] X. Gao, S. Mukherjee, P. M. Matthews, L. A. Hammad, D. B. Kearns, and C. E. Dann, “Functional Characterization of Core Components of the Bacillus subtilis Cyclic-Di-GMP Signaling Pathway,” *Journal of Bacteriology*, vol. 195, pp. 4782–4792, Nov. 2013.

[130] S. Subramanian, X. Gao, C. E. Dann, and D. B. Kearns, “MotI (DgrA) acts as a molecular clutch on the flagellar stator protein MotA in Bacillus subtilis,” *Proceedings of the National Academy of Sciences*, vol. 114, pp. 13537–13542, Dec. 2017.

[131] V. M. Worlitzer, A. Jose, I. Grinberg, M. Bär, S. Heidenreich, A. Eldar, G. Ariel, and A. Be’er, “Biophysical aspects underlying the swarm to biofilm transition,” *Science Advances*, June 2022.

[132] P. Bedrunka and P. L. Graumann, “New Functions and Subcellular Localization Patterns of c-di-GMP Components (GGDEF Domain Proteins) in B. subtilis,” *Frontiers in Microbiology*, vol. 8, p. 794, May 2017.

[133] P. Bedrunka and P. L. Graumann, “Subcellular clustering of a putative c-di-GMP-dependent exopolysaccharide machinery affecting macro colony architecture in *B acillus subtilis*: Localization of a novel putative exopolysaccharide machinery in *Bacillus subtilis*,” *Environmental Microbiology Reports*, vol. 9, pp. 211–222, June 2017.

[134] M. W. Orr, M. Y. Galperin, and V. T. Lee, “Sustained sensing as an emerging principle in second messenger signaling systems,” *Current Opinion in Microbiology*, vol. 34, pp. 119–126, Dec. 2016.

[135] C. M. Thompson and J. G. Malone, “Nucleotide second messengers in bacterial decision making,” *Current Opinion in Microbiology*, vol. 55, pp. 34–39, June 2020.

[136] A. Petchiappan, S. Y. Naik, and D. Chatterji, “Tracking the homeostasis of second messenger cyclic-di-GMP in bacteria,” *Biophysical Reviews*, vol. 12, pp. 719–730, June 2020.

[137] H. Zhou, C. Zheng, J. Su, B. Chen, Y. Fu, Y. Xie, Q. Tang, S.-H. Chou, and J. He, “Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter,” *Scientific Reports*, vol. 6, p. 20871, Feb. 2016.

[138] M. T. Rybtke, B. R. Borlee, K. Murakami, Y. Irie, M. Hentzer, T. E. Nielsen, M. Givskov, M. R. Parsek, and T. Tolker-Nielsen, "Fluorescence-Based Reporter for Gauging Cyclic Di-GMP Levels in Pseudomonas aeruginosa," *Applied and Environmental Microbiology*, vol. 78, pp. 5060–5069, Aug. 2012.

[139] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, "A guide to choosing fluorescent proteins," *Nature Methods*, vol. 2, pp. 905–909, Dec. 2005.

[140] M. Christen, H. D. Kulasekara, B. Christen, B. R. Kulasekara, L. R. Hoffman, and S. I. Miller, "Asymmetrical Distribution of the Second Messenger c-di-GMP upon Bacterial Cell Division," *Science*, vol. 328, pp. 1295–1297, June 2010.

[141] N. V. Mushnikov, A. Fomicheva, M. Gomelsky, and G. R. Bowman, "Inducible asymmetric cell division and cell differentiation in a bacterium," *Nature Chemical Biology*, vol. 15, pp. 925–931, Sept. 2019.

[142] A. Kepes, "KINETICS OF INDUCED ENZYME SYNTHESIS. DETERMINATION OF THE MEAN LIFE OF GALACTOSIDASE-SPECIFIC MESSENGER RNA," *Biochimica Et Biophysica Acta*, vol. 76, pp. 293–309, Oct. 1963.

[143] O. Yarchuk, N. Jacques, J. Guillerez, and M. Dreyfus, "Interdependence of translation, transcription and mRNA degradation in the lacZ gene," *Journal of Molecular Biology*, vol. 226, pp. 581–596, Aug. 1992.

[144] P. Nicolas, U. Mäder, E. Dervyn, T. Rochat, A. Leduc, N. Pigeonneau, E. Bidnenko, E. Marchadier, M. Hoebeke, S. Aymerich, D. Becher, P. Bisicchia, E. Botella, O. Delumeau, G. Doherty, E. L. Denham, M. J. Fogg, V. Fromion, A. Goelzer, A. Hansen, E. Härtig, C. R. Harwood, G. Homuth, H. Jarmer, M. Jules, E. Klipp, L. Le Chat, F. Lecointe, P. Lewis, W. Liebermeister, A. March, R. A. T. Mars, P. Nannapaneni, D. Noone, S. Pohl, B. Rinn, F. Rügheimer, P. K. Sappa, F. Samson, M. Schaffer, B. Schwikowski, L. Steil, J. Stülke, T. Wiegert, K. M. Devine, A. J. Wilkinson, J. Maarten van Dijl, M. Hecker, U. Völker, P. Bessières, and P. Noirot, "Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus subtilis," *Science*, vol. 335, pp. 1103–1106, Mar. 2012.

[145] N. De, M. V. A. S. Navarro, R. V. Raghavan, and H. Sondermann, "Determinants for the Activation and Autoinhibition of the Diguanylate Cyclase Response Regulator WspR," *Journal of Molecular Biology*, vol. 393, pp. 619–633, Oct. 2009.

[146] F. Rao, R. Y. See, D. Zhang, D. C. Toh, Q. Ji, and Z.-X. Liang, "YybT Is a Signaling Protein That Contains a Cyclic Dinucleotide Phosphodiesterase Domain and a GGDEF Domain with ATPase Activity*," *Journal of Biological Chemistry*, vol. 285, pp. 473–482, Jan. 2010.

[147] M. E. Sharpe, P. M. Hauser, R. G. Sharpe, and J. Errington, “Bacillus subtilis Cell Cycle as Studied by Fluorescence Microscopy: Constancy of Cell Length at Initiation of DNA Replication and Evidence for Active Nucleoid Partitioning,” *Journal of Bacteriology*, vol. 180, pp. 547–555, Feb. 1998.

[148] N. Nordholt, J. H. van Heerden, and F. J. Bruggeman, “Biphasic Cell-Size and Growth-Rate Homeostasis by Single Bacillus subtilis Cells,” *Current Biology*, vol. 30, pp. 2238–2247.e5, June 2020.

[149] H. E. Kubitschek, “Growth During the Bacterial Cell Cycle: Analysis of Cell Size Distribution,” *Biophysical Journal*, vol. 9, pp. 792–809, June 1969.

[150] J. W. Young, J. C. W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. S. Swain, E. Mjolsness, and M. B. Elowitz, “Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy,” *Nature Protocols*, vol. 7, pp. 80–88, Jan. 2012.

[151] S. S. Branda, J. E. González-Pastor, S. Ben-Yehuda, R. Losick, and R. Kolter, “Fruiting body formation by Bacillus subtilis,” *Proceedings of the National Academy of Sciences*, vol. 98, pp. 11621–11626, Sept. 2001.

[152] J. Dervaux, J. C. Magniez, and A. Libchaber, “On growth and form of Bacillus subtilis biofilms,” *Interface Focus*, vol. 4, p. 20130051, Dec. 2014.

[153] O. M. O’Connor, R. N. Alnahhas, J.-B. Lugagne, and M. J. Dunlop, “DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics,” *PLOS Computational Biology*, vol. 18, p. e1009797, Jan. 2022.

[154] A. I. Kostyuk, A. D. Demidovich, D. A. Kotova, V. V. Belousov, and D. S. Bilan, “Circularly Permuted Fluorescent Protein-Based Indicators: History, Principles, and Classification,” *International Journal of Molecular Sciences*, vol. 20, p. 4200, Aug. 2019.

[155] N. Doi and H. Yanagawa, “Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution,” *FEBS Letters*, vol. 453, no. 3, pp. 305–307, 1999.

[156] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, and K. V. Wood, “HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis,” *ACS Chemical Biology*, vol. 3, pp. 373–382, June 2008.

[157] J. B. Grimm, B. P. English, J. Chen, J. P. Slaughter, Z. Zhang, A. Revyakin, R. Patel, J. J. Macklin, D. Normanno, R. H. Singer, T. Lionnet, and L. D. Lavis, "A general method to improve fluorophores for live-cell and single-molecule microscopy," *Nature Methods*, vol. 12, pp. 244–250, Mar. 2015.

[158] J. B. Grimm, T. A. Brown, B. P. English, T. Lionnet, and L. D. Lavis, "Synthesis of Janelia Fluor HaloTag and SNAP-Tag Ligands and Their Use in Cellular Imaging Experiments," in *Super-Resolution Microscopy: Methods and Protocols* (H. Erfle, ed.), Methods in Molecular Biology, pp. 179–188, New York, NY: Springer, 2017.

[159] K.-T. Chou, D.-y. D. Lee, J.-g. Chiou, L. Galera-Laporta, S. Ly, J. Garcia-Ojalvo, and G. M. Süel, "A segmentation clock patterns cellular differentiation in a bacterial biofilm," *Cell*, vol. 185, pp. 145–157.e13, Jan. 2022.

[160] J. Narula, A. Kuchina, D.-y. D. Lee, M. Fujita, G. M. Süel, and O. A. Igoshin, "Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication," *Cell*, vol. 162, pp. 328–337, July 2015.

[161] J. H. Levine and M. B. Elowitz, "Polyphasic feedback enables tunable cellular timers," *Current Biology*, vol. 24, pp. R994–R995, Oct. 2014.

[162] C. R. Gliech and A. J. Holland, "Keeping track of time: The fundamentals of cellular clocks," *Journal of Cell Biology*, vol. 219, p. e202005136, Sept. 2020.

[163] J. Gerardin, N. R. Reddy, and W. A. Lim, "The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation," *Cell Systems*, vol. 9, pp. 297–308.e2, Sept. 2019.

[164] C. Lori, S. Ozaki, S. Steiner, R. Böhm, S. Abel, B. N. Dubey, T. Schirmer, S. Hiller, and U. Jenal, "Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication," *Nature*, vol. 523, pp. 236–239, July 2015.

[165] C. K. Lee, W. C. Schmidt, S. S. Webster, J. W. Chen, G. A. O'Toole, and G. C. L. Wong, "Broadcasting of amplitude- and frequency-modulated c-di-GMP signals facilitates cooperative surface commitment in bacterial lineages," *Proceedings of the National Academy of Sciences*, vol. 119, p. e2112226119, Jan. 2022.

[166] M. A. Konkol, K. M. Blair, and D. B. Kearns, "Plasmid-Encoded ComI Inhibits Competence in the Ancestral 3610 Strain of Bacillus subtilis," *Journal of Bacteriology*, vol. 195, pp. 4085–4093, Sept. 2013.

[167] B.-M. Koo, G. Kritikos, J. D. Farelli, H. Todor, K. Tong, H. Kimsey, I. Wapinski, M. Galardini, A. Cabal, J. M. Peters, A.-B. Hachmann, D. Z. Rudner,

K. N. Allen, A. Typas, and C. A. Gross, “Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis,” *Cell Systems*, vol. 4, pp. 291–305.e7, Mar. 2017.

[168] M. H. Kabir, R. Patrick, J. W. K. Ho, and M. D. O’Connor, “Identification of active signaling pathways by integrating gene expression and protein interaction data,” *BMC Systems Biology*, vol. 12, p. 120, Dec. 2018.

[169] Y. Hu, T. Peng, L. Gao, and K. Tan, “CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data,” *Science Advances*, vol. 7, p. eabf1356, Apr. 2021.

[170] S. Srinivasan, I. D. Vladescu, S. A. Koehler, X. Wang, M. Mani, and S. M. Rubinstein, “Matrix Production and Sporulation in Bacillus subtilis Biofilms Localize to Propagating Wave Fronts,” *Biophysical Journal*, vol. 114, pp. 1490–1498, Mar. 2018.

[171] M. Morrison, M. Razo-Mejia, and R. Phillips, “Reconciling kinetic and thermodynamic models of bacterial transcription,” *PLOS Computational Biology*, vol. 17, p. e1008572, Jan. 2021.

[172] S. Mangan and U. Alon, “Structure and function of the feed-forward loop network motif,” *Proceedings of the National Academy of Sciences*, vol. 100, pp. 11980–11985, Oct. 2003.

[173] W. Kolch, M. Halasz, M. Granovskaya, and B. N. Kholodenko, “The dynamic control of signal transduction networks in cancer cells,” *Nature Reviews Cancer*, vol. 15, pp. 515–527, Sept. 2015.

[174] E. Palomer, J. Buechler, and P. C. Salinas, “Wnt Signaling Deregulation in the Aging and Alzheimer’s Brain,” *Frontiers in Cellular Neuroscience*, vol. 13, 2019.

[175] T. A. Baeumler, A. A. Ahmed, and T. A. Fulga, “Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors,” *Cell Reports*, vol. 20, pp. 2639–2653, Sept. 2017.

[176] Y. Ma, M. W. Budde, M. N. Mayalu, J. Zhu, A. C. Lu, R. M. Murray, and M. B. Elowitz, “Synthetic mammalian signaling circuits for robust cell population control,” *Cell*, vol. 185, pp. 967–979.e12, Mar. 2022.

[177] H.-S. Li, D. V. Israni, K. A. Gagnon, K. A. Gan, M. H. Raymond, J. D. Sander, K. T. Roybal, J. K. Joung, W. W. Wong, and A. S. Khalil, “Multidimensional control of therapeutic human cell function with synthetic gene circuits,” *Science*, vol. 378, pp. 1227–1234, Dec. 2022.

[178] X. Wei, V. G. Henke, C. Strübing, E. B. Brown, and D. E. Clapham, "Real-Time Imaging of Nuclear Permeation by EGFP in Single Intact Cells," *Biophysical Journal*, vol. 84, pp. 1317–1327, Feb. 2003.

[179] N. Hundt, W. Steffen, S. Pathan-Chhatbar, M. H. Taft, and D. J. Manstein, "Load-dependent modulation of non-muscle myosin-2A function by tropomyosin 4.2," *Scientific Reports*, vol. 6, p. 20554, Feb. 2016.

[180] M. Rief, R. S. Rock, A. D. Mehta, M. S. Mooseker, R. E. Cheney, and J. A. Spudich, "Myosin-V stepping kinetics: A molecular model for processivity," *Proceedings of the National Academy of Sciences*, vol. 97, pp. 9482–9486, Aug. 2000.

[181] A. Flamholz, R. Phillips, and R. Milo, "The quantified cell," *Molecular Biology of the Cell*, vol. 25, pp. 3497–3500, Nov. 2014.

[182] S. Preibisch, S. Saalfeld, and P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics*, vol. 25, pp. 1463–1465, June 2009.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]