Multiscale analyses of cellular signaling and regulation in response to multiple stress conditions

By

GW McElfresh

Submitted to the graduate degree program in The Center for Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

_____

Chairperson: Christian Ray, Ph.D.

_____

Joanna Slusky, Ph.D.

_____

Yinglong Miao, Ph.D.

_____

Ilya Vakser, Ph.D.

_____

Josephine Chandler, Ph.D.

Date Defended: 05/26/2020

The dissertation committee for GW McElfresh certifies that this is the approved version of the following dissertation:

Multiscale analyses of cellular signaling and regulation in response to multiple stress conditions

_____

Chairperson: Christian Ray, Ph.D.

Date Approved: 05/26/2020

ABSTRACT

Understanding the relationship between signaling and its corresponding cellular response is critical to combating stress responses, especially responses related antibiotic resistance and non-genetic phenotypic transitions to antibiotic tolerance. However, bacterial signal responses are notoriously noisy and difficult to predict. This work first develops a multiscale cell cycle-aware signal modeling framework to explore the energetics and dynamics of the phosphate starvation stress response two-component system, PhoBR, to better understand the relationship between stress response proteins and the bounds of cellular memory in stress response. I found that the transcription factor responsible for stress response remains nominally "active" for 2-4 generations after the stress response is relieved due to sequestration effects, with differential memory in offspring cells dictated by stochastic protein inheritance. Next, I studied a novel antibiotic persister phenotype that arises in non-canonical conditions. This phenotype exhibited a previously unknown stress response that resulted in growth arrest, granting it antibiotic tolerance. The tolerance seems to be imparted by a global stress response arising from toxic excessive lactose import, seemingly opposite of the starvation response that induces canonical persister cell formation. Finally, I improved the PhoBR stress response model to measure stochastic fluctuations of proteins within the two-component system to identify the principles of signal fluctuations and how they drive variability in the bacterial cell cycle (i.e., growth rate). The downstream regulon of the PhoB response regulator is the main driver of the growth rate, but the transcriptionally active dimerized PhoB acts as the link between fast molecular fluctuations and slower gene expression fluctuations within the system. Finally, I present a vision for future developments of this style of modeling to include spatial information.

To Holden

ACKNOWLEDGMENTS

First, I want to thank Christian for his guidance over these years. I was incredibly lucky to find someone whose values, personally and scientifically, would guide me through my transition to “becoming a biologist” and help me grow as both a scientist and a person. It was truly a pleasure to do science “in the weeds” and I wish there were more scientists like you.

I would also like to thank my committee members: Dr. Ilya Vakser, Dr. Joanna Slusky, Dr. Yinglong Miao and Dr. Josie Chandler. Thank you for your feedback on my projects during seminars and always being receptive to my questions. Dr. Eric Deeds, thank you for our enlightening discussions during your time in Kansas.

I want to thank all of my colleges for their ideas, discussions, advice, and friendship. I am especially thankful to Jaden Anderson, Meghan Franklin, and Nathan Jenkins for their camaraderie, commiseration through classes, and endless patience for my curiosity about their projects. Huijing, Nishantha, Andrew, and Adam, you have been fantastic lab mates and I’ve learned so much from you. Thank you.

To my friends: Ethan, Stewart, Justin, Imba, Justin, Garrett, and Odos, thank you all for providing me a space to relax and laugh. Chris, thank you for being a constant source of encouragement and commiseration. Tyler, thank you for being a brother to me. Shannel, thank you for your unwavering support, kindness, and constant accommodations.

To the influential teachers in my life: Mr. Chris Peterson, Mr. Kent Sameshima, Mrs. Phyllis Farmer, Mrs. Jessica Schober, Dr. Bob Robertson, and Dr. Christos Deligkaris. Thank you all for challenging me, believing in me, and guiding me to this point.

To my Mom and Dad, thank you for your constant love and support and for always encouraging my curiosity. To my grandparents, thank you for instilling a love of science in me and teaching me to nourish it. Katie, thank you for being the best sister anyone could ask for. I'm always learning from you and you've been my role model since day one (literally). One day I will forgive you for cutting up my Pokémon toys, but not yet.

To Holden, thank you for showing me that sometimes the smallest things can mean the most and for always reminding me that the world is full of wonder. To Sarah, thank you for being a wonderful mother to Holden and reminding me that I'm more than the sum of my productivity. To Gauge, thank you for always encouraging me before I left for the lab at 3:00 AM to perform growth experiments. I was always much happier after a quick reminder that you were waiting for me when I got back.

Finally, to Michelle, I couldn't have done it without you. Thank you for your time, patience, energy, compassion, companionship, sick jokes, and countless other acts of kindness that I will always be thankful for. You've helped me accomplish the impossible and then some.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding how bacteria sense and respond to stress is a long-standing goal in biology for therapeutics and synthetic biology. However, stochastic variability in cell to cell gene expression makes it difficult to predict how cells *en masse* will respond to a given amount of stimulus [1, 2]. Further work has been done to quantify the effect that noise has on the highly variable bacterial signal-response networks [2-5], but even the most basic all-or-nothing cellular response schemas are capable of wide distributions of population-scale responses. This dissertation shows the progress made in understanding how *E. coli* responds to various stressors and reveals novel ways to investigate future cellular phenomena.

The first work in this dissertation focuses on understanding signal transference in the *E. coli* PhoBR phosphate starvation response network. PhoBR belongs to a class of signal transduction pathways named two-component system (TCS) in which a typically integral membrane sensor histidine kinase (PhoR) autophosphorylates in response to a stimulus and performs a phosphotransfer to activate a response regulator (PhoB) that acts as a transcription factor for genes that will respond to the stimulus. We investigated and modeled PhoBR because of its role in sensing a critical component of central metabolism, its historical characterization [6-10], and the availability of experiments capturing PhoBR sensory and response behavior [11, 12].

The primary challenge in modeling a biochemical network that is connected to central metabolism is the inherent feedback between the larger cellular scale phenomenon (*e.g.* gene expression, growth rate, cell division times) and the smaller scale molecular fluctuations (Fig. 2.1).

To achieve the multiple scales necessary to model the mesoscale feedback, we use the Doob-Gillespie algorithm [13-16] to simulate the biochemical reaction network of PhoBR and its regulatory role in binding to DNA. The reaction network is iteratively simulated ($\Delta t = 1$ second) and the cellular growth rate is calculated based on resource allocation [12, 17]. This approach allows for a cell to grow until it achieves double its initial size and divide by binomially distributing the molecular species between the daughter cells, which will continue growing.

The second work in this dissertation tackles a non-intuitive antibiotic tolerance in *E. coli*. Understanding bacterial stress responses, and particularly antibiotic tolerance and resistance, are critical to the development of proper infection treatment plans. However, the stochasticity involved in those stress responses makes it difficult to predict how bacteria will react. Antibiotic tolerance in *E. coli* has traditionally been linked to nutrient limitation via the stringent response [18], but our lab discovered an enrichment in tolerant cells in excessive nutrient concentrations [19]. In order to discover the cause of the enriched tolerance, we performed RNA-seq on replicates of cells grown in three concentrations of lactose (0.1 mg/mL, 2.5 mg/mL, and 50 mg/mL) respectively representing starvation, satiety, and toxicity. However, significant growth rate heterogeneity was observed in populations grown in the highest lactose concentration consisting of a large, fast growing subpopulation and a smaller, slower growing subpopulation.

In the third part, I revisited our multiscale cell model with deeper analysis of the timing of molecular fluctuations, how they drive gene expression fluctuations, and how the entire multiscale system alters fluctuations in the cell cycle. The results of this analysis demonstrate that the transcriptionally active gene regulator is the key point linking molecular fluctuations and cell growth-driven fluctuations.

1.1 Sensing and growth in E. coli

1.1.1 Two component systems

The TCS is a common sensory motif in bacteria; *E. coli* has evolved several that respond to different stimuli and modulate diverse responses. A canonical TCS is composed of a typically integral membrane sensor histidine kinase that undergoes a conformational change in order to autophosphorylate in response to a stimulus and perform a phosphotransfer to activate a cognate response regulator. The majority are bifunctional, containing phosphatase activity that deactivates the response regulator as well. The response regulator acts as a transcription factor for genes that will respond to the stimulus. Once the stimulus is removed and the sensor histidine kinase returns to its inactive conformation, the sensor histidine kinase deactivates the response regulator; sometimes it is deactivated by a separate phosphatase protein (as in the bacterial chemotaxis TCS CheAY [20]).

Input-output robustness is a significant feature of the TCS. As previously discussed, bacterial responses in dose-response curves have wide distributions [1, 2], and ATP variation can be as high as a factor of 10 [21] between cells. However, TCSs with a bifunctional sensor histidine kinase can be nearly independent from concentrations of the proteins involved [7, 22]. This input-output robustness has limitations when the response regulator protein is at small concentrations compared to the signal, in which case the response is ATP dependent. This limitation is mitigated by the downstream regulon of a TCS having on the order of 10 binding sites, meaning that the active response regulator quickly achieves regulatory concentrations.

1.1.2 The PhoBR two component system

PhoB and PhoR make up the phosphate stress response starvation two-component system. PhoBR is a typical autoregulated TCS, wherein PhoR is a bifunctional sensor histidine kinase that phosphorylates and dephosphorylates PhoB, which dimerizes in its activated form and acts as a transcription factor for the phosphate response regulon. The regulon consists of phosphate assimilation proteins, such as phosphate channels and alkaline phosphatase.

We chose this system for the multiscale model for several compelling reasons. Because the majority of energetics in the cell depend on maintaining a high level of phosphate for ATP, GTP, etc, the system is critically important for bacterial survival in changing environments. As a result, it has been extensively characterized, including quantitation of each biochemical step, gene regulation studies, and physiological growth studies in *E. coli* [9, 11, 12, 23, 24]. Transcriptional autoregulation in this system creates a multigenerational memory – multiple timescales of effects have already been measured experimentally. Finally, virtually all of these studies have been done at the bulk population level: the effects of single-cell fluctuations in PhoBR were almost completely uncharacterized before the work I present here (Figs. 2.3, 4.4 – 4.9).

1.1.3 Allocation of cellular resources in *E. coli*

Amino acid expenditure in *E. coli* can be partitioned into metabolic proteins, synthesizing ribosomes, increasing biomass, and essential housekeeping proteins [17]. Thus, responding to a stress (*i.e.* expressing specialized genes in order to perform a niche function) comes at a direct cost to one of those partitions. Generally, the resource allocation spent on expressing stress response proteins comes at a cost to the growth rate of the cell as ribosome concentration and the amount of ribosomal proteins are correlated with growth rate [25] and large changes in cellular regulation are inversely correlated with growth rate [26].

1.1.4 Adders, timers, and sizers

E. coli growth rate and division size varies based on growth conditions. Under ideal conditions, *E. coli* will usually double its initial volume, regardless of size, and then divide. These cells are considered “adders” [27]. However, there are alternative growth strategies which have been observed in stressful conditions and other species [28, 29]. Timers are cells that seem to divide on a temporal periodicity and have a linear dependence between birth and division sizes. The last category, sizers, have an inverse linear dependence on birth size in which the population divides at an optimal size, where initially small cells divide later than cells already near the optimal size. At the beginning of these studies, it was unknown if phosphate-starved *E. coli* cells are still adders, or if a different stress-responsive growth strategy is present. Working with colleagues, I analyzed growth trajectories of single cells grown in a microfluidic device in a microscope with an environmental control chamber (Fig 4.1).

Chromosomal replication in different *E. coli* division regimes varies as well, but primarily due to growth rate. In fast growing adder-type cells, the chromosome is likely undergoing multiple replications at once, since the division time of the cells is shorter than the initiation and duplication speed of the chromosome [30].

1.1.5 Bet hedging and population heterogeneity

Bet hedging is a common feature of bacterial populations where a subpopulation of cells phenotypically differs from other genetically identical cells, often through a stochastic switch [31]. In *E. coli*, this behavior can be seen in temporarily growth arrested cells, where the local fitness cost of arresting a few cells can provide resilience against β-lactam antibiotics, granting a global fitness advantage. Bet hedging is related to bistability in the phenotypic landscape.

1.2 Multiscale modeling methodologies

Modeling signaling pathways in *E. coli* is intrinsically difficult. Approaches have been made that seek to simulate biomolecular reactions [2, 32-35] but constricting a model to biomolecular timescales specifically prevents a model from studying a core bacterial signaling paradigm: the varied dose-response distributions in bacterial populations caused by highly noisy cellular environments and low signal fidelity [36-38]. The biochemical models used to study such networks are frequently limited to the molecular scale [2, 32-35, 39], yet bacterial protein degradation is primarily dependent on dilution via cellular growth and division and thus on the cellular scale. To address this, we developed a mesoscale model [40] of the PhoB-PhoR two component system that would provide insights at the molecular, cellular, and population scales that are otherwise unable to be simulated with current computational resources.

1.2.1 Small scale modeling

In order to model the smallest scale of bacterial stress response, biochemical reactions, we created a mean-field/stochastic schema. Our mean field model represents cellular species as a set of ordinary differential equations and allows us to calibrate for experimentally unknown values at a higher scale, where we can fit modeled population growth rates to experimental values. Then, to observe stress responses and potential lineage dependence, we use the calibrated mean-field model parameters and the Doob-Gillespie algorithm [13, 14, 16] to create a stochastic version of the model. The stochastic version operates by simulating cellular reactions in one-second long quasi-constant volume simulations. At the end of each simulation, the change in ATP consumptive species production is calculated and provided as input into a growth function [40, 41] to provide a predicted change in cellular volume and the next second is simulated.

1.2.2 Medium to large scale modeling

Modeling processes that take on the order of seconds to minutes is less established, but they are still able to be modeled quantitatively. Gene transcription and translation are modeled within the Doob-Gillespie algorithm, but the rate constants are slow (10^{-2} – 10^{-5} slower) in comparison to the biomolecular reactions. Cellular division occurs via a binomial partitioning of non-chromosomal cellular species and then the two cells are simulated independent of each other. By simulating longer scale processes, we can predict the cellular costs of operating and expressing the stress response and the effect of inherited stress response protein in cellular lineages.

1.3 Bacterial stress responses

Bacteria have several ways to respond to stress including sporulation and biofilm formation, but an increasingly relevant phenotype is the antibiotic tolerant, or persister, cell. There is significant disagreement within the field for the criteria of a persister cell, but the basic characteristics are tolerance to antibiotics on the order of days, growth arrest, and the ability to recover from growth arrest and resume growing. The persister phenotype can arise from many stimuli, including diauxic shift [42], nutrient limitation [43-45], indole signaling [46], and antibiotic induced growth arrest that can lead to eventual resistance [47].

1.3.1 Stringent Response

The stringent response is a stress response in *E. coli* in which growth and division processes are inhibited in favor of preserving amino acid availability and survival [43-45]. The stringent response is mediated by the alarmone guanosine tetra/pentaphosphate, (p)ppGpp [48, 49]. The mechanisms behind (p)ppGpp's role in the stringent response are still being studied, but (p)ppGpp acts as a modulator for RNA polymerase affinity [50-53] and regulates on the order of hundreds

of genes [44]. Linking generalized antibiotic persistence/tolerance to the stringent response has been controversial [54, 55] but there is evidence for β-lactam tolerance due to growth arrest [56].

1.3.2 Novel antibiotic tolerant bacteria in non-stringent conditions

β-lactam tolerance can be achieved in *E. coli* in excessive amounts of nutrients as well [19]. In the more antibiotic tolerant, high sugar cultures, population-level growth rate was bimodal, with one subpopulation growing and dividing quickly while the other subpopulation was growing much more slowly. This type of heterogeneity, or bet-hedging, [57, 58] is usually suggestive of a bi-stability within the population, but the most closely linked mechanism of persistence (the stringent response) paradoxically occurs at nutrient limitation.

We explored possible mechanisms for this novel antibiotic tolerance by analyzing RNA-Seq data from cultures that were treated with antibiotics to select for the antibiotic tolerant cells. We compared these antibiotically tolerant cultures to each other to discover their differences, and we compared the tolerant cultures to their bulk culture in order to discover potential mechanisms for the transition to antibiotic tolerance in each condition.

1.4 Living systems are not in equilibrium

Finding the minimum amount of energy for cellular decision making is a long-standing goal of cellular and molecular biology. Frequently, inspired by Landauer's principle [59], approaches to understanding cellular decision making assume that the minimum amount of energy required to perform a computation is the energy required to erase the effects of the previous computation. The extension to bacterial physiology is natural: growth-mediated dilution is the erasure of the computation (*i.e.* the expression of signal-response regulated genes), so it is reasonable that extremely efficient cellular processes may be close to Landauer's limit. We use

the bacterial two-component system (TCS) as a minimal model system, since the transmission of signal is comprised of a biologically simple steps: a conformational change in a protein, the autophosphorylation of that protein, a single phosphotransfer reaction to the response regulator, and a dimerization of the active response regulator. In modeling the system, we discovered that the system is significantly out of equilibrium, so that even the resting steady state is well above Landauer's limit. Our analysis of the energetics of signaling in an intact system suggests that a different scale of physiological limitations must be considered rather than the near-equilibrium limits identified previously by physicists.

References

[1] A. Levchenko and I. Nemenman, "Cellular noise and information transmission," (in eng), *Curr Opin Biotechnol,* vol. 28, pp. 156-164, 2014, doi: 10.1016/j.copbio.2014.05.002.

[2] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information Transduction Capacity of Noisy Biochemical Signaling Networks," *Science,* vol. 334, no. 6054, pp. 354-358, 2011, doi: 10.1126/science.1204553.

[3] M. A. Rowland, W. Fontana, and E. J. Deeds, "Crosstalk and competition in signaling networks," (in eng), *Biophys J,* vol. 103, no. 11, pp. 2389-98, Dec 5 2012, doi: 10.1016/j.bpj.2012.10.006.

[4] P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, "Information processing and signal integration in bacterial quorum sensing," (in eng), *Molecular systems biology,* vol. 5, pp. 325-325, 2009, doi: 10.1038/msb.2009.79.

[5] E. Ziv, I. Nemenman, and C. H. Wiggins, "Optimal signal processing in small stochastic biochemical networks," (in eng), *PloS one,* vol. 2, no. 10, pp. e1077-e1077, 2007, doi: 10.1371/journal.pone.0001077.

[6] B. L. Wanner and B. D. Chang, "The phoBR operon in Escherichia coli K-12," *Journal of Bacteriology*, vol. 169, no. 12, pp. 5569-5574, 1987. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC213987/.

[7] E. Batchelor and M. Goulian, "Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system," *Proc Natl Acad Sci U S A,* vol. 100, no. 2, pp. 691-6, Jan 21 2003, doi: 10.1073/pnas.0234782100.

[8] E. J. Capra and M. T. Laub, "Evolution of two-component signal transduction systems," *Annu Rev Microbiol,* vol. 66, pp. 325-47, 2012, doi: 10.1146/annurev-micro-092611-150039.

[9] S. M. Hoffer, H. V. Westerhoff, K. J. Hellingwerf, P. W. Postma, and J. Tommassen, "Autoamplification of a two-component regulatory system results in "learning" behavior," *J Bacteriol,* vol. 183, no. 16, pp. 4914-7, Aug 2001, doi: 10.1128/JB.183.16.4914-4917.2001.

[10] R. Gao and A. M. Stock, "Probing kinase and phosphatase activities of two-component systems in vivo with concentration-dependent phosphorylation profiling," *Proc Natl Acad Sci U S A,* vol. 110, no. 2, pp. 672-7, Jan 8 2013, doi: 10.1073/pnas.1214587110.

[11] R. Gao, K. A. Godfrey, M. A. Sufian, and A. M. Stock, "Counterbalancing Regulation in Response Memory of a Positively Autoregulated Two-Component System," *J Bacteriol,* vol. 199, no. 18, Sep 15 2017, doi: 10.1128/JB.00390-17.

[12] L. W. Marzan and K. Shimizu, "Metabolic regulation of Escherichia coli and its phoB and phoR genes knockout mutants under phosphate and nitrogen limitations as well as at acidic condition," *Microb Cell Fact,* vol. 10, p. 39, May 20 2011, doi: 10.1186/1475-2859-10-39.

[13] J. L. Doob, "Stochastic Processes and Statistics," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 20, no. 6, pp. 376-379, 1934. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1076423/.

[14] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics,* vol. 22, no. 4, pp. 403-434, 1976/12/01/ 1976, doi: https://doi.org/10.1016/0021-9991(76)90041-3.

[15] J. H. Abel, B. Drawert, A. Hellander, and L. R. Petzold, "GillesPy: A Python Package for Stochastic Model Building and Simulation," *IEEE Life Sciences Letters,* vol. 2, no. 3, pp. 35-38, 2016, doi: 10.1109/LLS.2017.2652448.

[16] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry,* vol. 81, no. 25, pp. 2340-2361, 1977/12/01 1977, doi: 10.1021/j100540a008.

[17] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa, "Emergence of robust growth laws from optimal regulation of ribosome synthesis," (in eng), *Mol Syst Biol,* vol. 10, p. 747, Aug 22 2014, doi: 10.15252/msb.20145379.

[18] E. Germain, P. Guiraud, D. Byrne, B. Douzi, M. Djendli, and E. Maisonneuve, "YtfK activates the stringent response by triggering the alarmone synthetase SpoT in Escherichia coli," (in eng), *Nat Commun,* vol. 10, no. 1, p. 5763, Dec 17 2019, doi: 10.1038/s41467-019-13764-4.

[19] J. C. Ray, M. L. Wickersheim, A. P. Jalihal, Y. O. Adeshina, T. F. Cooper, and G. Balazsi, "Cellular Growth Arrest and Persistence from Enzyme Saturation," *PLoS Comput Biol,* vol. 12, no. 3, p. e1004825, Mar 2016, doi: 10.1371/journal.pcbi.1004825.

[20] J. R. Kirby, "Chemotaxis-like regulatory systems: unique roles in diverse bacteria," (in eng), *Annu Rev Microbiol,* vol. 63, pp. 45-59, 2009, doi: 10.1146/annurev.micro.091208.073221.

[21] H. D. Murray, D. A. Schneider, and R. L. Gourse, "Control of rRNA expression by small molecules is dynamic and nonredundant," *Mol Cell,* vol. 12, no. 1, pp. 125-34, Jul 2003, doi: 10.1016/s1097-2765(03)00266-1.

[22] G. Shinar, R. Milo, M. R. Martínez, and U. Alon, "Input output robustness in simple bacterial signaling systems," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 50, pp. 19931-19935, 2007, doi: 10.1073/pnas.0706792104.

[23] A. M. Stock, V. L. Robinson, and P. N. Goudreau, "Two-Component Signal Transduction," *Annual Review of Biochemistry,* vol. 69, no. 1, pp. 183-215, 2000/06/01 2000, doi: 10.1146/annurev.biochem.69.1.183.

[24] J. C. Ray and O. A. Igoshin, "Adaptable functionality of transcriptional feedback in bacterial two-component systems," (in eng), *PLoS Comput Biol,* vol. 6, no. 2, p. e1000676, Feb 12 2010, doi: 10.1371/journal.pcbi.1000676.

[25] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of cell growth and gene expression: origins and consequences," (in eng), *Science,* vol. 330, no. 6007, pp. 1099-102, Nov 19 2010, doi: 10.1126/science.1192588.

[26] K. Potrykus, H. Murphy, N. Philippe, and M. Cashel, "ppGpp is the major source of growth rate control in E. coli," (in eng), *Environ Microbiol,* vol. 13, no. 3, pp. 563-575, Mar 2011, doi: 10.1111/j.1462-2920.2010.02357.x.

[27] P. Wang *et al.*, "Robust growth of Escherichia coli," *Curr Biol,* vol. 20, no. 12, pp. 1099-103, Jun 22 2010, doi: 10.1016/j.cub.2010.04.045.

[28] M. Wallden, D. Fange, E. G. Lundius, O. Baltekin, and J. Elf, "The Synchronization of Replication and Division Cycles in Individual E. coli Cells," *Cell,* vol. 166, no. 3, pp. 729-739, Jul 28 2016, doi: 10.1016/j.cell.2016.06.052.

[29] S. Taheri-Araghi *et al.*, "Cell-Size Control and Homeostasis in Bacteria," *Current Biology,* vol. 25, no. 3, pp. 385-391, 2015, doi: 10.1016/j.cub.2014.12.009.

[30] K. Skarstad, E. Boye, and H. B. Steen, "Timing of initiation of chromosome replication in individual Escherichia coli cells," *The EMBO Journal*, vol. 5, no. 7, pp. 1711-1717, 1986/07/01 1986, doi: 10.1002/j.1460-2075.1986.tb04415.x.
[31] E. Kussell and S. Leibler, "Phenotypic diversity, population growth, and information in fluctuating environments," (in eng), *Science*, vol. 309, no. 5743, pp. 2075-8, Sep 23 2005, doi: 10.1126/science.1114383.
[32] A. C. Barato, D. Hartich, and U. Seifert, "Efficiency of cellular information processing," *New Journal of Physics*, vol. 16, no. 10, p. 103024, 2014/10/16 2014, doi: 10.1088/1367-2630/16/10/103024.
[33] C. C. Govern and P. R. ten Wolde, "Optimal resource allocation in cellular sensing systems," *Proceedings of the National Academy of Sciences*, p. 201411524, 2014, doi: 10.1073/pnas.1411524111.
[34] S. Bo, M. D. Giudice, and A. Celani, "Thermodynamic limits to information harvesting by sensory systems," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 1, p. P01014, 2015/01/09 2015, doi: 10.1088/1742-5468/2015/01/p01014.
[35] D. Hartich, A. C. Barato, and U. Seifert, "Sensory capacity: An information theoretical measure of the performance of a sensor," *Physical Review E*, vol. 93, no. 2, p. 022116, 02/10/ 2016, doi: 10.1103/PhysRevE.93.022116.
[36] R. Suderman, J. A. Bachman, A. Smith, P. K. Sorger, and E. J. Deeds, "Fundamental trade-offs between information flow in single cells and cellular populations," *Proceedings of the National Academy of Sciences*, vol. 114, no. 22, pp. 5755-5760, 2017, doi: 10.1073/pnas.1615660114.
[37] R. Suderman and E. J. Deeds, "Intrinsic limits of information transmission in biochemical signalling motifs," (in eng), *Interface Focus*, vol. 8, no. 6, p. 20180039, Dec 6 2018, doi: 10.1098/rsfs.2018.0039.
[38] J. Paulsson, "Summing up the noise in gene networks," *Nature*, vol. 427, p. 415, 01/29/online 2004, doi: 10.1038/nature02257

https://www.nature.com/articles/nature02257#supplementary-information.
[39] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, "The energy–speed–accuracy trade-off in sensory adaptation," *Nature Physics*, vol. 8, no. 5, pp. 422-428, 2012/05/01 2012, doi: 10.1038/nphys2276.
[40] C. R. GW McElfresh, "Intergenerational Cellular Signal Transfer and Erasure," in *The Interplay of Thermodynamics and Computation in Natural and Artificial Systems*, C. K. D Wolpert, J Grochow, and P and Stadler Eds. Santa Fe, NM: SFI Press, 2019.
[41] A. Bandyopadhyay and C. Ray, "Lineage space and the propensity of bacterial cells to undergo growth transitions," *bioRxiv*, 2018, doi: 10.1101/256123.
[42] S. M. Amato, M. A. Orman, and M. P. Brynildsen, "Metabolic control of persister formation in Escherichia coli," *Mol Cell*, vol. 50, no. 4, pp. 475-87, May 23 2013, doi: 10.1016/j.molcel.2013.04.002.
[43] V. Jain, M. Kumar, and D. Chatterji, "ppGpp: stringent response and survival," *J Microbiol*, vol. 44, no. 1, pp. 1-10, Feb 2006. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/16554711.
[44] M. F. Traxler *et al.*, "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli," (in eng), *Mol Microbiol*, vol. 68, no. 5, pp. 1128-1148, 2008, doi: 10.1111/j.1365-2958.2008.06229.x.

[45] W. A. Haseltine and R. Block, "Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 70, no. 5, pp. 1564-1568, 1973, doi: 10.1073/pnas.70.5.1564.

[46] N. M. Vega, K. R. Allison, A. S. Khalil, and J. J. Collins, "Signaling-mediated bacterial persister formation," *Nat Chem Biol,* vol. 8, no. 5, pp. 431-3, Mar 18 2012, doi: 10.1038/nchembio.915.

[47] T. Dorr, M. Vulic, and K. Lewis, "Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli," (in eng), *PLoS Biol,* vol. 8, no. 2, p. e1000317, Feb 23 2010, doi: 10.1371/journal.pbio.1000317.

[48] M. Cashel, "Regulation of bacterial ppGpp and pppGpp," (in eng), *Annu Rev Microbiol,* vol. 29, pp. 301-18, 1975, doi: 10.1146/annurev.mi.29.100175.001505.

[49] D. Chatterji and A. K. Ojha, "Revisiting the stringent response, ppGpp and starvation signaling," (in eng), *Curr Opin Microbiol,* vol. 4, no. 2, pp. 160-5, Apr 2001, doi: 10.1016/s1369-5274(00)00182-x.

[50] Y. Zuo, Y. Wang, and T. A. Steitz, "The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex," (in eng), *Mol Cell,* vol. 50, no. 3, pp. 430-6, May 9 2013, doi: 10.1016/j.molcel.2013.03.020.

[51] W. Ross, C. E. Vrentas, P. Sanchez-Vazquez, T. Gaal, and R. L. Gourse, "The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation," (in eng), *Molecular cell,* vol. 50, no. 3, pp. 420-429, 2013, doi: 10.1016/j.molcel.2013.03.021.

[52] P. Sanchez-Vazquez, C. N. Dewey, N. Kitten, W. Ross, and R. L. Gourse, "Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase," (in eng), *Proc Natl Acad Sci U S A,* vol. 116, no. 17, pp. 8310-8319, Apr 23 2019, doi: 10.1073/pnas.1819682116.

[53] W. Ross, P. Sanchez-Vazquez, A. Y. Chen, J. H. Lee, H. L. Burgos, and R. L. Gourse, "ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response," (in eng), *Mol Cell,* vol. 62, no. 6, pp. 811-823, Jun 16 2016, doi: 10.1016/j.molcel.2016.04.029.

[54] D. O. Osbourne, V. W. Soo, I. Konieczny, and T. K. Wood, "Polyphosphate, cyclic AMP, guanosine tetraphosphate, and c-di-GMP reduce in vitro Lon activity," *Bioengineered,* vol. 5, no. 4, pp. 264-8, Jul-Aug 2014, doi: 10.4161/bioe.29261.

[55] J. S. Kim and T. K. Wood, "Persistent Persister Misperceptions," *Front Microbiol,* vol. 7, p. 2134, 2016, doi: 10.3389/fmicb.2016.02134.

[56] G. Bokinsky *et al.*, "HipA-Triggered Growth Arrest and β-Lactam Tolerance in Escherichia coli Are Mediated by RelA-Dependent ppGpp Synthesis," *Journal of Bacteriology,* vol. 195, no. 14, pp. 3173-3182, 2013, doi: 10.1128/jb.02210-12.

[57] A. Solopova *et al.*, "Bet-hedging during bacterial diauxic shift," *Proceedings of the National Academy of Sciences,* vol. 111, no. 20, p. 7427, 2014, doi: 10.1073/pnas.1320063111.

[58] J.-W. Veening, W. K. Smits, and O. P. Kuipers, "Bistability, Epigenetics, and Bet-Hedging in Bacteria," *Annual Review of Microbiology,* vol. 62, no. 1, pp. 193-210, 2008/10/01 2008, doi: 10.1146/annurev.micro.62.081307.163002.

[59] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Dev.*, vol. 5, no. 3, pp. 183-191, 1961, doi: 10.1147/rd.53.0183.

Chapter 2

Intergenerational Cellular Signal Transfer and Erasure

GW McElfresh[1]
J. Christian J. Ray[*,1,2]

[*]Corresponding author.
Email: jjray@ku.edu
Tel: 785-864-1506

Address:
[1]Center for Computational Biology
University of Kansas – Lawrence
2030 Becker Dr.
Lawrence, KS 66047

[2]Department of Molecular Biosciences
University of Kansas – Lawrence
1200 Sunnyside Ave.
Lawrence, KS 66045

Abstract

Cellular signaling serves as a means for cells to respond to external stimuli, but the stochasticity involved in the signal to response network questions exactly how sensitive the responses are. We investigate this question by modeling the phosphate stress response two-component system (TCS), a common sensory motif in bacteria composed of a sensor histidine kinase (SHK) and a response regulator (RR), and using empirically based statistical models to allow feedback with larger scale phenomena (growth rate, gene expression, effect of stress response). We implemented this model as both a mean field and stochastic model to observe the bounds of cellular populations: many identical cells recovering from stress as a collective (mean field) and a single stressed cell growing under unstressed conditions to observe its transition into a small colony. From observing a stress shut-off response, we discovered that there were three important time scales for cellular memory of the stress. On the seconds to minutes scale, the SHK responsible for sensing the phosphate starvation was deactivated. However, the RR species took 1-3 generations to cease transducing the signal due to sequestration in the active, dimerized form. The last relevant time scale shows us that the dilution of the stress response regulon is lineage dependent with some lineages diluting the response faster than others but still occurs on the generational scale. Additionally, in this work we summarized the energy costs associated with TCS signaling. The potential difference $\Delta\mu \approx 15k_bT$ varies slightly depending on the level of stress, but the ATP hydrolysis performed by the TCS is very sensitive to the stress.

Biological substrates for computation have been considered since before the advent of modern deterministic computers [1, 2]. Technological advances in measuring cellular responses to molecular signals have again raised the question of how stochastic networks compute.

Signaling pathways enable living cells to process responses to stimuli from the extracellular environment. The uncertainty of signal transmission in a single cell has prompted various research efforts to quantify how much a cell knows about its environment. Advances in nonequilibrium thermodynamics have arrived alongside analyses of biological signaling. Often, models of signaling that consider only the time scale of molecular fluctuations have been considered (see, among others, [3-8]), especially in relation to the bacterial chemotactic response [9].

We suggest that an important time scale for biological signaling should be on the order of gene expression (in the case of bacteria, potentially multiple generations). Growing cells invest energy to grow and divide, thereby diluting the results of previous computations. Because the remnants of previous responses are reduced but not necessarily completely erased, gradual dilution imparts a memory effect: a daughter cell is predisposed to respond in a qualitatively similar manner to its mother cell. Quantifying thermodynamic costs of molecular receptor signaling on short time scales reveals much about the extreme limits of the biological cost of computation, but such energy use is ultimately minor compared to the massive costs of gene expression that can arise as a result of such a signal. Here we seek to explore the effects of those costs on cellular information processing.

We analyze cellular memory in a broad class of bacterial information transfer systems: two-component system (TCS) modules. TCSs respond to information about modulation of the physicochemical environment in and around the cell. Our analysis places nongenetic

intergenerational information transfer in a computational context and raises the question of the appropriate scales for analyzing the thermodynamics of information in living systems.

2.1 Signaling Dynamics on the Time Scale of Generations

Two relevant time scales of cellular signaling responses are molecular kinetic fluctuations and gene expression programs. In a bacterial cell, the time scale of protein turnover (and thus shifts in gene expression) is set by the generation time for the majority of protein types. This is because most proteins are quite stable; the relevant quantity for protein kinetic activity is concentration, and growth of the cell is the fastest process that reduces the concentration. Considering a signal that activates a transcription factor, the loss of the signal depends on the elimination of the responding proteins. Thus, for the mean-field birth–death process with constant production α/τ and constant generation time τ , we have dynamics of protein concentration x(t) as

$$x = e^{-t/\tau}(x(0) + \alpha(e^{t/\tau} - 1)\ ,$$

and protein half-life after loss of signal is τ ln 2, or about 70 percent of a cell's lifetime, due to growth-mediated dilution. Positive feedback on the activation signal can promote the transcriptionally activated state of the cell, further exaggerating the effect. Many studies have explored the implications of such phenotypic memory [10-16].

To make the conditions underlying cost and benefit more concrete, we introduce a common signaling pathway in bacteria: the TCS. Our goal here is to create a biologically realistic model that allows numerical determination of thermodynamic and informational quantities.

2.2 Models of Bacterial Two-Component Signaling

TCSs are a common sensing mechanism in bacteria that have a notable level of conservation across phyla (for a review of TCS evolution, see [17]). Though many variations on the core motif exist, the canonical TCS has a dimeric sensor histidine kinase (SHK) and a cognate response regulator (RR). The sensor responds to stimuli by increasing phosphorylated RR [18]. Once phosphorylated, RR dimerization is stabilized, allowing it to become a transcription factor for genes that are typically relevant to the original stimulus. In many TCSs, one of the operons regulated by the RR is the TCS operon itself, providing feedback and potentially affecting the regulatory activity of the TCS [15, 19-22]. TCS operons have strong gene expression polarity, meaning the expression level of the gene closest to the transcription start site is higher than expression of the subsequent gene(s). Because of this effect, [RR] exceeds [SHK] by orders of magnitude to maintain a sensitive, yet reproducible, response to stimuli [23, 24]. There are multiple distinct TCSs in most characterized bacterial species, each responding to distinct stimuli and inducing distinct responses [25-27]. However, TCSs are integrated into global responses. For example, phosphate limitation depends on a complex between multiple sensors, including a TCS sensor called PhoR [28]. We developed coarse-grained models for the TCS core motif that were parameterized to represent a large class of them approximately, but with special reference to the PhoBR system in Escherichia coli, which has been extensively studied ([29-31], and references therein).

2.3 Coarse-Grained Kinetic Model

The sensor of a TCS is a dimer composed of two inactive monomers. It matures into a dimeric form that is usually in the cell membrane and senses changes in environment. The mature sensor has two reaction pathways: one that favors creating the active regulator and one that favors

the inactivation of the regulator. The result is a dynamic balance between the competing processes of activation and deactivation. Which one dominates at a given time depends on how much stress signal is present. Figure 2.1a depicts this process. Conformational states in figure 2.1a represent ensembles of protein structure conformations that are functionally equivalent in terms of the reaction kinetics, which is why we refer to this as a coarsegrained kinetic model. All of the depicted reaction rates follow mass action kinetics at this scale.

We have inferred that the SHK component of E. coli PhoBR switches between kinase-active and kinase-inactive conformational ensembles because phosphatase activity is unaffected in mutants lacking kinase activity [32]. Because ATP or ADP is bound in close proximity to the phosphorylated site on the SHK, the kinase and phosphotransfer reactions are reversible. The step in a TCS that truly dissipates energy is phosphatase activity: effectively irreversible dephosphorylation of a phosphorylated RR monomer.

The cytoplasm contains ATP at approximately one hundredfold excess over ADP [20]. Our model assumes that ATP quickly replaces ADP in the binding pocket of SHK molecules. SHK reversibly binds its cognate RR. SHK is then capable of reversibly transferring the phosphoryl group to the RR. 5

In the limit of large numbers of molecules, the steady-state fraction of active SHK is $\frac{k_2}{k_2+k_{-2}}$. In this model, we can say that the rate of kinase phosphorylation is $\frac{k_3 k_2}{k_2+k_{-2}}$. We can find the potential difference [33], $\Delta\mu = k_b \mathrm{T} \ln \frac{J^+}{J^-}$, where J^+ represents the flux toward transcriptionally active RRP2, J^- represents the reverse flux toward the inactive state, and k_bT is the Boltzmann constant times the temperature. In the equilibrium state, the two fluxes balance, and we have detailed

balance. Deviations of $\Delta\mu$ from zero quantify how far out of equilibrium the system is being driven by mass and energy input from the rest of the cell.

Our TCS model has the following mean-field fluxes:

$$J^{+} = k_1[\mathrm{SHKm}]^2 \times \frac{k_2 k_3}{k_{-2}+k_2}\,[\mathrm{SHK}] \times (k_{-5} + k_4)[\mathrm{SHK.RRP}] \times k_6[\mathrm{RRP}]^2$$

$$J^{-} = k_{-1}[\mathrm{SHK}] \times k_{-3}[\mathrm{SHK.RRP}] \times (k_5 + k_{-4})[\mathrm{RRP}] \times k_{-6}[\mathrm{RRP_2}],$$

where [SHK_m] represents SHK monomers, [SHK] is SHK dimers, [SHK.RRP] is the SHK + RR complex, [RRP] is phosphorylated RR monomer, and [RRP_2] is transcriptionally active, phosphorylated RR dimer. We have identified specific parameter values for each rate constant that reflect the PhoBR system (table 1).

In practice, living cells constantly produce ATP; the TCS has a constant source of energy in ATP and a sink in ADP + Pi. The gene regulatory activity of the TCS, including its autoregulation, also contributes to the total energy in the system. The steady state of a functional TCS is intrinsically out of equilibrium: $\Delta\mu > 0$.

2.4 Connections to Cellular Physiology

Activation of a TCS upregulates a regulon, the set of genes that are the target of the regulator. The cell pays a metabolic cost for the response, but also benefits from the ameliorative activities of the regulon. For example, in the case of phosphate starvation, the PhoBR TCS induces expression of alkaline phosphatase (phoA), recovering phosphorus from phosphate ester. However, the complete regulon of PhoBR consists of approximately forty upregulated genes; the metabolic cost of expressing it is significant. Lynch and Marinov [34] give a sense of the scale of

a regulon. They estimated the absolute cost per gene to be 103 − 108 hydrolyzed phosphate bonds in bacteria. This is likely to be the majority of the metabolic cost of TCS activation.

We consider the fraction of the growth budget dedicated to the TCS to be $1 - \phi(\rho) = 1 - \chi$ ρ ρmax , where ρ represents the size of the total regulon, ρmax is the maximal hypothetical induction, and χ is the maximal fraction of the growth budget that the regulon can take. We have defined the "growth budget" somewhat amorphously so that we can use $1 - \phi(\rho)$ as a multiplier to limit growth rate. Then we have a growth multiplier that determines the growth benefit from expressing the TCS regulon, f(ρ) + fb, where fb is the basal growth rate without the benefit of the TCS signal, while f(ρ) is the ameliorative contribution of the regulon. We take a linear benefit, f(ρ) = α ρ ρmax , where α + fb represents the maximal growth rate attainable in a given condition without accounting for TCS regulon cost. The net growth rate accounting for both cost and benefit is then

$$\gamma = \big(1 - \phi(\rho)\big)\big(f(\rho) + f_b\big) = (1 - \chi\frac{\rho}{\rho_{max}})(\alpha\frac{\rho}{\rho_{max}} + f_b)$$

The trade-off effect naturally arises because this form is quadratic in ρ, with a predicted optimal regulon size at the point where $\frac{\partial\gamma}{\partial\rho}|_{\chi,\alpha,f_b,\rho_{max}} = 0$, which gives $\rho = \frac{\rho_{max}(\alpha+f_b\chi)}{2\alpha\chi}$.

The situation is not that simple, however, because both α and fb depend on the same conditions that determine the activation state of the TCS: kinetic parameter k2. The relationship could take a variety of forms. We estimated the relationship empirically using biomass in a chemostat experiment in an E. coli strain that has had the phoB gene (response regulator) deleted [35]. This strain does not produce the TCS regulon. Its steady-state biomass in a chemostat at various levels of phosphate starvation therefore gives fb for the case of the PhoBR system. The

biomass data happen to fit an inverse logistic function with $r^2 > 0.999$. Assuming that TCS activation rate, k_2, is proportional to the degree of phosphate starvation in PhoBR, we have $f_b = \frac{a_{fit}}{b_{fit} + e^{-cfit+k_2}}$.

In the preceding equation, $\alpha + f_b$ is the maximum possible recovery from the signal-induced growth rate: with γ_u as the upper limit of the growth rate and ϵ (≤ 1) as the efficiency of the regulon to recover the growth rate,

$$\alpha + f_b = \epsilon \left(\gamma_u - \frac{a_{fit}}{b_{fit} + e^{-cfit+k_2}} \right) + \frac{a_{fit}}{b_{fit} + e^{-cfit+k_2}}.$$

For the PhoBR system, we have a growth model with free parameters ϵ, γ_u, χ, and ρ_{max}. The same study that gave data for the logistic fit of fb (Marzan and Shimizu 2011) also measured relative expression of selected PhoBR regulon genes in wildtype cultures. From this, we estimated $\chi \approx 0.37$. We assumed that the genes upregulated by the TCS were mostly capable of reducing phosphate stress ($\epsilon = 0.95$) and that the growth medium without phosphate starvation is relatively favorable (γ_u = 0.0004/s = 1.44/h). The hypothetical maximum induction of the regulon (ρ_{max} = 150 μM) was set by calibration with the average regulon transcription and translation rates, k_{txnR} and k_{tsnR} (table 1).

Using the growth model, we create two multiscale models of a TCS embedded in cellular physiology—one representing the average of many cells and a stochastic simulation that tracks the dynamics of signaling in single cells. We first describe the mean-field model TCS dynamics. We then use this to develop a stochastic model. We calibrate both models with the meanfield model, explore average responses with it, and then use the stochastic model to simulate the dynamics of signal transfer as the population recovers from signal loss.

2.5 Mean-Field Model

We represent two types of processes: reversible chemical reactions and irreversible reactions that represent dissipative processes, such as transcription, translation, cellular growth, and, in the TCS, the irreversible step in hydrolysis of ATP—phosphatase activity of SHK. We allow transcription and translation to be governed by mass-action kinetics. The complete model is a set of differential equations with twelve variables: bicistronic messenger RNA (mRNA), monocistronic RR mRNA, downstream regulon mRNA, downstream regulon protein, and the species represented in figure 2.1a.

The equations are omitted for brevity, but all interactions are assumed to be mass action, except for gene regulation processes, which take Michaelis–Menten form, with V_{max} given by k_{txn} (TCS operon) or k_{txnR} (regulon operons) and K_m given by K_{mtxn} (table 1). (We assume that most promoters of the TCS regulon are calibrated to typical concentrations of RRP_2, and therefore we assign the same K_m for the TCS and all regulon promoters.) mRNA is unstable and actively degraded by cells; degradation of mRNA is taken to be a massaction process. On the basis of the work of Aiso and Ohki [36], our model has an unstable bicistronic TCS mRNA species capable of initiating translation of both RR and SHK as well as a more stable monocistronic mRNA species that only initiates translation of RR. Dilution of molecules depends on the previously described growth model: loss of protein has a rate $\gamma[Protein] = (1 - \phi(\rho))(f(\rho) + fb) \times [Protein]$, and loss of mRNA has a rate $(k_{degRNA} + \gamma) \times [mRNA]$ for degradation rate constants that depend on the specific mRNA species.

2.6 Stochastic Model

The stochastic model is based on the mean-field model with the following additions. Reactions occur in individual cell agents that have a volume growing according to the growth model described earlier, based on Bandyopadhyay, Wang, and Ray [27]. Increments of stochastic simulation occur at approximately constant volume intervals, then the volume is updated based on the resulting growth rate. Increments in reaction volume affect any bimolecular interactions [27]. We chose a quasi-constant volume interval of 1 s, which is less than the expected time to add a single phospholipid in a cell that is growing relatively quickly.

For the stochastic growth model, we assume that the mean-field growth model holds, with the exception of regulon fluctuations. The downstream regulon of a TCS potentially undergoes significant fluctuations that are entrained to RRP_2 fluctuations. However, there is still an independent stochastic component: between the expression of multiple genes, upward fluctuations in expression in some genes may be counterbalanced by downward fluctuations in expression in other genes. We therefore represent gene expression from n = 40 independent loci, all assumed to have identical binding and gene expression kinetics, producing mRNA into a common pool that results in a common regulon.

In the stochastic model, we represent explicit promoters for the regulated genes, with binding/unbinding and irreversible transcription initiation events. We set the binding constants and transcription initiation constants to be equal to the Michaelis–Menten form of the mean-field model (Table 1).

Each cell agent grows at a rate set by the growth model (γ), and, when the initial volume has been doubled, it divides, partitioning all non-DNA species into two daughter cells with a binomial distribution. Jun et al. [37] (and references therein) suggest that the “adder” principle is an excellent phenomenological representation of cell volumes during the E. coli cell cycle: a

constant cell volume is added before division. In our model, each cell agent has a volume of 1 femtoliter and doubles to 2 femtoliters before division. Promoters/DNA are all deterministically inherited into both daughter cells. The cellular simulation is implemented in Python, with the stochastic simulations run in StochKit and using GillesPy [28] to interface the Python cell script with the stochastic simulations.

2.7 Results

2.7.1 Average Cellular Growth and Signal Dynamics

Simulations using the mean-field model reveal the effects of induction and shutoff of a TCS in E. coli (Fig. 2.2). The model suggests that intermediate levels of induction have a slightly lower growth rate than the fully induced system when the stress becomes more severe (Fig. 2.2c). The reason for the effect is clear looking at the model variants lacking transcriptional feedback with constant low and high TCS gene expression (shaded lines in figure 2.2). The system with transcriptional feedback switches from being nearly equivalent to the low-TCS expression feedbackless case to being nearly equivalent to the high-TCS-expression feedbackless case. It is the transcriptional feedback that allows the system to adapt to higher signal levels. Constant high TCS expression causes grossly more ATP hydrolysis (which is the same as the phosphatase flux, $k_{-3} \times ([SHKa.RRP] + [SHK.RRP])$) than the case with low expression or transcriptional feedback (Fig. 2.2d). This demonstrates a trade-off between cost and benefit: in the autoregulated TCS, it is possible to sacrifice large investments in stress responses at the cost of a slightly lower growth rate, unless the stress becomes severe.

Our model predicts that the TCS has a potential difference $\Delta\mu \approx 15\ k_bT$, varying slightly depending on signal level (Fig. 2.2b). The same is not true for the ATP dissipation rate of the TCS, which increases dramatically at the largest induction levels (Fig. 2.2d).

We find that shutting off the signal (initial conditions at the $k_2 = 10$ steady state, instantaneously switched to $k_2 = 10^{-3}$) reveals three relevant time scales (Figs. 2.2e–2.2f). On the generational time scale, ≫1000 s, the regulon is diluted, and normal growth resumes. Loss of TCS transcriptional activity (RRP_2) occurs at a faster rate. At very short time scales, the sensor shuts off to an intermediate quasi–steady state before being driven even lower by the effects of growth dilution (Fig. 2.2f).

2.7.2 Intergenerational Signal Transfer in A Two-Component System

The stochastic cell growth framework captures the rate of signal loss and the interaction between cell division and dynamics of the signal (Fig. 2.3). We used the same switch from high to low signal as earlier. Figure 2.3a confirms the mean-field results that signal shutoff is faster than loss of the regulon. The half-life of RRP_2 is less than half of a generation, the regulon half-life is more than one generation, and a purely growth-diluted molecule half-life would be ≈70 percent of a generation. Both species follow nearly deterministic trajectories. The same is not true for TCS total protein expression, where protein dilution is highly lineage dependent (Fig. 2.3b). The difference in time scales between the signal shutoff and the residual response illustrates an intergenerational memory effect.

2.8 Discussion

It is increasingly feasible to model time scales of cellular information processing that are relevant to fitness and evolution without them being oversimplified toy models. The disadvantage of this approach is the loss of generality: the necessary quantity of empirical information requires that they simulate a specific system. This is a small problem in the face of ever-increasing high-resolution physiological data. The ability to capture interactions accurately between the short time

scales of molecular fluctuations and the global physiological shifts in a cell is an unmistakable advantage. Here we have demonstrated how such models can address questions of energetics and cellular information processing, setting up a framework for future, more thorough studies of information flow.

Our model of TCS suggests that the vast majority of metabolic (ATP) cost lies in the production of the regulon, which has a higher ATP investment compared to the signaling system itself. Monte Carlo sampling of the TCS kinetic parameters shows that our empirical parameter set lies in the middle of possible responses (not shown). Though not precisely quantitative of any particular system, the numerical results are reliable.

In the intact system, the constant source of ATP along with material influx of TCS proteins maintains the TCS out of thermodynamic equilibrium in all conditions (Fig. 2.2b). At the same time, the system is driven by global physiological variables coupled to stochasticity effects, which diversifies the level of memory in a lineage-dependent manner: some cells and all of their daughters undergo rapid loss of TCS proteins, while other cells maintain a longer-lived high expression level that may be metastable (Fig. 2.3).

Table 2.1. Calibrated parameters

Parameter	Estimated value	Notes/reference
k_1	$10\ (\mu\text{M s})^{-1}$	Fast SHK dimerization
k_{-1}	$0.00001\ \text{s}^{-1}$	Rare SHK dedimerization
k_2	$\text{Conditional}, \text{s}^{-1}$	$k_2 \in [0.001, 10]$
k_{-2}	$0.1\ \text{s}^{-1}$	Assumed fast
k_3	$0.0004\ (\mu\text{M s})^{-1}$	Model calibration
k_{-3}	$0.0087\ \text{s}^{-1}$	[38]
$k_4$and k_{-5}	$1\ \text{s}^{-1}$	Model calibration
k_{-4} and k_5	$0.036\ (\mu\text{M s})^{-1}$	[20]
k_6	$1\ (\mu M\ \text{s})^{-1}$	Model calibration
k_{-6}	$4\ \text{s}^{-1}$	[20]
k_{txnb}	$0.00001\ \text{s}^{-1}$	Model calibration
k_{txn}	$0.00025\ \text{s}^{-1}$	Model calibration
k_{txni}	$0.15\ \text{s}^{-1}$	TCS transcription initiation rate when RRP_2 is bound
K_{mtxn}	$2.5\ \mu\text{M}$	P_{TCS} half-sat; [32]
k_{pb}	$1.66\ (\mu\text{M s})^{-1}$	P_{TCS} binding rate; [32]
k_{pu}	$3.86\ \text{s}^{-1}$	P_{TCS} unbinding rate; inferred from K_{mtxn} and k_{pb}
k_{degb}	$0.027\ \text{s}^{-1}$	[32]
k_{degr}	$0.0044\ \text{s}^{-1}$	[32]
k_{tsn}	$0.05\ \text{s}^{-1}$	Model calibration
χ	0.37	[32]
a_{fit}	$\approx 1.123\ \times 10^{-4}$	[39]
b_{fit}	≈ 1.77	[39]
c_{fit}	≈ 3.75	[40]

Note. P_{TCS} refers to the promoter of the two-component system operon

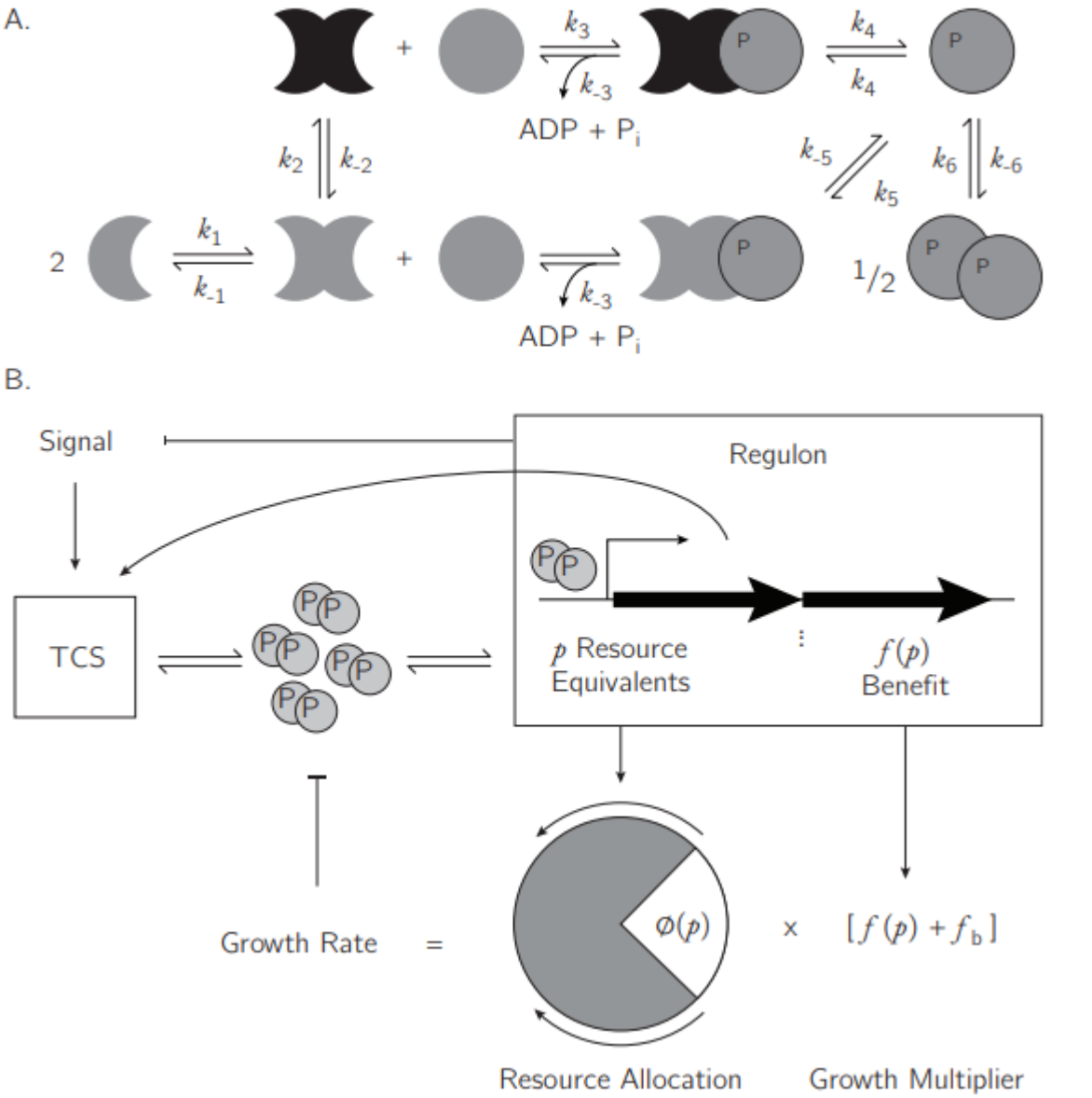

Figure 2.1*:* Coarse-grained multiscale model of a two-component system (TCS). (a) ATP associates with the sensor histidine kinase (SHK), , along with fast interchange between ADP and ATP. External signal stimulates the SHK conformational switch (k_2). Physical interaction between SHK and response regulator (RR) allows phosphotransfer to RR, stabilizing the dimeric RRP_2, an active transcription factor. SHK phosphatase activity is an ATP dissipative step. (b) Nested feedback loops involved in signals from TCSs. Signal stimulates production of RRP_2, which modulates a regulon (upregulation of several genes). Often, transcription of the TCS operon itself is induced: feedback that may affect the signal level. The regulon typically counteracts the signal, another feedback loop. Expression of the regulon entails a metabolic investment, reducing the fraction of resources devoted to growth. Growth dilutes the molecules, affecting bimolecular reaction propensities. The TCS maintains responsiveness by constantly dissipating ATP energy, but the major cost of the TCS during the signal is in the regulon.

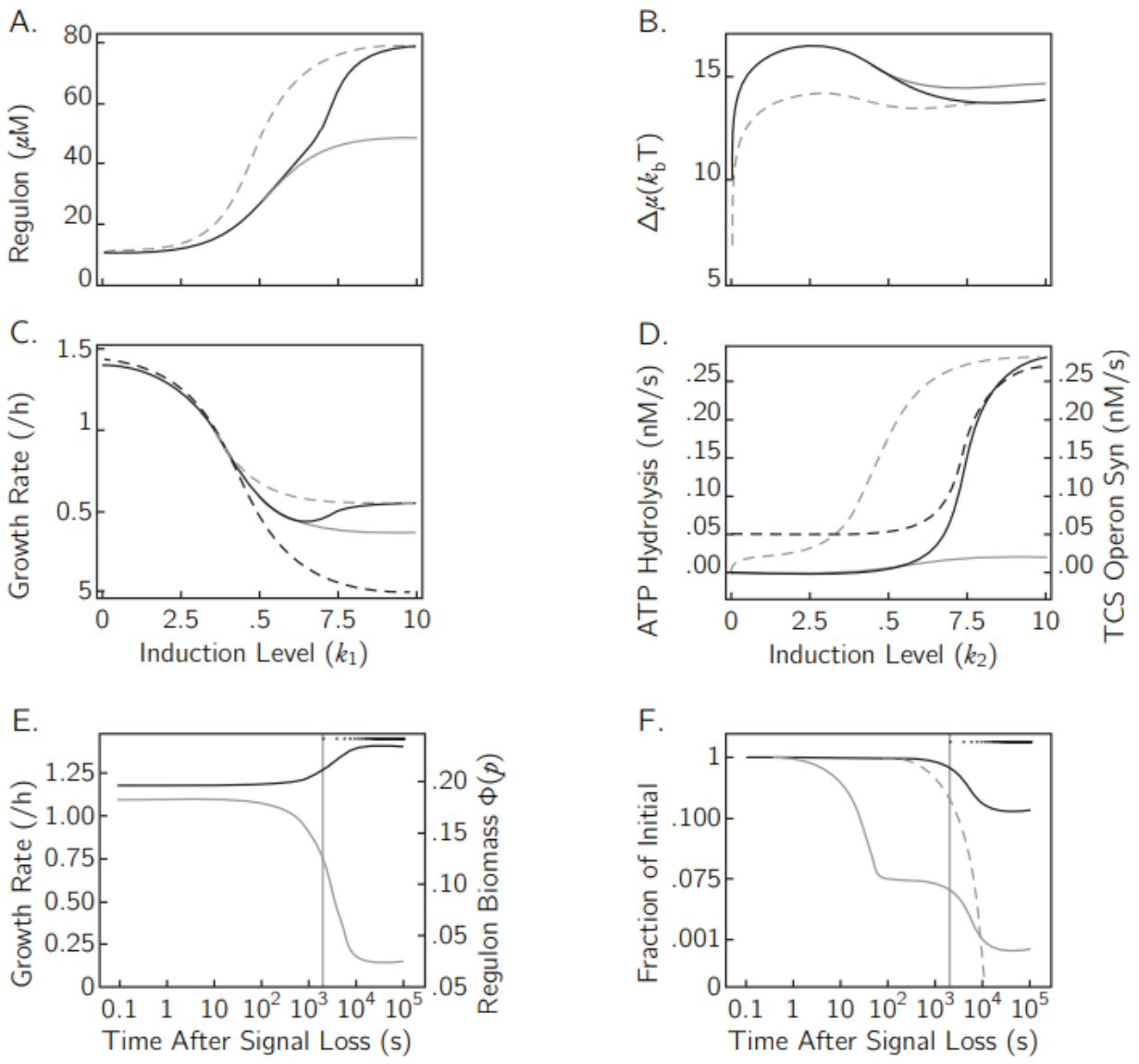

Figure 2.2*:* Predicted steady-state and dynamical physiological outcomes of activating a TCS, with parameters calibrated to represent the E. coli PhoBR system having a regulon containing approximately forty genes. Line styles represent the intact system (black solid) and transcriptional feedbackless system with basal (shaded solid) or maximal (shaded dashed) expression. (a) The level of induction is related to the size of the signal—in this case, phosphate limitation. (b) Potential difference Δμ in units of k_bT. (c) TCS induction recovers a fraction of the growth rate lost to the stress condition. (d) Rates of ATP hydrolysis by the TCS ($k_{-3}([SHKa.RRP] + [SHK.RRP])$) and TCS operon synthesis ($k_{txnb} + \frac{k_{txn}[RRP_2]}{K_{mtxn}+[RRP2]}$). (e) Dynamics of growth rate and biomass on recovery from fully induced to uninduced conditions. The solid black line is growth rate; the shaded line is biomass; and the vertical line is the tcrst generation of growth. Dots represent subsequent generations. (f) Dynamic loss of TCS activity. The solid black line is the regulon; the shaded solid line is SHKa; and the shaded dashed line is RRP_2.

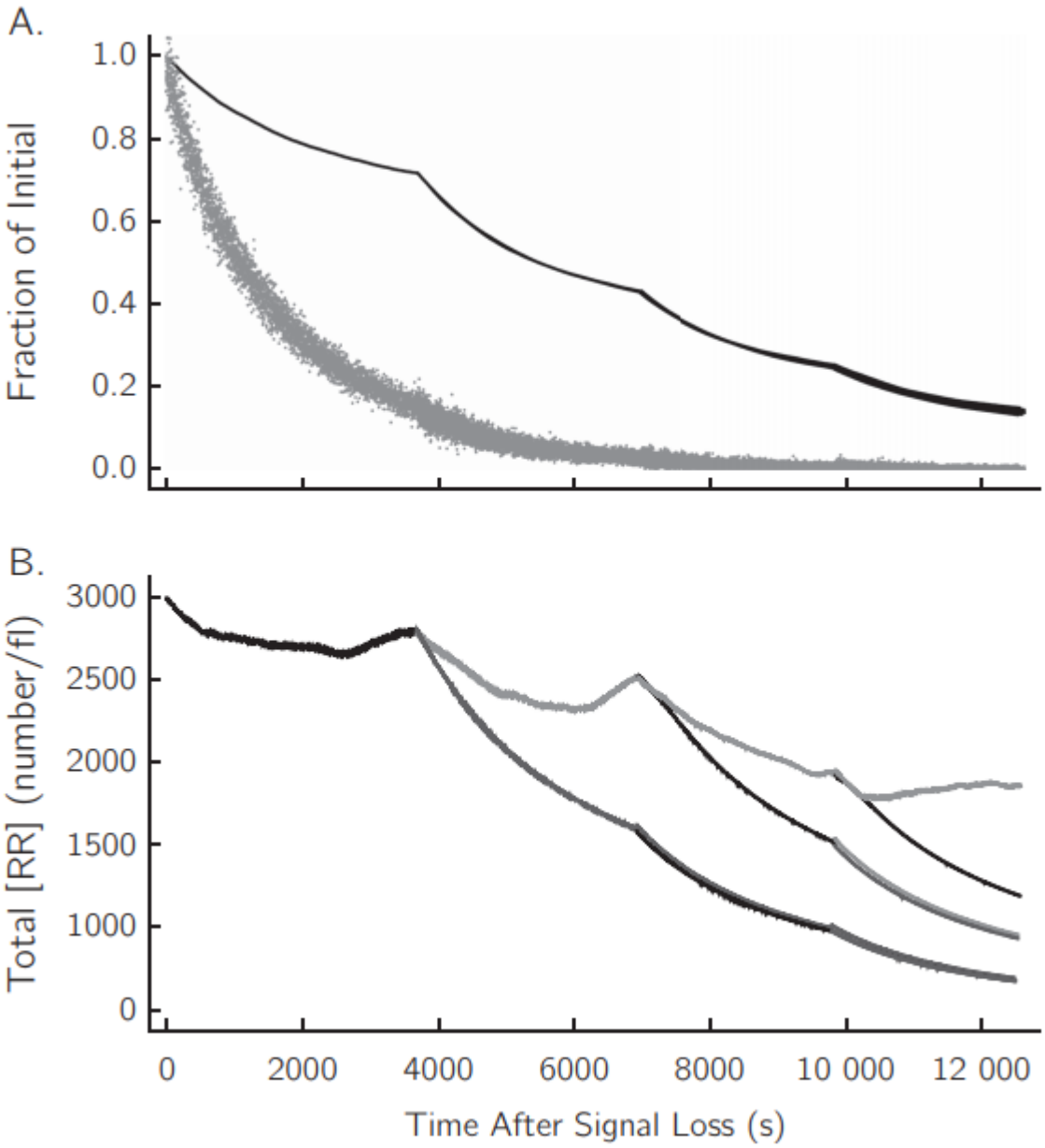

Figure 2.3*:* Stochastic dynamics of intergenerational signal loss. The simulation was started in a high activated steady state ($k_2 = 10$) and allowed to relax to an inactive state ($k_2 = 10^{-3}$). (a) Levels of the TCS regulon (black line) and transcriptionally active regulator RRP_2 (shaded line) follow different time scales. Cell division times are evident. These results superimpose the levels in all of the cell agents in the simulation. The results closely follow the expected deterministic mean. (b) Levels of the TCS proteins display striking heterogeneity that arises at cell birth. Different individual cells are represented by different shades of gray or black. Some cells are nearly to basal levels of the protein, while others still have substantial residual protein several generations later

References

[1] A. C. Barato, D. Hartich, and U. Seifert, "Efficiency of cellular information processing," *New Journal of Physics,* vol. 16, no. 10, p. 103024, 2014/10/16 2014, doi: 10.1088/1367-2630/16/10/103024.

[2] C. C. Govern and P. R. ten Wolde, "Optimal resource allocation in cellular sensing systems," *Proceedings of the National Academy of Sciences,* p. 201411524, 2014, doi: 10.1073/pnas.1411524111.

[3] S. Bo, M. D. Giudice, and A. Celani, "Thermodynamic limits to information harvesting by sensory systems," *Journal of Statistical Mechanics: Theory and Experiment,* vol. 2015, no. 1, p. P01014, 2015/01/09 2015, doi: 10.1088/1742-5468/2015/01/p01014.

[4] D. Hartich, A. C. Barato, and U. Seifert, "Sensory capacity: An information theoretical measure of the performance of a sensor," *Physical Review E,* vol. 93, no. 2, p. 022116, 02/10/ 2016, doi: 10.1103/PhysRevE.93.022116.

[5] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, "The energy–speed–accuracy trade-off in sensory adaptation," *Nature Physics,* vol. 8, no. 5, pp. 422-428, 2012/05/01 2012, doi: 10.1038/nphys2276.

[6] D. Nevozhay, R. M. Adams, E. Van Itallie, M. R. Bennett, and G. Balázsi, "Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit," *PLOS Computational Biology,* vol. 8, no. 4, p. e1002480, 2012, doi: 10.1371/journal.pcbi.1002480.

[7] B. B. Kaufmann, Q. Yang, J. T. Mettetal, and A. van Oudenaarden, "Heritable Stochastic Switching Revealed by Single-Cell Genealogy," *PLOS Biology,* vol. 5, no. 9, p. e239, 2007, doi: 10.1371/journal.pbio.0050239.

[8] P. L. Frick, B. B. Paudel, D. R. Tyson, and V. Quaranta, "Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation," (in eng), *J Cell Physiol,* vol. 230, no. 7, pp. 1403-12, Jul 2015, doi: 10.1002/jcp.24888.

[9] D. R. Burrill, M. C. Inniss, P. M. Boyle, and P. A. Silver, "Synthetic memory circuits for tracking human cell fate," (in eng), *Genes Dev,* vol. 26, no. 13, pp. 1486-97, Jul 1 2012, doi: 10.1101/gad.189035.112.

[10] M. C. Inniss and P. A. Silver, "Building synthetic memory," (in eng), *Curr Biol,* vol. 23, no. 17, pp. R812-6, Sep 9 2013, doi: 10.1016/j.cub.2013.06.047.

[11] G. Lambert and E. Kussell, "Memory and fitness optimization of bacteria under fluctuating environments," *PLoS Genet,* vol. 10, no. 9, p. e1004556, Sep 2014, doi: 10.1371/journal.pgen.1004556.

[12] J. C. Ray, "Survival of Phenotypic Information during Cellular Growth Transitions," (in eng), *ACS Synth Biol,* vol. 5, no. 8, pp. 810-6, Aug 19 2016, doi: 10.1021/acssynbio.5b00229.

[13] E. J. Capra and M. T. Laub, "Evolution of two-component signal transduction systems," *Annu Rev Microbiol,* vol. 66, pp. 325-47, 2012, doi: 10.1146/annurev-micro-092611-150039.

[14] C. P. Zschiedrich, V. Keidel, and H. Szurmant, "Molecular Mechanisms of Two-Component Signal Transduction," (in eng), *J Mol Biol,* vol. 428, no. 19, pp. 3752-75, Sep 25 2016, doi: 10.1016/j.jmb.2016.08.003.

[15] E. Batchelor and M. Goulian, "Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system," *Proc Natl Acad Sci U S A,* vol. 100, no. 2, pp. 691-6, Jan 21 2003, doi: 10.1073/pnas.0234782100.

[16] D. Shin, E. J. Lee, H. Huang, and E. A. Groisman, "A positive feedback loop promotes transcription surge that jump-starts Salmonella virulence circuit," (in eng), *Science,* vol. 314, no. 5805, pp. 1607-9, Dec 8 2006, doi: 10.1126/science.1134930.

[17] G. Shinar, R. Milo, M. R. Martínez, and U. Alon, "Input output robustness in simple bacterial signaling systems," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 50, pp. 19931-19935, 2007, doi: 10.1073/pnas.0706792104.

[18] E. S. Groban, E. J. Clarke, H. M. Salis, S. M. Miller, and C. A. Voigt, "Kinetic buffering of cross talk between bacterial two-component sensors," (in eng), *J Mol Biol,* vol. 390, no. 3, pp. 380-93, Jul 17 2009, doi: 10.1016/j.jmb.2009.05.007.

[19] J. C. Ray and O. A. Igoshin, "Adaptable functionality of transcriptional feedback in bacterial two-component systems," (in eng), *PLoS Comput Biol,* vol. 6, no. 2, p. e1000676, Feb 12 2010, doi: 10.1371/journal.pcbi.1000676.

[20] T. Aiso and R. Ohki, "Instability of sensory histidine kinase mRNAs in Escherichia coli," (in eng), *Genes Cells,* vol. 8, no. 2, pp. 179-87, Feb 2003, doi: 10.1046/j.1365-2443.2003.00624.x.

[21] J. M. Skerker *et al.*, "Rewiring the specificity of two-component signal transduction systems," (in eng), *Cell,* vol. 133, no. 6, pp. 1043-54, Jun 13 2008, doi: 10.1016/j.cell.2008.04.040.

[22] M. T. Laub and M. Goulian, "Specificity in Two-Component Signal Transduction Pathways," *Annual Review of Genetics,* vol. 41, no. 1, pp. 121-145, 2007/12/01 2007, doi: 10.1146/annurev.genet.41.042007.170548.

[23] M. A. Rowland and E. J. Deeds, "Crosstalk and the evolution of specificity in two-component signaling," *Proceedings of the National Academy of Sciences,* vol. 111, no. 15, p. 5550, 2014, doi: 10.1073/pnas.1317178111.

[24] S. G. Gardner, K. D. Johns, R. Tanner, and W. R. McCleary, "The PhoU protein from Escherichia coli interacts with PhoR, PstB, and metals to form a phosphate-signaling complex at the membrane," (in eng), *J Bacteriol,* vol. 196, no. 9, pp. 1741-52, May 2014, doi: 10.1128/jb.00029-14.

[25] S. M. Hoffer, H. V. Westerhoff, K. J. Hellingwerf, P. W. Postma, and J. Tommassen, "Autoamplification of a two-component regulatory system results in "learning" behavior," *J Bacteriol,* vol. 183, no. 16, pp. 4914-7, Aug 2001, doi: 10.1128/JB.183.16.4914-4917.2001.

[26] R. Gao, K. A. Godfrey, M. A. Sufian, and A. M. Stock, "Counterbalancing Regulation in Response Memory of a Positively Autoregulated Two-Component System," *J Bacteriol,* vol. 199, no. 18, Sep 15 2017, doi: 10.1128/JB.00390-17.

[27] R. Gao and A. M. Stock, "Probing kinase and phosphatase activities of two-component systems in vivo with concentration-dependent phosphorylation profiling," *Proc Natl Acad Sci U S A,* vol. 110, no. 2, pp. 672-7, Jan 8 2013, doi: 10.1073/pnas.1214587110.

[28] R. Gao and A. M. Stock, "Temporal hierarchy of gene expression mediated by transcription factor binding affinity and activation dynamics," (in eng), *mBio,* vol. 6, no. 3, pp. e00686-15, May 26 2015, doi: 10.1128/mBio.00686-15.

[29] D. O. Carmany, K. Hollingsworth, and W. R. McCleary, "Genetic and biochemical studies of phosphatase activity of PhoR," (in eng), *J Bacteriol,* vol. 185, no. 3, pp. 1112-5, Feb 2003, doi: 10.1128/jb.185.3.1112-1115.2003.

[30] H. Qian, "Phosphorylation energy hypothesis: open chemical systems and their biological functions," (in eng), *Annu Rev Phys Chem,* vol. 58, pp. 113-42, 2007, doi: 10.1146/annurev.physchem.58.032806.104550.
[31] M. Lynch and G. K. Marinov, "The bioenergetic costs of a gene," (in eng), *Proc Natl Acad Sci U S A,* vol. 112, no. 51, pp. 15690-5, Dec 22 2015, doi: 10.1073/pnas.1514974112.
[32] L. W. Marzan and K. Shimizu, "Metabolic regulation of Escherichia coli and its phoB and phoR genes knockout mutants under phosphate and nitrogen limitations as well as at acidic condition," *Microb Cell Fact,* vol. 10, p. 39, May 20 2011, doi: 10.1186/1475-2859-10-39.
[33] A. Bandyopadhyay and C. Ray, "Lineage space and the propensity of bacterial cells to undergo growth transitions," *bioRxiv,* 2018, doi: 10.1101/256123.
[34] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics,* vol. 22, no. 4, pp. 403-434, 1976/12/01/ 1976, doi: https://doi.org/10.1016/0021-9991(76)90041-3.
[35] S. Jun, F. Si, R. Pugatch, and M. Scott, "Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: a review," (in eng), *Rep Prog Phys,* vol. 81, no. 5, p. 056601, May 2018, doi: 10.1088/1361-6633/aaa628.
[36] J. H. Abel, B. Drawert, A. Hellander, and L. R. Petzold, "GillesPy: A Python Package for Stochastic Model Building and Simulation," *IEEE Life Sciences Letters,* vol. 2, no. 3, pp. 35-38, 2016, doi: 10.1109/LLS.2017.2652448.
[37] T. R. Mack, R. Gao, and A. M. Stock, "Probing the roles of the two different dimers mediated by the receiver domain of the response regulator PhoB," (in eng), *J Mol Biol,* vol. 389, no. 2, pp. 349-64, Jun 5 2009, doi: 10.1016/j.jmb.2009.04.014.
[38] J. Elf, G. W. Li, and X. S. Xie, "Probing transcription factor dynamics at the single-molecule level in a living cell," (in eng), *Science,* vol. 316, no. 5828, pp. 1191-4, May 25 2007, doi: 10.1126/science.1141967.
[39] A. Levchenko and I. Nemenman, "Cellular noise and information transmission," (in eng), *Curr Opin Biotechnol,* vol. 28, pp. 156-164, 2014, doi: 10.1016/j.copbio.2014.05.002.
[40] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information Transduction Capacity of Noisy Biochemical Signaling Networks," *Science,* vol. 334, no. 6054, pp. 354-358, 2011, doi: 10.1126/science.1204553.

Chapter 3

A core conserved bacterial stimulon in response to divergent stresses

Huijing Wang[1,*], GW McElfresh[1,*], Nishantha Wijesuriya[2], Adam Podgorny[1], Andrew D. Hecht[2,†], J. Christian J. Ray[1,2,‡]

[*]Equal contributions.
[†]Current address: Department of Bioengineering, Rice University, Houston, TX
[‡]Corresponding author.

Address:
[1]Center for Computational Biology
[2]Department of Molecular Biosciences
University of Kansas
2030 Becker Dr.
Lawrence, KS 66047

Email: jjray@ku.edu
Tel: 785-864-1506

Abstract

Responses to stress can cause a similar overall reduction in population growth rate in bacteria even if the stresses are different. Lactose as a sole carbon source for *Escherichia coli* is an established example: too little causes starvation, and too much causes toxicity. *E. coli* B strains are robust to osmotic and ionic differences in growth media. The resulting growth dynamic is heterogeneous in lactose-toxified conditions as compared to starving or non-toxified conditions. Both toxicity and starvation cause reduced population growth rates and enrichment for antibiotic-tolerant persister cells in comparison to less stressful intermediate concentrations of lactose. Similarities between starvation and toxification raise the question of how the global stress response stimulon differs between opposite types of stress. We hypothesized that a core generalized stress response is conserved between the two conditions, but that the average cells in starvation conditions would have a key difference in gene expression indicative of the stringent response. To test this, we performed RNA-Seq in three representative conditions for differential expression analysis. In comparison to the non-starving, non-toxified cultures, both showed global shifts in gene expression, though the number of differentially regulated genes is higher in starving conditions. Together with emerging evidence from other studies, our results suggest that many possible pathways can contribute to stress responses and persistence. Therefore, we propose imagining phenomena such as persistence as a set of responses without a single unifying mechanism, but rather as a set of overlapping responses. Our approach provides a resource for identifying diverse mechanisms of stress tolerance from diverse stimuli.

3.1. **Introduction**

Fitness and survival of single-celled organisms in diverse environments is a nuanced topic: a functional response in one environment may incur penalties in another. Mesophilic bacteria encounter this problem: if temperature is too high or too low, different stress responses can arise in opposite conditions. Similarly, osmotic pressure, chemical concentrations, and other conditions typically require a stress response if they get either too high or too low. The primary mechanism attributed to non-genetic antibiotic tolerance in the form of persister cells is starvation or loss of metabolic activity [1-5], but multiple mechanisms induce persister formation [6-10]. This is a chief concern for how to handle infectious disease during the rise of antibiotic resistance because persistence may allow resistance to evolve more quickly [11, 12].

We previously discovered that excessively high or low lactose concentrations (as a sole carbon source) can predispose *Escherichia coli* B REL606 populations to lowered death rates in antibiotics [13]. Varying the concentration of lactose as the sole carbon source in minimal media results in a non-monotonic population growth rate, from a slow uniform growth to a plateau at intermediate concentrations and drastically reduced growth rate at higher concentrations [13]. Lactose has established toxic effects on *E. coli* cultures, often attributed to membrane depolarization via excessive proton symport with lactose through the permease LacY of the major facilitator superfamily [14-16]. In B REL606, toxic lactose levels create a heterogeneous population dynamic with a chance of fast-growing cells to enter growth arrest [13, 17]. We believe that these growth-arrested cells represent a persister-prone subpopulation. This system provides a model for examining how cells respond to conflicting, or opposite, stresses: is the global transcriptional program overall the same, or fundamentally different? In this model system, both

starving and toxified cultures exhibit increased stress tolerance (Fig. 3.1). The mechanisms for these conflicting stresses to attain a similar phenotype are unknown.

We proposed three possible hypotheses: one, that starving and toxified cultures exhibit overall similarities in gene expression response profiles; two, that transcriptional responses to conflicting stresses are fundamentally different; three, that there is a core conserved response alongside distinct responses in each condition. To put these hypotheses to the test, we cultured *E. coli* in low, intermediate, and high concentrations of lactose minimal medium, grew them to mid-exponential phase, and harvested cells from each culture for RNA-Seq.

The resulting transcriptional profiles were subjected to differential expression analysis with the intermediate lactose concentration as the reference condition. Our results showed decisively different gene expression profiles between the starving and toxified cultures. The response regulons were distinct but of similar size in toxified and starving conditions, the latter of which also showed signs of the stringent response. While there was a core conserved set of genes with similar responses in both conditions, toxified cultures showed a distinct gene expression pattern suggestive of a unique stress response. Our results show compelling evidence that persister-prone conditions can arise from many distinct pathways. We suggest that the persister phenotype as we have defined it here is a phenomenological property arising from multiple distinct pathways, and that many phenotypes may similarly arise from a combination of overlapping regulons.

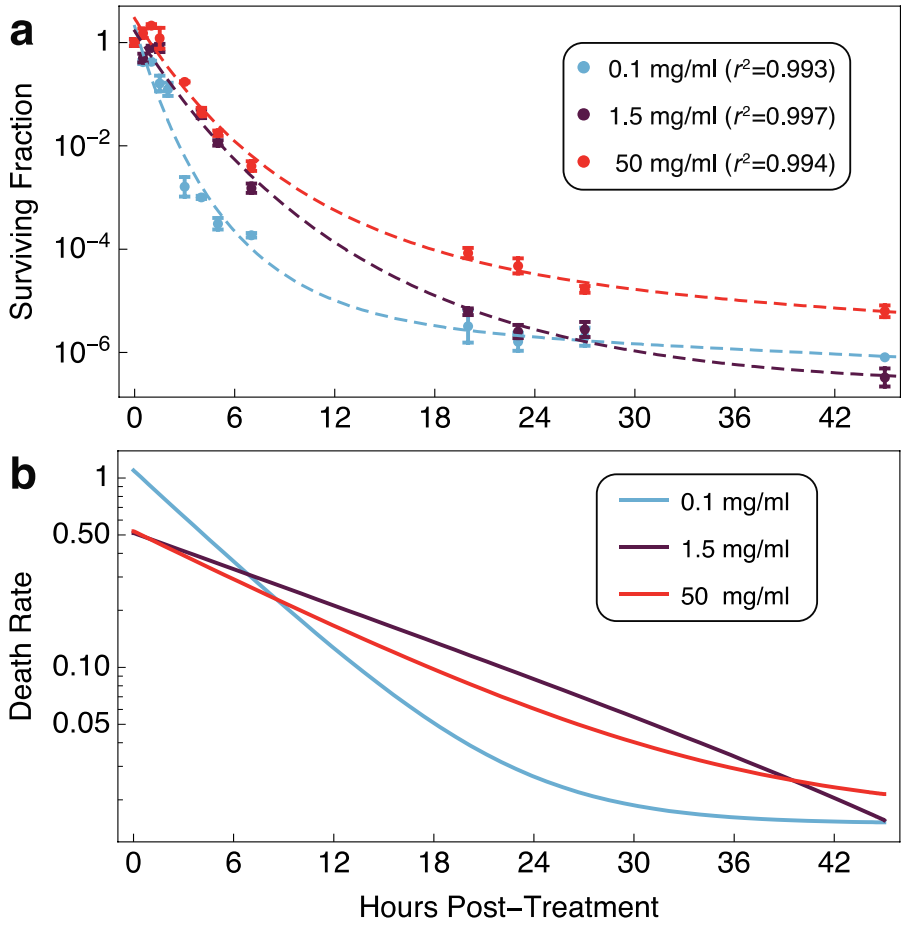

Figure 3.1: The killing rate of *E. coli* B REL606 in ampicillin is lowest in starving and toxified culture conditions. a. Surviving fractions in low, intermediate, and high lactose conditions after ampicillin treatment (100 mg/ml) during mid-exponential phase (mean ± standard deviation, N = 3; final point in 1.5 mg/ml, N = 1). Data from [13, 18] were fit to a mixed linear-exponential model $y = a\,t + b\,e^{-g\,t} + c$ with r^2 as reported in the figure. b. Time derivative $-\frac{dy}{dt}$ of the statistical model parameterized for each fit in panel a shows a lowered killing rate for both starving and toxified cultures between approximately 7 and 40 hours post-treatment.

3.2. Results

To analyze gene expression profiles, we purified total RNA from the cell culture from different lactose concentrations. We mapped sequencing reads to the reference genome *E. coli* B REL606 NC_012967.1 [17] with kallisto [19] and subsequently analyzed the count data in Python and DESeq2[20]. Setting the moderate lactose concentration (2.5 mg/ml) condition as the reference, we defined genes that were differentially expressed (DEGs) in starvation (lactose conc. 0.1 mg/ml) and toxified (lactose conc. 50 mg/ml) conditions as genes with an adjusted p-value, i.e. false detection rate (FDR), of below 0.05 and a log2 fold change (LFC) greater than 1 (Fig. 3.S1). Out of 4490 genes in the genome, 913 DEGs were upregulated, and 1668 DEGs were downregulated in the starvation condition, while 1638 DEGs were upregulated, and 853 DEGs were done regulated in toxified condition. For further comparing our RNA sequence analysis results with the classic stringent response, we refer to RNA sequence analysis results from induced ppGpp expression by Patricia et al. [21], where 638 DEGs were upregulated, and 682 DEGs were downregulated out of 4318 genes when cells were stressed. Gene differential expression clustering revealed that the phenotypes of starvation cells and toxified cells were quite different (Figs. 3.2 and 3.S2).

3.2.1. Slowed killing rate in starving and toxified *E. coli* cultures

Re-analysis of our previous time course of exponential-phase *E. coli* treated with ampicillin demonstrates a lowered rate of death in starving and toxified cultures (Fig. 3.1). To show this, we fit the data to a mixed linear-exponential statistical model in logarithmic coordinates on the *y* axis: $y = a\,t + b\,e^{-g\,t} + c$. All three were fit with low error ($r^2 > 0.99$ in all three cases). Taking the time derivative revealed the estimated rate of killing for each culture condition (Fig. 3.1b), which

was highest in non-stressful conditions. Therefore, both the starving and toxified cultures are prone to produce higher levels of antibiotic tolerance than less stressful intermediate conditions.

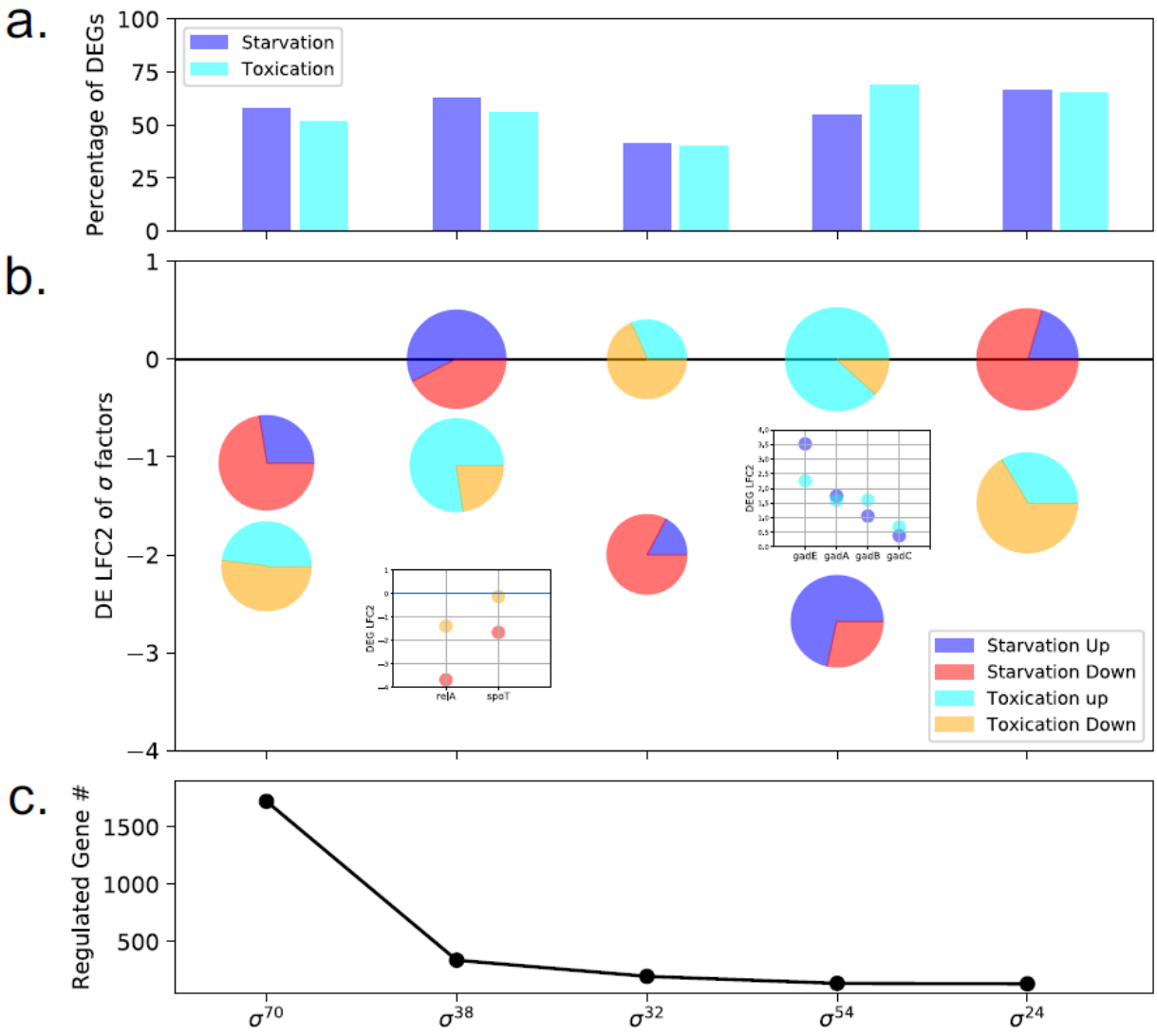

Figure 3.2: σ factor regulation differs responding to starvation and toxification. In both starvation and toxification, σ^{70}, which is the proliferation sigma factor, is downregulated. a. Percentage of differentially expressed genes among all genes regulated by that σ factor. b. The $\log_2$ fold change (LFC2) of the σ factor with corresponding percentage of upregulated genes and down regulated genes among all differentially expressed genes. c. The number of genes that are regulated by the σ factor.

3.2.2. Divergent expression profiles with a core conserved transcriptional response

The central metabolism of the high and low lactose cultures share a non-intuitive utilization of glucarate and galactarate (both oxidized derivatives of lactose [21]) as a carbon source and arginine degradation as a nitrogen source (Table S1). The utilization of these non-ideal metabolic staples could be the cause of the decreased growth rates of the high and low lactose cultures compared to the intermediate lactose condition [22]. Additionally, both cultures shift away from ammonia mediated nitrogen acquisition, tRNA charging, and general amino acid biosynthesis, potentially stemming from non-ideal carbon metabolism (Table 3.S2).

However, the starved and toxified cultures are not identical in their transcriptional profile. Unique to the low (0.1 mg/mL) lactose condition, the culture upregulates fatty acid beta oxidation and down regulates galactose degradation (Table 3.S4). This suggests that the lactose-starved cells are catabolizing previously made fatty acids and relying less on the free sugar in the medium. The lactose-toxified cells have a confounding transcriptional profile, as the galactose derivative metabolism and alternative sugar metabolic pathways are upregulated, yet many fatty acid and membrane-associated lipid pathways are downregulated. A potential explanation of this result is that the cells are undergoing a diauxic shift from using glucose as their primary sugar to previously exported galactose derivatives (Table 3.S5, 3.S6). The downregulation of the TCA cycle in lactose-toxified cultures could also be evidence of an overflow metabolism state [22], wherein sugar concentrations are in excess.

3.2.3. Similarities and differences in global transcriptional regulation responding to opposing stresses

Cells deploy fast regulatory responses to cope with varying environmental changes. Genetic regulation is upstream of the central dogma, thus having a global impact on subsequent regulatory responses on protein expression, metabolism, and cell phenotype. To understand the genetic regulation for cells confronting different stresses, we looked into regulation of DEGs. We classify DEGs into genetically regulated genes and constitutively expressing genes. Regulated genes are further divided into auto-regulated and non-auto-regulated gene categories. The classified DEGs are shown in Figure 3.3. In general, we observe similar distributions for regulated DEGs and non-regulated DEGs for cells in both starved and toxified conditions. To quantify this similarity, we first encoded the DEG LFC2 data with -1, 0 and 1 representing indices of negative DE, non-DE, and positive DE, then calculated the Pearson correlation for the encoded index between starvation and toxification condition. Pearson correlation for non-regulated DEGs is 0.66, for regulated but not auto-regulated DEGs is 0.67, and for auto-regulated DEGs is 0.56. This similarity may due to biased expression based on the gene's position on the genome, such as higher expression near the origin. Negative auto-regulation in gene circuits has been shown to decrease response times to stimuli and generally increase system stability [23, 24]. Auto-regulated genes showing lower correlations may be due to their functionality to maintain different steady states needed to respond to different stresses. The Hellinger distance between the regulated gene distribution and the non-regulated gene distribution for starvation is 0.22, and for toxicity is 0.21, meaning that there are still differences between the regulated gene and non-regulated gene distributions. Observing the distributions in Figure 3.3, we find that regulated genes have wider distributions in both conditions, confirming specified directional regulation as cells responding to stress. The LFC2 range for starvation is wider than that of toxification, implying more severe stress on starving cells compared to toxified cells. This result is consistent with our previous results [13]

and the model shown in Figure 3.1, where cells initially have a higher death rate under lactose starvation than when compared to cells grown in lactose toxicity.

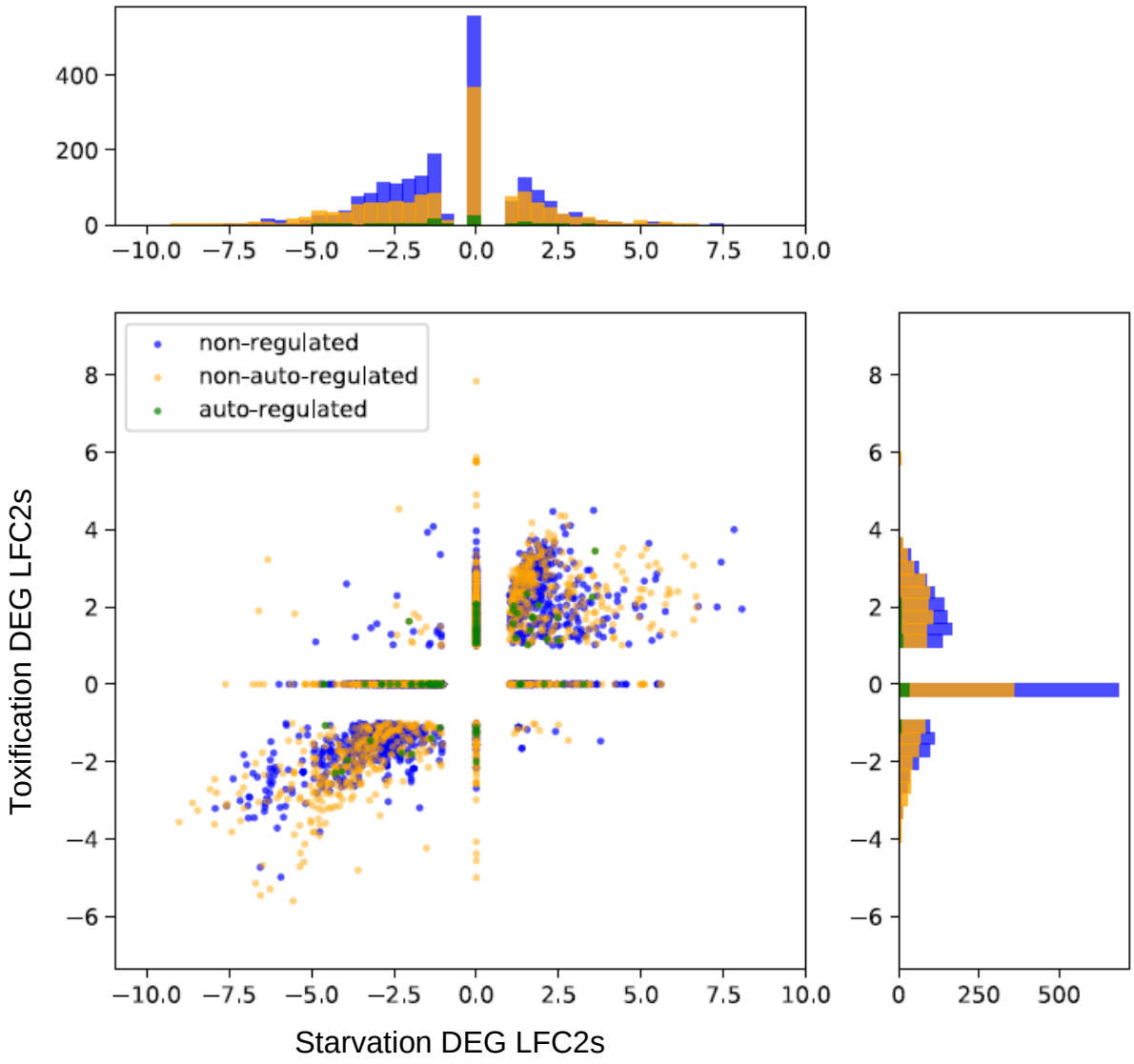

Figure 3.3: Differentially expressed gene LFC2 for starved cells and toxified cells. Histograms for non-regulated DEGs, regulated DEGs that are not autoregulated and autoregulated DEGs are show in blue, orange and green.

3.2.4. Stress responses regulated by sigma factors

Transcription initiation for *E. coli* promoters requires subunits of RNA polymerase (RNAP) called sigma (σ) factors [25]. Six σ factors exist in the *E. coli* B REL606 strain. Comparing to K-12 MG1655, σ^{28} (RpoF/FliA) is missing in REL606 strain, which is the flagellar synthesis sigma factor. σ^{19} (FecI), the ferric citrate sigma factor regulating iron transport and metabolism, is present in the REL606 strain, but does not interact with any genes according to RegulonDB [26]. To understand how a σ factor regulates cellular response to environmental stimuli, we looked at the differential expression for σ factors and their regulated genes (Fig. 3.2).

Distinct promoter classes depend on different σ factors to recruit RNAP for gene expression. Figure 3.2 shows that σ factor regulation corresponds with LFC2 above the differential expression cutoff. A hypothesis for this effect is that the competitive binding between different σ factors can switch RNAP binding regimes and subsequently result in growth phase changes, where ppGpp upregulates σ^{38} to activate stringent response gene expression [27]. In our data, compared to the stationary phase of the moderate lactose condition, there is a significant decrease of σ^{70} expression (rpoD) for both toxification and starvation conditions. This σ factor initiates by far the largest gene set of the sigma factors: 1723 genes, involving cell proliferation related behavior such as substrate update, DNA replication, membrane synthesis, and ribosome production. Consistent with the σ^{70} LFC2 values, most genes initiated by σ^{70} were downregulated in both toxified and starved conditions. Interestingly, σ^{38} (rpoS) is not differentially expressed in starving cells and is downregulated in toxified cells. This result is consistent with the relA and spoT expression level, where downregulated relA and spot expression can lead to lower ppGpp concentrations, and thus lower σ^{38} expression. The ppGpp stringent response may have been attenuated due to a strong σ^{70} presence in the moderate lactose condition. On the contrary, σ^{70} presence is downregulated in

starvation and toxification. Thus, even with relatively similar or lower σ^{38} expression, cells in starvation and toxification appear to undergo a stress response. This view is confirmed by the observation that the majority of DEGs initiated by σ^{38} are downregulated for starvation and toxification conditions.

σ^{54} (rpoN) is the only σ factor sharing little sequence homology to σ70 and it regulates 136 genes involving nutrient limitation such as nitrogen assimilation, substrate-specific transport systems, and utilization of alternative carbon and energy sources. σ^{54} has similar behavior to σ^{38}: the majority genes initiated by σ^{54} are downregulated, while σ^{54} itself is down- or un-regulated in starvation and toxification conditions compared to the reference. Toxified conditions have downregulated σ^{38} expression, but upregulated σ^{54} expression, which is opposite of the starvation condition.

3.3. Discussion

3.3.1. Excess lactose-induced persistence via diauxic shift toxicity due to a combination of overflow metabolism and Leloir pathway intermediates

One hypothesis for non-stringent persistence is metabolic toxicity induced by critical proteomic concentrations. In high glucose, *E. coli* undergoes a metabolic shift from primarily aerobic metabolism to incomplete oxidation of metabolites, including ATP synthesis [28]. The cause seems to be linked to a proteomic optimization, as anaerobic ATP synthesis requires a smaller fraction of the proteome to synthesize an equivalent amount of ATP at the cost of more sugar [22]. The smaller proteome allows for higher growth rates due to the reduced size of the necessary metabolome, allowing more transcriptional/translational machinery to be devoted to ribosome synthesis [22, 29].

We performed preliminary metabolic transcriptomic data analysis and found that the non-persistent bulk culture of cells utilize the Entner-Doudoroff shunt, which shifts pyruvate into phosphoenolpyruvate and downregulates the enzymes responsible for oxoalacetate and acetyl-CoA entering the citric acid cycle, which is consistent with cells undergoing overflow metabolism. The only sugar present in the media is lactose so proteomic stress could induce a state where GalE fluctuation leads to UDP-Galactose toxicity [20], and indeed, *galE* is downregulated in high lactose antibiotic tolerant cells, but not significantly differentially expressed in untreated cultures.

3.3.2. Different phenotypes could be due to sigma factor competition

Regulation of stress responses in bacteria is generally considered to be robust. Despite being exposed to opposite stress, we show that differential expression of genes in starvation and toxification have similar distributions. This may due to leaky expression of genes in the different locus on the genome. As expected, the regulated differentially expressed genes have wider $\log_2$ fold change distributions compared to constitutively expressed genes. With differential expression analysis, we find the global proliferation regulator σ^{70} is downregulated in both starvation and toxification conditions. As σ^{38} is not downregulated in starved cells, the sigma factor competition balance leans toward σ^{38}, and the starved cells are more stressed than that of toxified cells. The nutrient limitation-responsive sigma factor σ^{54} is downregulated, aiding glutamate-dependent acid resistance (GDAR). In toxified conditions, downregulation of σ^{70} is more drastic than that of σ^{38}, moving the sigma factor competition balance towards σ^{38}. Though σ^{54} is not downregulated, GDAR is again upregulated in toxified cells. Thus, nutrient-poor and nutrient-rich conditions both stress the cells with clear regulatory responses.

3.4. **Methods**

Our model system uses the *E. coli* strain B REL606, which has a unique enriched survival profile in toxic lactose concentrations compared to the K-12 strain [13]. This difference may arise from a more robust cell wall and has been shown to increase survival of this strain compared to others in ionically and osmotically suboptimal media [13, 30, 31].

3.4.1. Persister Enriched RNA-Seq

An E. coli REL606 *lacI*$^{-}$ strain transformed with Tn7::PlacO1GFP(KanR) used in previous experiments in the lab[32] [19] was inoculated in LB medium from a -80^{O}C bacterial stock and grown for 16 hours in a 37^{O}C shaking incubator. The LB culture was then resuspended (1:1000) into 5mL of Davis Minimal medium (DM; Difco) supplemented with thiamine and one of three lactose concentrations (0.1 mg/mL, 2.5 mg/mL, and 50 mg/mL). The cultures were allowed to acclimatize for 24 hours before being resuspended (1:1000) into 5mL of the same Davis Minimal medium and lactose concentration. Cultures were grown long enough (8 hours in 2.5 mg/mL lactose, 10 hours in 50 mg/mL lactose, and 12 hours in 0.1 mg/mL lactose) to provide enough biomass for RNA-Seq after antibiotic treatment. After the initial growth phase, 1.5 mL of untreated (no antibiotic) cell culture was RNA isolated according to the RNA isolation procedure below, dosed to 50 μg/mL concentration of ampicillin, and incubated for 24 hours in a 37 ^{O}C shaking incubator for persister cell enrichment. 3.0 mL of persister-enriched culture was then RNA isolated (procedure below) for lactose concentrations 0.1 mg/mL and 50 mg/mL. Lactose concentration 2.5 mg/mL required large experimental deviations to achieve enough post-ampicillin biomass requirements and was omitted.

3.4.2. RNA Sequencing

Cell cultures were pelleted in a microcentrifuge (10,000 G for 2 minutes), washed in PBS buffer twice, and resuspended in 500 μL of RNA-Later (ThermoFisher) and stored in at -20^{O}C for one week. The persister-enriched 50 mg/mL lactose culture was unable to be preserved by RNA-Later and proceeded to RNA isolation immediately. RNA isolation was performed using Direct-zol (Zymo) and TRIzol reagent (ThermoFisher) and stored in a -80^{O}C freezer overnight. Isolated RNA was ribo-depleted by RiboZero (Illumina) using ethanol washing to precipitate the RNA. Library preparation was completed using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) and sequenced using MiSeq v3 Paired End 150 bp (Illumina).

3.4.3. Genome re-annotation with Ecocyc and RegulonDB

The RNA transcript quantification was performed by kallisto [19] and genome NC_012967.1 [17] (Escherichia coli B str. REL606) was used as the reference genome. Kallisto was run using paired-end data and 10 bootstrap samples.As there is a recent update to REL606 sequence on NCBI, sequence offsets exist between Ecocyc annotations and NCBI sequence annotations for rel606 strain. Thus annotations from different databases were recognized by pairing annotation with gene locus. Furthermore, lacking extensive annotation of gene regulation in REL606, we used the K-12 MG1655 strain annotation based on gene and gene product similarities. Thus the annotation crossing strains are mapped by sequence alignment (Supplementary Methods), where databases Ecocyc [33] and RegulonDB [26] were combined for further analyses.

3.4.4. Differential Expression Analysis

We used R package DESeq2 for gene differential expression analysis. The RNA transcription quantification data were firstly clustered for isolating outliers using principal component analysis (PCA) (Figure S2). Two samples were taken out from subsequent analysis due to wrong clustering in the PCA plot with a high number of missing values. The rest samples were regarded as reliable and went into the DESeq2 pipeline.

DESeq2 pipeline includes size factor estimation, dispersion estimation, and DEG tests. Low-count RNA quantifications are noisy and may decrease the sensitivity of DEGs detection [34]. Size factors were calculated with a subset of control genes, which are non-regulated genes according to RegulonDB[26] and have expression higher than threshold 10 across all replicas. RNA sequencing profiles obtained from lactose starvation (0.1 mg/ml), moderate (2.5 mg/ml) and toxification (50 mg/ml) conditions were analyzed. Setting transcriptome quantification from the moderate lactose condition, we applied the adaptive-T prior shrinkage estimator "apeglm" and used Wald significance tests for detecting DEGs and obtaining the $\log_2$ fold changes (LFC).

3.5. Acknowledgments

This project was supported by Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant numbers P20GM103418 and P20GM103638. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

[1] P. Cui, H. Niu, W. Shi, S. Zhang, W. Zhang, and Y. Zhang, "Identification of Genes Involved in Bacteriostatic Antibiotic-Induced Persister Formation," (in eng), *Front Microbiol,* vol. 9, p. 413, 2018, doi: 10.3389/fmicb.2018.00413.

[2] S. Matsumoto *et al.*, "Unique transcriptional profile of native persisters in Escherichia coli," *J Biosci Bioeng,* vol. 125, no. 1, pp. 15-22, Jan 2018, doi: 10.1016/j.jbiosc.2017.07.015.

[3] A. Harms, C. Fino, M. A. Sorensen, S. Semsey, and K. Gerdes, "Prophages and Growth Dynamics Confound Experimental Results with Antibiotic-Tolerant Persister Cells," (in eng), *MBio,* vol. 8, no. 6, Dec 12 2017, doi: 10.1128/mBio.01964-17.

[4] S. M. Amato, M. A. Orman, and M. P. Brynildsen, "Metabolic control of persister formation in Escherichia coli," *Mol Cell,* vol. 50, no. 4, pp. 475-87, May 23 2013, doi: 10.1016/j.molcel.2013.04.002.

[5] B. W. Kwan, J. A. Valenta, M. J. Benedik, and T. K. Wood, "Arrested protein synthesis increases persister-like cell formation," (in eng), *Antimicrob Agents Chemother,* vol. 57, no. 3, pp. 1468-73, Mar 2013, doi: 10.1128/aac.02135-12.

[6] N. Hofsteenge, E. van Nimwegen, and O. K. Silander, "Quantitative analysis of persister fractions suggests different mechanisms of formation among environmental isolates of E. coli," (in eng), *BMC Microbiol,* vol. 13, p. 25, Feb 4 2013, doi: 10.1186/1471-2180-13-25.

[7] A. G. Tkachenko, N. M. Kashevarova, E. A. Karavaeva, and M. S. Shumkov, "Putrescine controls the formation of Escherichia coli persister cells tolerant to aminoglycoside netilmicin," (in eng), *FEMS Microbiol Lett,* vol. 361, no. 1, pp. 25-33, Dec 2014, doi: 10.1111/1574-6968.12613.

[8] Y. Pu *et al.*, "Enhanced efflux activity facilitates drug tolerance in dormant bacterial cells," *Mol Cell,* vol. 62, no. 2, pp. 284-294, 04/22/ 2016, doi: 10.1016/j.molcel.2016.03.035.

[9] E. Simsek and M. Kim, "Power-law tail in lag time distribution underlies bacterial persistence," (in eng), *Proc Natl Acad Sci U S A,* vol. 116, no. 36, pp. 17635-17640, Sep 3 2019, doi: 10.1073/pnas.1903836116.

[10] N. M. Vega, K. R. Allison, A. S. Khalil, and J. J. Collins, "Signaling-mediated bacterial persister formation," *Nat Chem Biol,* vol. 8, no. 5, pp. 431-3, Mar 18 2012, doi: 10.1038/nchembio.915.

[11] I. Levin-Reisman, I. Ronin, O. Gefen, I. Braniss, N. Shoresh, and N. Q. Balaban, "Antibiotic tolerance facilitates the evolution of resistance," (in eng), *Science,* vol. 355, no. 6327, pp. 826-830, Feb 24 2017, doi: 10.1126/science.aaj2191.

[12] I. El Meouche and M. J. Dunlop, "Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation," (in eng), *Science (New York, N.Y.),* vol. 362, no. 6415, pp. 686-690, 2018, doi: 10.1126/science.aar7981.

[13] J. C. Ray, M. L. Wickersheim, A. P. Jalihal, Y. O. Adeshina, T. F. Cooper, and G. Balazsi, "Cellular Growth Arrest and Persistence from Enzyme Saturation," *PLoS Comput Biol,* vol. 12, no. 3, p. e1004825, Mar 2016, doi: 10.1371/journal.pcbi.1004825.

[14] D. Dykhuizen and D. Hartl, "Transport by the lactose permease of *Escherichia coli* as the basis of lactose killing," *J Bacteriol,* vol. 135, no. 3, pp. 876-882, 1978. [Online]. Available: citeulike-article-id:12415087

http://jb.asm.org/content/135/3/876.abstract
[15] M. Eames and T. Kortemme, "Cost-benefit tradeoffs in engineered *lac* operons," *Science*, vol. 336, no. 6083, pp. 911-915, 2012. [Online]. Available: citeulike-article-id:10682501

http://dx.doi.org/10.1126/science.1219083
[16] J. Abramson, S. Iwata, and H. R. Kaback, "Lactose permease as a paradigm for membrane transport proteins," *Mol Membr Biol*, vol. 21, no. 4, pp. 227-236, 2004. [Online]. Available: citeulike-article-id:12575051

http://dx.doi.org/10.1080/09687680410001716862
[17] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res*, vol. 35, no. Database issue, pp. D61-5, Jan 2007, doi: 10.1093/nar/gkl842.
[18] P. Sanchez-Vazquez, C. N. Dewey, N. Kitten, W. Ross, and R. L. Gourse, "Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase," (in eng), *Proc Natl Acad Sci U S A*, vol. 116, no. 17, pp. 8310-8319, Apr 23 2019, doi: 10.1073/pnas.1819682116.
[19] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nat Biotechnol*, vol. 34, no. 5, pp. 525-7, May 2016, doi: 10.1038/nbt.3519.
[20] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, p. 550, 2014/12/05 2014, doi: 10.1186/s13059-014-0550-8.
[21] B. Y. Yang and R. Montgomery, "Oxidation of lactose with bromine," *Carbohydrate Research*, vol. 340, no. 17, pp. 2698-2705, 2005/12/12/ 2005, doi: https://doi.org/10.1016/j.carres.2005.05.025.
[22] M. Basan *et al.*, "Overflow metabolism in Escherichia coli results from efficient proteome allocation," *Nature*, vol. 528, no. 7580, pp. 99-104, Dec 3 2015, doi: 10.1038/nature15765.
[23] A. Becskei and L. Serrano, "Engineering stability in gene networks by autoregulation," *Nature*, vol. 405, no. 6786, pp. 590-593, 2000/06/01 2000, doi: 10.1038/35014651.
[24] N. Rosenfeld, M. B. Elowitz, and U. Alon, "Negative autoregulation speeds the response times of transcription networks," *J Mol Biol*, vol. 323, no. 5, pp. 785-93, Nov 8 2002, doi: 10.1016/s0022-2836(02)00994-4.
[25] M. S. Paget, "Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution," (in eng), *Biomolecules*, vol. 5, no. 3, pp. 1245-1265, 2015, doi: 10.3390/biom5031245.
[26] A. Santos-Zavaleta *et al.*, "RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12," (in eng), *Nucleic acids research*, vol. 47, no. D1, pp. D212-D220, 2019, doi: 10.1093/nar/gky1077.
[27] T. Nystrom, "Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition?," *Mol Microbiol*, vol. 54, no. 4, pp. 855-62, Nov 2004, doi: 10.1111/j.1365-2958.2004.04342.x.
[28] A. J. Wolfe, "The acetate switch," *Microbiol Mol Biol Rev*, vol. 69, no. 1, pp. 12-50, Mar 2005, doi: 10.1128/MMBR.69.1.12-50.2005.

[29] K. Peebo, K. Valgepea, A. Maser, R. Nahku, K. Adamberg, and R. Vilu, "Proteome reallocation in Escherichia coli with increasing specific growth rate," *Mol Biosyst,* vol. 11, no. 4, pp. 1184-93, Apr 2015, doi: 10.1039/c4mb00721b.
[30] K. Tomasek, T. Bergmiller, and C. C. Guet, "Lack of cations in flow cytometry buffers affect fluorescence signals by reducing membrane stability and viability of Escherichia coli strains," *J Biotechnol,* vol. 268, pp. 40-52, Jan 20 2018, doi: 10.1016/j.jbiotec.2018.01.008.
[31] F. W. Studier, P. Daegelen, R. E. Lenski, S. Maslov, and J. F. Kim, "Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes," (in eng), *J Mol Biol,* vol. 394, no. 4, pp. 653-80, Dec 11 2009, doi: 10.1016/j.jmb.2009.09.021.
[32] S. Quan *et al.*, "Adaptive evolution of the lactose utilization network in experimentally evolved populations of Escherichia coli," (in eng), *PLoS Genet,* vol. 8, no. 1, p. e1002444, Jan 2012, doi: 10.1371/journal.pgen.1002444.
[33] I. M. Keseler *et al.*, "The EcoCyc database: reflecting new knowledge about Escherichia coli K-12," *Nucleic Acids Research,* vol. 45, no. D1, pp. D543-D550, 2016, doi: 10.1093/nar/gkw1003.
[34] Y. Sha, J. H. Phan, and M. D. Wang, "Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2015, pp. 6461-4, 2015, doi: 10.1109/EMBC.2015.7319872.

	Pathway ID	NAME	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	PWYCQD-3	N-acetyl-galactosamine degradation	9	-0.25	1.877689053	1.786498
1	GLUCARDEG-PWY	D-glucarate degradation I	4	0	1.760165246	1.987105
2	GALACTARDEG-PWY	D-galactarate degradation I	4	0	1.760165246	2.03351
3	AST-PWY	arginine degradation II (AST pathway)	5	0	1.723538403	2.39796
4	GLUCARGALACTSUPER-PWY	superpathway of D-glucarate and D-galactarate degradation	5	0	1.898099336	2.199492
5	GLYCOLATEMET-PWY	glycolate and glyoxylate degradation I	7	0	1.962030992	2.178856
6	LYXMET-PWY	L-lyxose degradation	7	0	1.955055932	1.914805
7	PWY0-301	L-ascorbate degradation I (bacterial, anaerobic)	8	0	1.843739519	1.914805
8	PWY-6961	L-ascorbate degradation II (bacterial, aerobic)	9	0	2.125790811	2.004214
9	GLYCOL-GLYOXDEG-PWY	superpathway of glycol metabolism and degradation	11	0.240562612	1.940871818	2.539813
10	PWY0-1321	nitrate reduction III (dissimilatory)	12	0.483493778	1.937141283	1.903222
11	PWY0-42	2-methylcitrate cycle I	6	1	1.759956	2.19355

Table 3.S1. Pathways enriched in both starvation and toxification conditions measured by Normalized Enrichment Score (NES), normalized to the gene set size of the pathways. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

	Pathway ID	NAME	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	GLUTSYNIII-PWY	glutamate biosynthesis III	3	0	-1.9084	-1.78044
1	PWY0-1329	succinate to cytochrome b_o oxidase electron transfer	8	0	-1.9627	-1.69627
2	THRESYN-PWY	threonine biosynthesis	7	0	-2.2458	-1.97413
3	PYRIDOXSYN-PWY	pyridoxal 5'-phosphate biosynthesis I	7	0	-1.9821	-1.76787
4	LEUSYN-PWY	leucine biosynthesis	6	0	-2.1543	-1.72495
5	PWY-6277	superpathway of 5-aminoimidazole ribonucleotide biosynthesis	6	0	-2.5038	-2.06556
6	PWY-6123	inosine-5'-phosphate biosynthesis I	5	0	-2.2916	-1.8642
7	PWY-6121	5-aminoimidazole ribonucleotide biosynthesis I	5	0	-2.3507	-1.94256
8	PHESYN	phenylalanine biosynthesis I	5	0	-1.8685	-1.68616
9	PWY-7219	adenosine ribonucleotides de novo biosynthesis	3	0	-1.79118	-1.69118
10	AMMASSIM-PWY	ammonia assimilation cycle III	3	0	-1.8713	-1.78044
11	PWY-6122	5-aminoimidazole ribonucleotide biosynthesis II	5	0	-2.2843	-1.94256
12	ECASYN-PWY	enterobacterial common antigen biosynthesis	11	0	-1.8383	-1.68834
13	PWY0-781	aspartate superpathway	26	0.185164	-2.7615	-1.87522
14	P4-PWY	superpathway of lysine, threonine and methionine biosynthesis I	20	0.229416	-2.5674	-1.79183
15	1CMET2-PWY	formylTHF biosynthesis	13	0.232737	-2.4992	-1.95406
16	SULFATE-CYS-PWY	superpathway of sulfate assimilation and cysteine biosynthesis	14	0.258199	-1.8539	-1.96565
17	TRNA-CHARGING-PWY	tRNA charging	107	0.26773	-4.5769	-2.18056
18	MET-SAM-PWY	superpathway of *S*-adenosyl-L-methionine biosynthesis	11	0.288675	-2.0786	-1.64504
19	PWY0-1335	NADH to cytochrome b_o oxidase electron transfer	17	0.298807	-2.629	-2.04831
20	VALSYN-PWY	valine biosynthesis	9	0.395285	-1.9346	-1.81595
21	PWY0-162	superpathway of pyrimidine ribonucleotides *de novo* biosynthesis	11	0.430331	-2.2873	-1.66733
22	COMPLETE-ARO-PWY	superpathway of phenylalanine, tyrosine, and tryptophan biosynthesis	21	0.467707	-2.7401	-2.27617
23	ALL-CHORISMATE-PWY	superpathway of chorismate metabolism	52	0.473036	-2.8953	-1.97799
24	PWY-6629	superpathway of tryptophan biosynthesis	16	0.492366	-2.3182	-2.07576
25	PWY0-845	superpathway of pyridoxal 5'-phosphate biosynthesis and salvage	9	0.5	-2.0641	-1.7769
26	PRPP-PWY	superpathway of histidine, purine, and pyrimidine biosynthesis	43	0.551151	-3.8135	-2.78562

27	PWY-6628	superpathway of phenylalanine biosynthesis	16	0.5547	-2.5400	-2.1732
28	ARO-PWY	chorismate biosynthesis I	11	0.5863	-1.9821	-1.8532
29	GLUTAMINDEG-PWY	glutamine degradation I	9	0.6123	-2.1606	-1.7888
30	BRANCHED-CHAIN-AA-SYN-PWY	superpathway of leucine, valine, and isoleucine biosynthesis	16	0.6299	-2.5208	-2.0252
31	SER-GLYSYN-PWY	superpathway of serine and glycine biosynthesis I	6	0.7071	-1.8709	-1.6995
32	PWY-6126	superpathway of adenosine nucleotides *de novo* biosynthesis II	9	0.75	-2.4288	-1.8348
33	DENOVOPURINE2-PWY	superpathway of purine nucleotides *de novo* biosynthesis II	21	0.8167	-3.6885	-2.5339
34	PWY-6125	superpathway of guanosine nucleotides *de novo* biosynthesis II	9	0.8183	-2.3273	-1.7532
35	PWY0-1544	proline to cytochrome b_o oxidase electron transfer	5	1	-1.9282	-1.9021
36	PWY-7221	guanosine ribonucleotides *de novo* biosynthesis	4	1	-1.9537	-1.7658

Table 3.S2. Pathways that are down regulated in both starvation and toxification conditions. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

	Pathway ID	NAME	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	FAO-PWY	fatty acid β-oxidation I	9	-0.125	0	2.569134
1	PWY-4261	glycerol degradation I	5	0	0	2.21895
2	PWY0-381	glycerol and glycerophosphodiester degradation	7	0.645497	0	2.216135
3	PWY0-1182	trehalose degradation II (trehalase)	3	1	0	1.776402
4	GLYOXDEG-PWY	glycolate and glyoxylate degradation II	5	1	0	2.2208
5	PWY-6952	glycerophosphodiester degradation	6	1	0	2.090168

Table 3.S3. Pathways uniquely enriched in starvation condition. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

	Pathway ID	NAME	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	PYRUVDEHYD-PWY	pyruvate decarboxylation to acetyl CoA	3	0	0	-1.59135
1	PWY-5340	sulfate activation for sulfonation	3	0	0	-1.55612
2	PWY-5084	2-oxoglutarate decarboxylation to succinyl-CoA	3	0	0	-1.59135
3	GLYCLEAV-PWY	glycine cleavage	4	0	0	-1.645
4	DTDPRHAMSYN-PWY	dTDP-L-rhamnose biosynthesis I	6	0	0	-1.65113
5	PWYCQD-2	dTDP-*N*-acetylviosamine biosynthesis	6	0	0	-1.65113
6	SO4ASSIM-PWY	sulfate reduction I (assimilatory)	6	0	0	-1.93692
7	HISTSYN-PWY	histidine biosynthesis	9	0	0	-2.02013
8	OANTIGEN-PWY-1	*O*-antigen building blocks biosynthesis (E. coli)	12	0	0	-1.78812
9	GALACTMETAB-PWY	galactose degradation I (Leloir pathway)	5	0.408248	0	-1.78613
10	ILEUSYN-PWY	isoleucine biosynthesis I (from threonine)	10	0.625463	0	-1.81948
11	PWY0-1348	NADH to dimethyl sulfoxide electron transfer	16	0.722529	0	-1.6492
12	PWY0-1347	NADH to trimethylamine *N*-oxide electron transfer	17	0.767123	0	-1.65104
13	THREOCAT-PWY	superpathway of threonine metabolism	19	0.792948	0	-1.77919

Table 3.S4. Pathways that are uniquely down regulated in starvation condition. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

	Pathway ID	Name	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	PYRUVDEHYD-PWY	pyruvate decarboxylation to acetyl CoA	3	0	0	-1.59135
1	PWY-5340	sulfate activation for sulfonation	3	0	0	-1.55612
2	PWY-5084	2-oxoglutarate decarboxylation to succinyl-CoA	3	0	0	-1.59135
3	GLYCLEAV-PWY	glycine cleavage	4	0	0	-1.645
4	DTDPRHAMSYN-PWY	dTDP-L-rhamnose biosynthesis I	6	0	0	-1.65113
5	PWYCQD-2	dTDP-*N*-acetylviosamine biosynthesis	6	0	0	-1.65113
6	SO4ASSIM-PWY	sulfate reduction I (assimilatory)	6	0	0	-1.93692
7	HISTSYN-PWY	histidine biosynthesis	9	0	0	-2.02013

Table 3.S5. Pathways that are uniquely enriched in toxification condition. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

	Pathway ID	NAME	Pathway Size	Similarity	lac50 NES	lac0.1 NES
0	PWY-5188	tetrapyrrole biosynthesis I (from glutamate)	10	-0.666	-2.1521	0
1	FASYN-INITIAL-PWY	superpathway of fatty acid biosynthesis initiation (E. coli)	4	0	-1.8267	0
2	PWY0-881	superpathway of fatty acid biosynthesis I (E. coli)	9	0	-2.4322	0
3	PWY-6387	UDP-*N*-acetylmuramoyl-pentapeptide biosynthesis I (*meso*-DAP-containing)	9	0	-1.9031	0
4	PWY-6284	superpathway of unsaturated fatty acids biosynthesis (*E. coli*)	8	0	-2.1111	0
5	PWY0-163	salvage pathways of pyrimidine ribonucleotides	7	0	-1.9229	0
6	FASYN-ELONG-PWY	fatty acid elongation -- saturated	7	0	-2.1727	0
7	PWY-5971	palmitate biosynthesis II (bacteria and plants)	6	0	-1.9823	0
8	PWY-5973	*cis*-vaccenate biosynthesis	7	0	-2.1518	0
9	NAGLIPASYN-PWY	lipid IV_A biosynthesis	6	0	-1.8287	0
10	PWY-5989	stearate biosynthesis II (bacteria and plants)	5	0	-1.9988	0
11	PWY0-862	*cis*-dodecenoyl biosynthesis	5	0	-1.9823	0
12	PWY-6282	palmitoleate biosynthesis I	6	0	-1.7539	0
13	DAPLYSINESYN-PWY	lysine biosynthesis I	11	0	-2.0096	0
14	METSYN-PWY	homoserine and methionine biosynthesis	10	0.2721	-1.9882	0
15	LPSSYN-PWY	superpathway of lipopolysaccharide biosynthesis	20	0.3429	-2.2445	0
16	KDO-NAGLIPASYN-PWY	superpathway of $(KDO)_2$-lipid A biosynthesis	16	0.4285	-2.4428	0
17	PWY0-1479	tRNA processing	8	0.4879	-2.2374	0
18	PWY0-1061	superpathway of alanine biosynthesis	8	0.5204	-1.8253	0
19	TCA	TCA cycle I (prokaryotic)	19	0.5244	-1.8710	0
20	PWY0-166	superpathway of pyrimidine deoxyribonucleotides *de novo* biosynthesis (E. coli)	12	0.5345	-1.8572	0
21	PWY-5686	UMP biosynthesis	8	0.5773	-2.1137	0
22	TCA-GLYOX-BYPASS	superpathway of glyoxylate bypass and TCA	22	0.5980	-1.7823	0
23	PWY-7220	adenosine deoxyribonucleotides *de novo* biosynthesis II	6	0.7745	-1.9416	0
24	PWY-7222	guanosine deoxyribonucleotides *de novo* biosynthesis II	6	0.7745	-1.9416	0
25	GLYCOCAT-PWY	glycogen degradation I	6	1	-1.7583	0
26	PWY-5188	tetrapyrrole biosynthesis I (from glutamate)	10	-0.666	-2.1521	0

Table 3.S6. Pathways that are uniquely downregulated in toxification condition. Pathway ID and name are consistent with EcoCyc database annotations, and normalized enrichment score is calculated with R package fgsea.

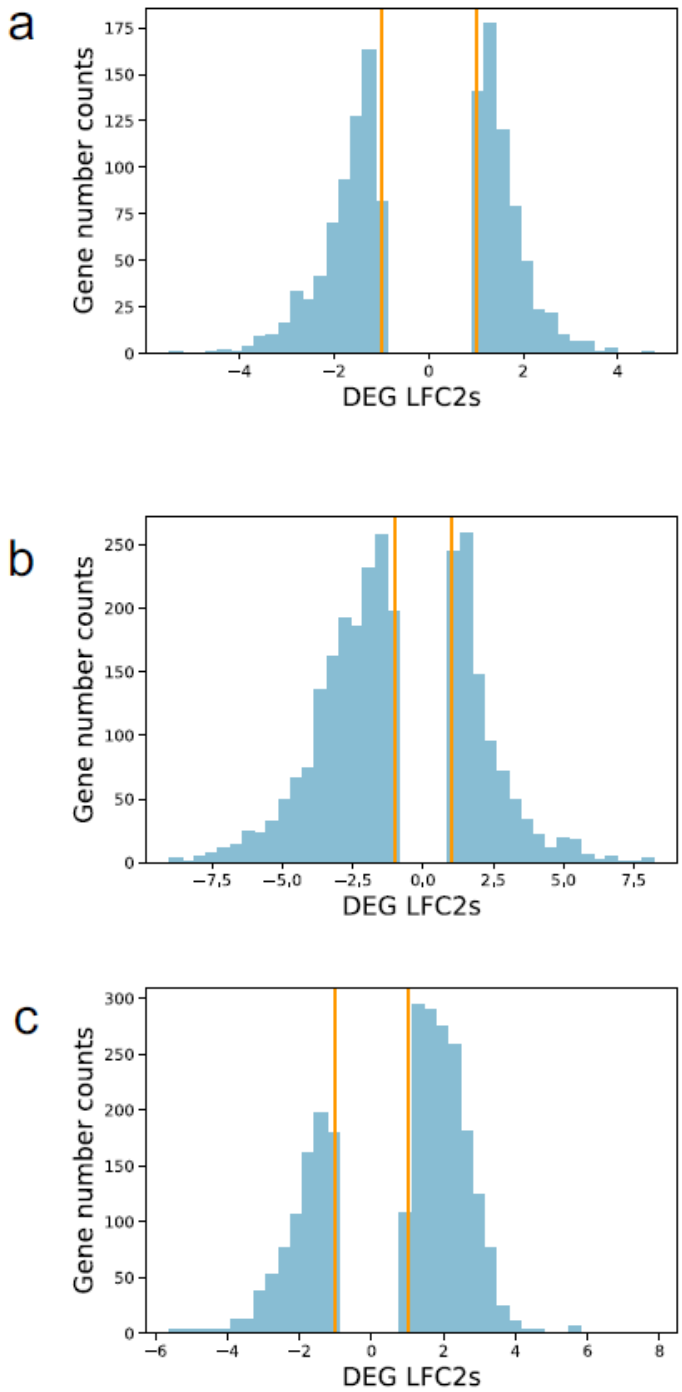

Figure 3.S1: Distribution of statistically significant differential expressed gene $\log_2$ fold change for different culture conditions, including **a.** induced ppGpp stress response, **b.** starvation condition (lactose conc. 0.1mg/ml), **c.** toxicity condition (lactose conc. 50 mg/ml). Vertical line in orange indicate expression level fold change of 2, corresponding to log2 foldchange of 1.

Chapter 4

Modification of the Bacterial Cell Cycle by Metabolic Signaling Costs

GW McElfresh[1]
Nishantha Wijesuriya[2]
Brian Drawert[3]
J. Christian J. Ray[*,1,2]

[*]Corresponding author.
Email: jjray@ku.edu
Tel: 785-864-1506

Address:
[1]Center for Computational Biology
University of Kansas – Lawrence
2030 Becker Dr.
Lawrence, KS 66047

[2]Department of Molecular Biosciences
University of Kansas – Lawrence
1200 Sunnyside Ave.
Lawrence, KS 66045

[3]Department of Computer Science
University of North Carolina – Asheville

Abstract

Cell physiological studies demonstrate systemic quantitative relationships between growth rate, metabolic state, and various structural parameters of bacterial cells. Regulatory networks are embedded in this context. Moreover, stochastic effects that arise at the scale of single bacterial cells result in uncertainty about how subnetworks interact with systemic physiological variables. Experimentally measuring all functionally relevant variables in the same cell is not technically feasible. However, building accurate computational or mathematical models capturing all of these interactions has also proven difficult. We use a hybrid statistical-stochastic modeling approach that simulates a well-characterized and detailed regulatory network with empirically precise parameter values that has key points of interaction with a knowledge-based statistical model of how cellular physiology responds to fluctuations in the network. This approach maximizes realism of the specific regulatory subnetwork within the context of a defined set of conditions captured by the statistical model. We apply the approach to the phosphate starvation response induced by the PhoBR two-component system in *E. coli*. The results reveal that fast molecular fluctuations are transmitted through an intermediate-timescale active transcriptional regulator to affect slower physiological processes that fluctuate with the growth rate. We thus reveal a principle for how PhoBR affects the cell cycle that could not be discovered with experiments alone.

4.1. **Introduction**

Cellular homeostasis is an energy-intensive process. The cost of gene expression is sufficiently costly [1] that some metabolic processes have evolved to run inefficiently in terms of metabolite usage to minimize protein production [2]. In bacteria, particularly the model organism *E. coli*, the cell cycle has been shown to be a coordinated process that depends on nutrient signals (energy) to determine its rate [3]. Chromosome replication and segregation is an energy-consuming, regulated process during the cell cycle [4-8] that continually alters gene dosage. The resulting coupled system has revealed well-known and newly emerging quantitative principles that couple the rate of growth to cell volume, chromosome segregation, rates of gene expression, and ribosome content. Sensory and signaling responses are inextricably embedded in this milieu. At the same time, gene expression and all other chemical reaction processes in single cells have a stochastic component that arises from the underlying nature of physical chemistry when the number of reacting components is limited. Stochasticity adds an element of uncertainty into how regulatory responses to environmental changes interacts with the systemic cell cycle.

The complexities of cellular physiology make it difficult to reconstruct and predict overall cellular responses to changing environments, despite the clear opportunities this would create for novel drug discovery and exploration of developmental processes. Current experimental methods restrict the number of measurable variables at single-cell resolution [9]. It is a remarkable fact that state-of-the-art understanding of the high-dimensional single cell has depended on measurements only a few at a time combined with extensive inference from large-population studies and models. We have developed a hybrid statistical-stochastic modeling approach to create robust, data-driven inferences of how regulatory subnetworks interact with systemic physiology and the cell cycle [10].

The first step is to identify a stimulon of interest, and then to determine the core subnetwork driving the response and any feedback resulting from changes in concentration of molecular species, and then determine points of contact with other physiological processes in the cell. The core network and any molecular feedback processes are encoded into a detailed stochastic simulation framework. Molecular fluctuations are coarse grained into discrete, functionally equivalent markovian variables. Within a set of conditions of interest (i.e. relevant to the subnetwork), experimental data are used to determine the quantitative effect of cellular responses to physiology. A specific example is detailed below, concerning the phosphate starvation response in *E. coli*. The advantage of this approach is that the high costs of whole-cell modeling are significantly reduced by data, in a similar manner to complexity-aware modeling [11, 12]. At the same time, it gives a detailed enough representation of a specific response that the dynamics are realistic, allowing quantitative simulation of the network of interest and inference of the relationship between global physiology and a given response.

We applied this approach to the phosphate starvation response in *E. coli*, regulated by the well-studied two-component system (TCS) PhoBR. PhoBR is a sensor histidine kinase – response regulator pair. The PhoR protein is membrane-bound and senses cytoplasmic phosphate starvation. On phosphate starvation, PhoR can autophosphorylate and transfer the phosphate group to PhoB, which stabilizes its dimerization and enables it to become an active transcriptional regulator of 32 known operons (one of which is cryptic in the popular *E. coli* strain K-12) containing 60 genes (46 in K-12). Among the members of the $(PhoB{\sim}P)_2$ regulon are the genes *phoU*, encoding a negative regulator of the PhoBR response, and *phoA*, a periplasmic alkaline phosphatase that recovers inorganic phosphate from multiple sources and accounts for a substantial fraction of the proteome during phosphate starvation.

The kinetics of PhoB-PhoR interactions have been quantitatively characterized in biochemical experiments. Furthermore, studies have measured global physiological responses to phosphate starvation in wild-type and strains lacking the PhoBR response, allowing robust estimation of the effects of the response on physiology, and in strains lacking such a response because of genetic inactivation of PhoBR responses.

With this model we analyzed the effects of stochastic fluctuations in PhoBR and variations in the *E. coli* cell cycle. Our results reveal that the key molecule in the system – the transcriptionally active PhoB~P dimer – functions as a translational layer between the fast molecular conformational fluctuations in the biochemical network and the slower fluctuations of large-scale physiological variables.

4.2. **Results**

4.2.1 Phosphate-Stressed *E. coli* Cells are Adders

The long-standing question of how the timing of cellular division is central to cellular physiology, and critical for an accurate model of the bacterial cell cycle. Such studies have previously been performed in many growth conditions. While *E. coli* growing normally at steady state are known to be "adders" (determining division based on the addition of a given quantity of biomass [13]), one study suggested that starving *E. coli* may be more like "timers", which divide after a certain amount of time regardless of the biomass gain [14]. However, whether this result is robust or general has come under question [15]|. Specifically, to our knowledge it is unknown whether the adder property of *E. coli* holds under phosphate starvation. Knowing this mechanism is critical to the accuracy of our cellular physiology model.

To address this question, we performed time-lapse microscopy experiments of *E. coli* K-12 MG1655 in a microfluidic flow device that provided MOPS medium with high (1mM) or low (50

μM) supplemented phosphate. We segmented the images and measured the relevant physiological variables: cellular division size, initial cell size, and the amount of time before division. The results clearly show that both phosphate-starved and phosphate-replete conditions are adders, with no significant difference between them (Fig. 4.1).

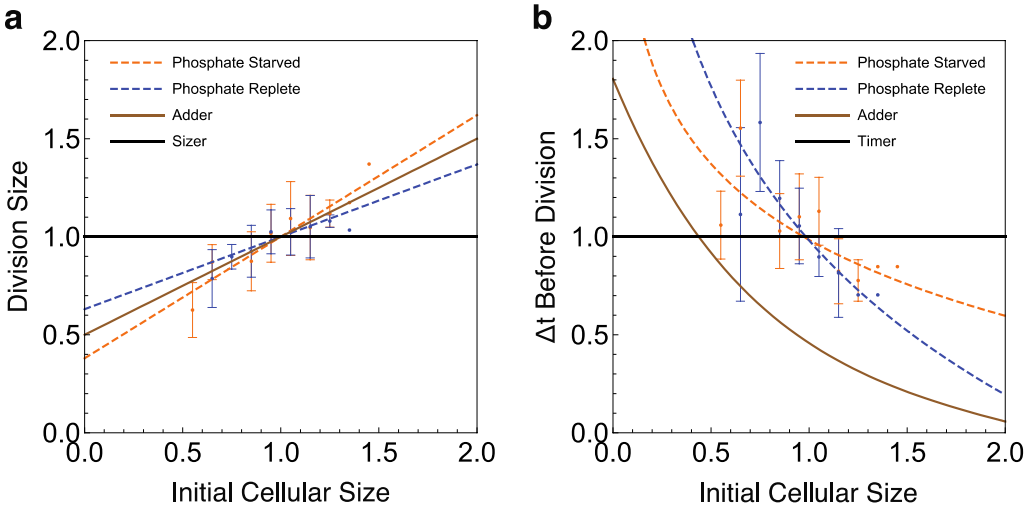

Figure 4.1. The effect of phosphate starvation on cell division size and time in *E. coli*. **a.** Relationship between initial cell size and division size shows that both phosphate-starved and phosphate-replete cultures are indistinguishable from the adder model. **b.** Relationship between initial cell size and time before division again supports the adder model. For phosphate-replete conditions, the number of cells analyzed was 50. For phosphate-starved conditions, the number of analyzed cells was 51.

4.2.2 Dynamics of the *E. coli* Cell Cycle During Phosphate Starvation

Simulations of the *E. coli* PhoBR stress model revealed a non-constant growth rate over the course of the cell cycle (Fig. 4.2a). The replicates exhibit the same non-monotonic pattern in their growth rate for cells grown under constant stress. Correlations on a principle component

analysis (PCA) of growth rate variation show that there are different drivers of the cell growth rate depending on the progress through the cell cycle (Fig. 4.2).

To understand the role of each molecular species in determining the growth rate, we measured the correlation between each molecular species and the first principal component of the cellular and molecular information (Figs. 4.3-4.9). The correlations reveal that the variability of transitioning through the cell cycle has several drivers at different rates of fluctuation. The primary cause of changes in the cell cycle is the proteins in the phosphate stress response regulon. This result makes intuitive sense, as the regulon is by far the most costly (aggregate) variable and also directly alleviates the penalties to growth rate introduced by the stress response. Other relevant, but less significant, species include the precursors to the active response regulator dimer (RR, RR~P) and the activated sensor histidine kinase (SHK_2P). Additionally, the sensitivity analysis revealed discrete timescales of growth rate effectors (Fig. 2b). The four classes of timescale and effect on the cell cycle can be broken down into physiological variables, precursors to response, inactive protein conformations and, in its own unique category, the transcriptionally active response regulator.

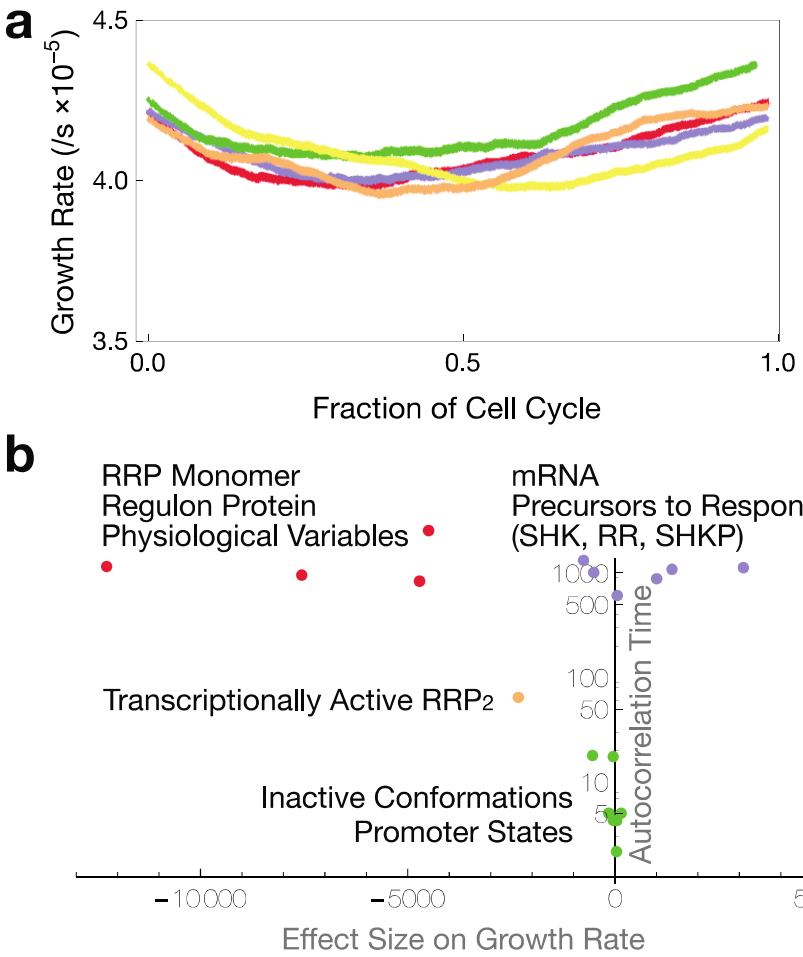

Figure 4.2. Cell cycle-synchronized simulations reveal new relationships between biochemical and physiological processes. **a.** Multiple cell cycle simulations under phosphate stress are synchronized, allowing stochastic sensitivity analysis of what processes contribute to a phenotype of interest (here, growth rate). **b.** Sensitivity analysis reveals discrete timescales of effects that different model variables have on growth rate. The timescales fall into discrete categories, reflecting the diverse ways that molecular fluctuations can emerge as effectors of overall cellular phenotype.

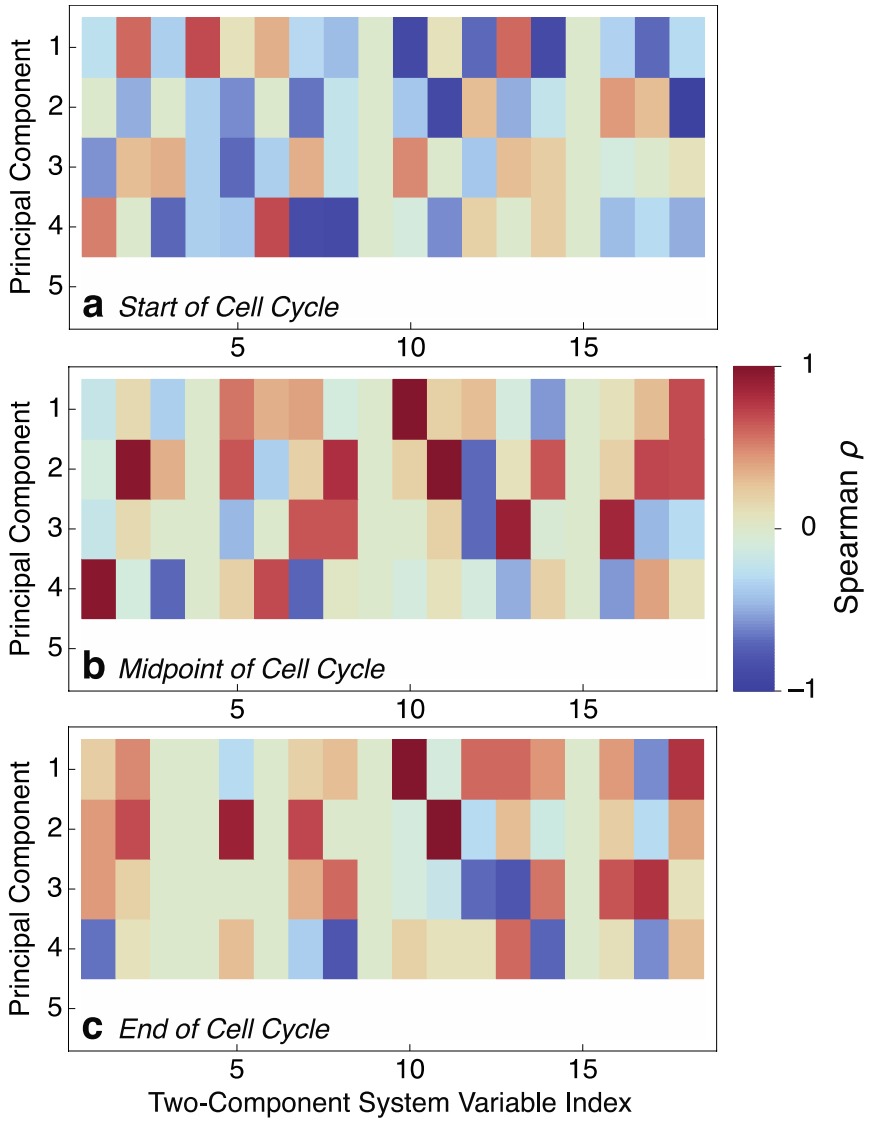

Figure 4.3. Principal component analysis to reveal the primary contributors of cell cycle variation during a phosphate stress response. The first principal component explains the most variance. Correlations (Spearman ρ) show the relationship between each variable and each principal component. Note that the 18 variables considered are consistently explained by four principal components throughout the cell cycle. **a.** Start of cell cycle immediately after division. **b.** Midpoint of cell cycle. **c.** End of cell cycle immediately before division.

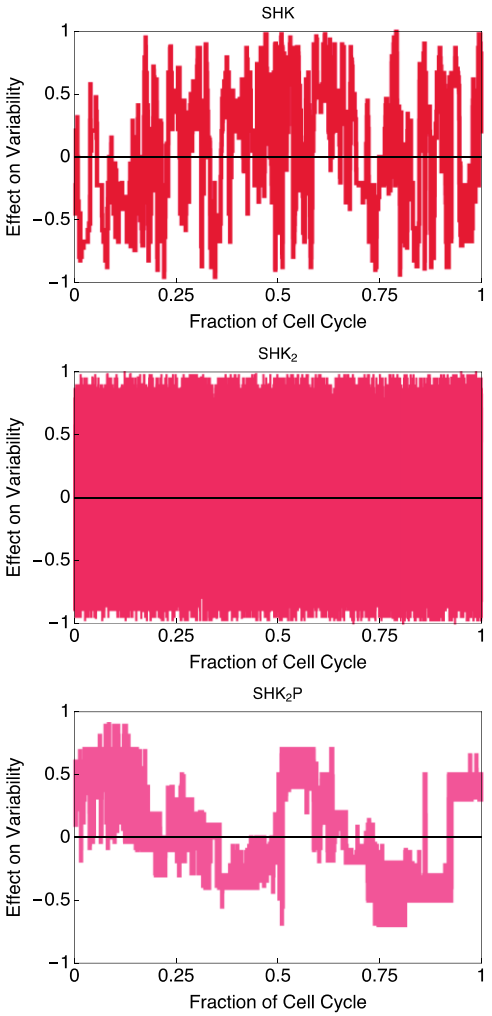

Figure 4.4. Internal variable correlation analysis of sensor kinase protein effects on cell cycle variability.

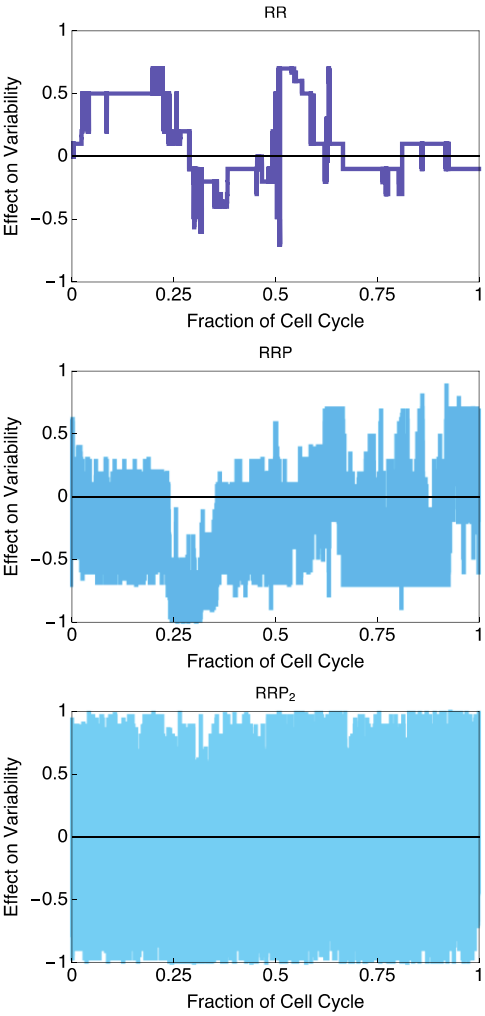

Figure 4.5. Internal variable correlation analysis of response regulator protein effects on cell cycle variability.

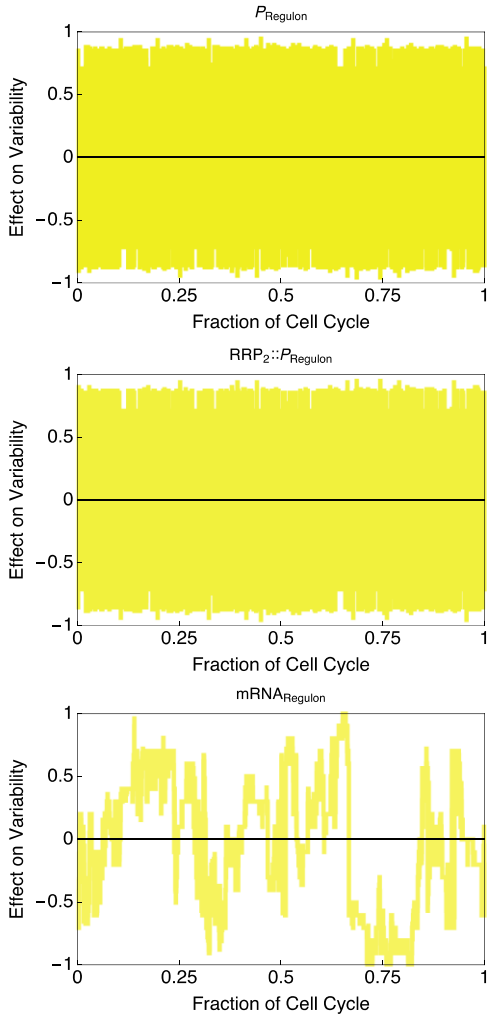

Figure 4.6. Internal variable correlation analysis of TCS regulon gene expression effects on cell cycle variability.

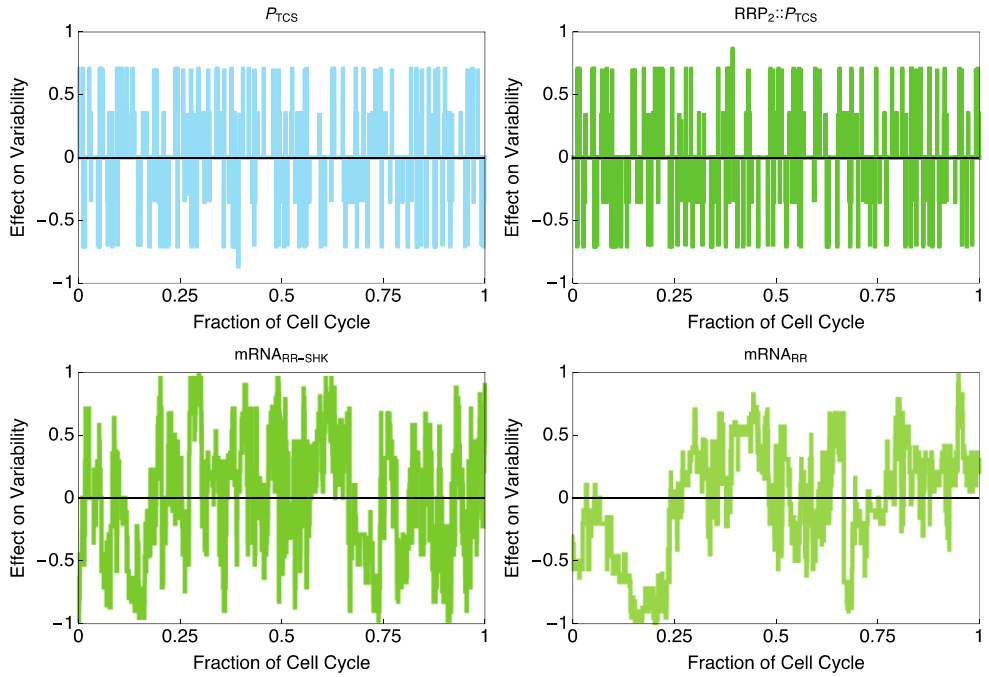

Figure 4.7. Internal variable correlation analysis of TCS gene expression effects on cell cycle variability.

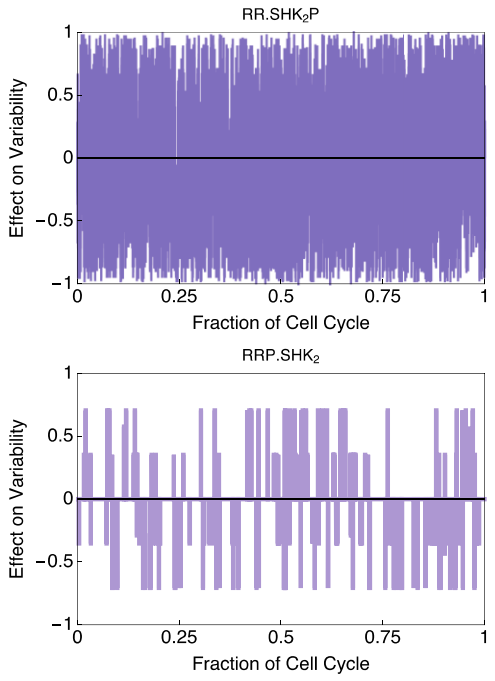

Figure 4.8. Internal variable correlation analysis of sensor-regulator complex protein effects on cell cycle variability.

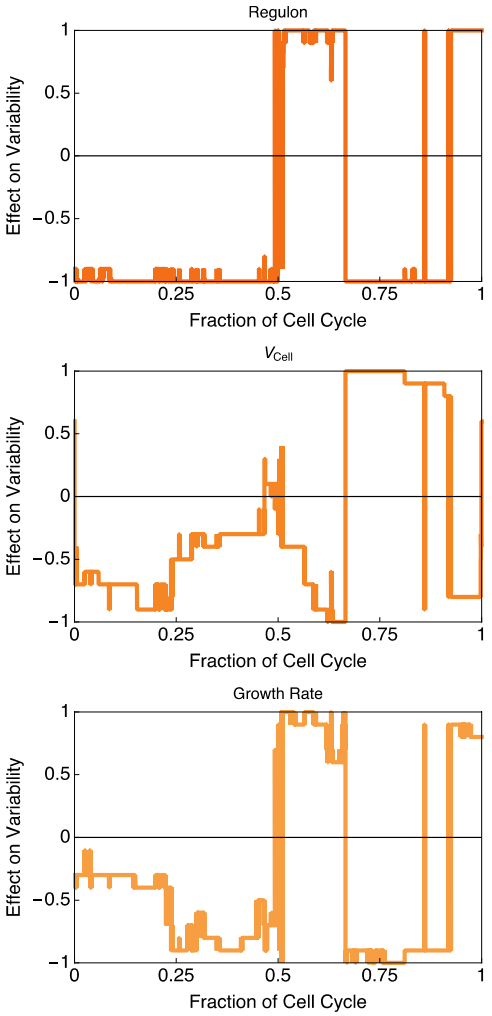

Figure 4.9. Internal variable correlation analysis of physiological variable effects on cell cycle

4.3 Discussion

4.3.1. Transcriptionally active response regulator as signal transducer in stress response

Wide distributions of response are characteristic of bacterial stress responses. The stochastic nature of bacterial stress responses could be due to the biological constraints of using a protein as an information channel. Our results show that the autocorrelation time for the transcriptionally active response regulator (RRP_2) is in an intermediary scale between the inactive protein conformation

states and the regulon proteins expressed. Naturally, the role of the response regulator is to transfer molecular scale fluctuations of proteins into the expression of genes, but the significance of this result is that the response regulator is the only signaling species that occupies an intermediate scale in autocorrelation time. This implies that the only dynamically relevant species for transferring the state of the short-scale molecular fluctuations to the regulon is the active transcriptional regulator RRP_2. Furthermore, since RRP_2 is effectively the channel by which information transfer occurs in the signaling network, the highly variable stress responses in bacteria may be explained by complex physical interactions that comprise response regulator-mediated expression. Specifically, there is a poorly understood relationship between the PhoB-regulated promoter and the ultimate expression level of the regulated genes [16]. The expression level also seems independent of the number of PhoB-regulated promoters directly before the gene, but instead the expression level depends on the function of the gene [16].

Previous studies of two-component systems revealed that some exhibit input-output robustness: the ability of the active transcriptional regulator to reflect signal levels outside the cell regardless of fluctuations in the number of TCS proteins present in the cell. Our study fill in a new aspect of that result, highlighting why input-output robustness may be an important trait selected for in evolution.

4.4 Methods

4.4.1 *E. coli* Growth Conditions

The *E. coli* strain K-12 MG1655 was inoculated into LB medium from a -80^{O} C bacterial stock and grown for 16 hours in a 37 ^{O}C shaking incubator (Corning LSE 71L). The LB culture was then resuspended (1:1000) into 1mL of MOPS EZ Rich Defined Medium (VWR) with 0.4% added glucose and high (1 mM) or low (50μM) phosphate. The cultures were allowed to

acclimatize for 24 hours before being resuspended (1:1000) into 1mL of the same MOPS medium and phosphate concentration. Cultures were grown for 4 hours before being transferred into the microfluidic device for imaging.

4.4.2 Integrated Computational Model

The basis of the integrated computational model has been described previously [10]. The concept is related to complexity-aware models that quantitatively account for physiological effects even if their precise mechanisms are not directly represented in the model [11, 12]. In this framework, we have statistical models representing the effects of phosphate starvation on the *E. coli* cell cycle and a detailed mechanistic stochastic simulation of the primary signaling system that responds to phosphate starvation – the PhoBR two-component system.

4.4.3 Analysis of Simulated Cell Cycle Trajectories

Simulated trajectories were subjected to two stages of analysis: (i) internal variable correlation analysis (IVCA) and (ii) autocorrelation analysis. The former allows us to quantify the relationships between any pair of variables in the model and the latter gives the timescales of those effects.

4.4.4 Internal Variable Correlation Analysis

IVCA is related to the popular approach of sensitivity analysis used in large-scale simulations in many fields [17, 18]. The difference is that while sensitivity analysis is concerned with how constant, but uncertain, parameters affect the prediction of a model (i.e. epistemic uncertainty), the analysis here is concerned with how stochastic fluctuations in different model predictions are related to each other (i.e. aleatory uncertainty).

The procedure is as follows: independent cell cycle trajectories were aligned at the same starting time, principal component analysis was performed at each timepoint, and the Spearman ρ

(non-parametric correlation) was computed between the first principal component of each pair of dependent variables of interest.

4.4.5 Autocorrelation Analysis

To determine the timescales at which different correlations fluctuate in the model, we performed autocorrelation analysis on the IVCA trajectories using standard methods [19]

4.5 Acknowledgments

This project was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM103418. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

[1] M. Lynch and G. K. Marinov, "The bioenergetic costs of a gene," (in eng), *Proc Natl Acad Sci U S A*, vol. 112, no. 51, pp. 15690-5, Dec 22 2015, doi: 10.1073/pnas.1514974112.

[2] M. Basan *et al.*, "Overflow metabolism in Escherichia coli results from efficient proteome allocation," *Nature*, vol. 528, no. 7580, pp. 99-104, Dec 3 2015, doi: 10.1038/nature15765.

[3] L. Dewachter, N. Verstraeten, M. Fauvart, and J. Michiels, "An integrative view of cell cycle control in Escherichia coli," (in eng), *FEMS Microbiol Rev*, vol. 42, no. 2, pp. 116-136, Mar 1 2018, doi: 10.1093/femsre/fuy005.

[4] Y. Sakiyama, K. Kasho, Y. Noguchi, H. Kawakami, and T. Katayama, "Regulatory dynamics in the ternary DnaA complex for initiation of chromosomal replication in Escherichia coli," (in eng), *Nucleic Acids Res*, vol. 45, no. 21, pp. 12354-12373, Dec 1 2017, doi: 10.1093/nar/gkx914.

[5] S. M. Mangiameli, B. T. Veit, H. Merrikh, and P. A. Wiggins, "The Replisomes Remain Spatially Proximal throughout the Cell Cycle in Bacteria," (in eng), *PLoS Genet*, vol. 13, no. 1, p. e1006582, Jan 2017, doi: 10.1371/journal.pgen.1006582.

[6] M. Stouf, J.-C. Meile, and F. Cornet, "FtsK actively segregates sister chromosomes in Escherichia coli," *Proceedings of the National Academy of Sciences*, 06/18/ 2013, doi: citeulike-article-id:12438716

doi: 10.1073/pnas.1304080110.

[7] N. J. Kuwada, K. C. Cheveralls, B. Traxler, and P. A. Wiggins, "Mapping the driving forces of chromosome structure and segregation in Escherichia coli," (in eng), *Nucleic Acids Res*, vol. 41, no. 15, pp. 7370-7, Aug 2013, doi: 10.1093/nar/gkt468.

[8] J. K. Fisher and N. Kleckner, "Magnetic force micropiston: an integrated force/microfluidic device for the application of compressive forces in a confined environment," *The Review of scientific instruments*, vol. 85, no. 2, 02// 2014, doi: citeulike-article-id:13934704.

[9] E. Levy and N. Slavov, "Single cell protein analysis for systems biology," *Essays in Biochemistry*, vol. 62, no. 4, pp. 595-605, 2018, doi: 10.1042/EBC20180014.

[10] C. R. GW McElfresh, "Intergenerational Cellular Signal Transfer and Erasure," in *The Interplay of Thermodynamics and Computation in Natural and Artificial Systems*, C. K. D Wolpert, J Grochow, and P and Stadler Eds. Santa Fe, NM: SFI Press, 2019.

[11] I. Lestas, G. Vinnicombe, and J. Paulsson, "Fundamental limits on the suppression of molecular fluctuations," *Nature*, vol. 467, no. 7312, pp. 174-178, 2010. [Online]. Available: citeulike-article-id:7801529

http://dx.doi.org/10.1038/nature09333

[12] M. Gomez-Schiavon and H. El-Samad, "Complexity-aware simple modeling," (in eng), *Curr Opin Microbiol*, vol. 45, pp. 47-52, Oct 2018, doi: 10.1016/j.mib.2018.01.004.

[13] S. Taheri-Araghi *et al.*, "Cell-Size Control and Homeostasis in Bacteria," *Current Biology*, vol. 25, no. 3, pp. 385-391, 2015, doi: 10.1016/j.cub.2014.12.009.

[14] M. Wallden, D. Fange, E. G. Lundius, O. Baltekin, and J. Elf, "The Synchronization of Replication and Division Cycles in Individual E. coli Cells," *Cell*, vol. 166, no. 3, pp. 729-739, Jul 28 2016, doi: 10.1016/j.cell.2016.06.052.

[15] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, "Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria," *Current Biology*, vol. 29, no. 11, pp. 1760-1770.e7, 2019/06/03/ 2019, doi: https://doi.org/10.1016/j.cub.2019.04.062.
[16] R. Gao and A. M. Stock, "Temporal hierarchy of gene expression mediated by transcription factor binding affinity and activation dynamics," (in eng), *mBio*, vol. 6, no. 3, pp. e00686-15, May 26 2015, doi: 10.1128/mBio.00686-15.
[17] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd, 2002.
[18] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," (in eng), *J Theor Biol*, vol. 254, no. 1, pp. 178-96, Sep 7 2008, doi: 10.1016/j.jtbi.2008.04.011.
[19] A. Sokal, "Monte Carlo methods in statistical mechanics: foundations and new algorithms," in *Functional integration*, C. DeWitt-Morette, P. Cartier, and A. Folacci Eds. Boston, MA: Springer, 1997, pp. 131-192.

Chapter 5

Conclusions and Future Directions

5.1. Summary of Results

The work presented here uses both experimental and computational and mathematical models to study the principal effectors in stress responses.

5.1.1. Responses in *E. coli* are noisy due to regulatory constraints

The work presented in Chapters 2 and 4 show that the transcriptionally active response regulator RRP_2 is principally responsible for short term cellular memory and acting as the information channel for expression of the regulon. The memory effect is due to a sequestration of activated response regulator, since the response regulator cannot be dephosphorylated while it is in its active, dimerized form, allowing the stress response to be fractionally "active" for 2-4 generations after the stress is no longer present. Sensitivity analysis performed in Chapter 4 on the same model under constant stress shows that the response regulator occupies an intermediary bridge state between the fast fluctuations of the inactive protein conformation states and slower scale gene expression products.

The work presented in Chapter 3 shows that the regulatory differences in persistent and non-persistent cells may be explained by the difference in σ factor expression. The competitive binding between the σ factors is a compelling hypothesis for phenotype determination due to the wide number of genes each σ factor regulates. Each condition showed a unique σ factor combination, but the persister enriched conditions showed a higher representation of nutrient starvation σ factors compared to intermediate lactose level, which had higher σ^{70} expression. The differential

combinations of these sigma factors may be the underlying mechanism of high-lactose mediated persistence.

5.1.2. Proteomic pressures in *E. coli* may cause bistability in bacterial populations

Bistable phenotypic responses are common in bacterial cultures, but the mechanisms are usually an internal system that causes growth arrest or a drastic change in growth rate. Here, we have presented evidence that *E. coli*, grown in high concentrations of lactose, undergoes population level heterogeneity through a combination of toxic intermediates of the Leloir pathway and the proteomic pressure provided by utilizing overflow-like metabolism. The diauxic shift from glucose to left over galactose via lactose degradation could be responsible for the essential condition for this type of bistability to occur.

5.2. Future Directions

5.2.1. Coarse graining of chromosomal modeling in *E. coli*

To extend our computational model presented in Chapter 2, we modeled the chromosomal replication cycle. The model uses the chromosomal locations of the genes in the PhoB operon and the speed of the replication fork to determine when the number of promoters for each gene should increment. Our model makes use of known trinormal replication fork speeds [1] to sample the speed at which DNA polymerase replicates the chromosome, and the model is capable of active replication forks to daughter cells in the event of multiple initiation events in fast growing cells.

To further improve the chromosome replication dynamics, the options are to implement previously developed models [2] that assume a time independent probability of replication or to add a more complex and rigorous DnaA dependence on initiation. Replication initiation occurs when DnaA reaches a critical concentration within the cell [3] so it would be an excellent extension

of the model to include. However, in high-nutrient simulations, the model's ability to support multiple origins of replication requires further experimental testing to consider multiple-origin DnaA criticality, as origins of replication typically initiate together [4].

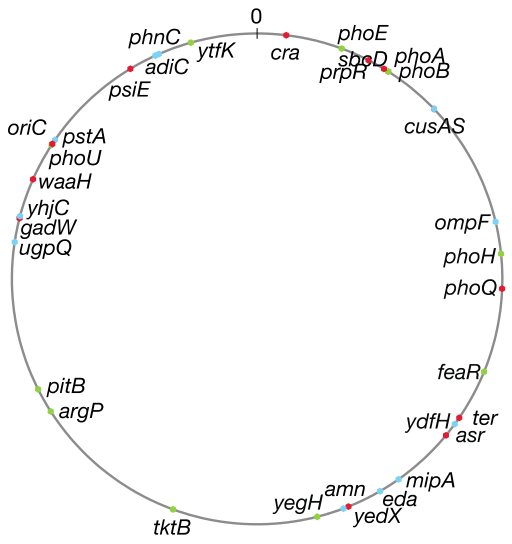

Figure 5.1: Representation of the locations of PhoB regulated genes in *E. coli.*

5.2.2. Spatial Stochastic Models

Moving forward, we seek to implement our whole-cell mesoscale *E. coli* stochastic model in a spatial context. Using nonspatial SSA methodology has specific limitations that we would like to explore. Principally, PhoR is an integral membrane protein and the kinetic effects of sequestering a protein in SSA-like simulations is currently unknown. Additionally, there is a gradual accumulation and localization of degraded protein in the old pole of *E. coli* [5] that is unable to be modeled in nonspatial simulations where there are no cell poles.

We have done preliminary work on the implementation of the mesoscale *E. coli* model as a spatial model. Using SpatialPy [6] we first create a mesh network of nodes (fig) in which each node is a nonspatial stochastic simulation. We impose spatiality by connecting the nodes and allowing proteins and protein complexes to diffuse between nodes. The nodes can be classified into regions subject to different parameter sets (*e.g.* cytoplasm, nucleoid, membrane, among others as needed) and diffusion rates can be altered to adhere to specific biological diffusions such as the asymmetric diffusion of a membrane bound protein moving to the membrane but then remaining confined.

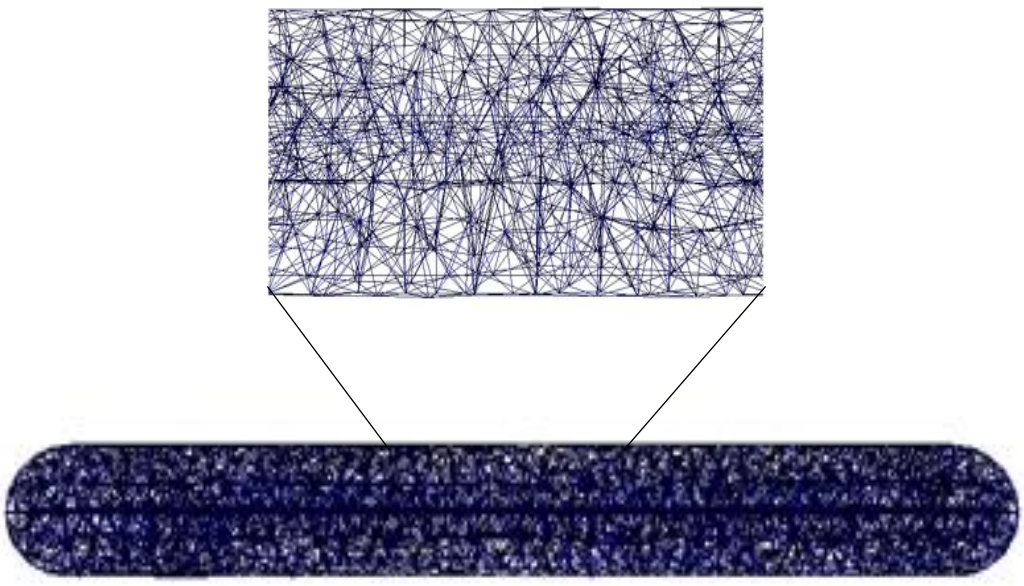

Figure 5.2: Spatial mesh of connected nodes in an *E. coli* mesh. Each node represents a voxel that performs an SSA simulation, but allows diffusion of molecules between locally connected nodes.

References

[1] T. M. Pham, K. W. Tan, Y. Sakumura, K. Okumura, H. Maki, and M. T. Akiyama, "A single-molecule approach to DNA replication in Escherichia coli cells demonstrated that DNA polymerase III is a major determinant of fork speed," (in eng), *Mol Microbiol,* vol. 90, no. 3, pp. 584-96, Nov 2013, doi: 10.1111/mmi.12386.

[2] M. Wallden, D. Fange, E. G. Lundius, O. Baltekin, and J. Elf, "The Synchronization of Replication and Division Cycles in Individual E. coli Cells," *Cell,* vol. 166, no. 3, pp. 729-739, Jul 28 2016, doi: 10.1016/j.cell.2016.06.052.

[3] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, "Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria," *Current Biology,* vol. 29, no. 11, pp. 1760-1770.e7, 2019/06/03/ 2019, doi: https://doi.org/10.1016/j.cub.2019.04.062.

[4] S. R. Khan, T. Mahaseth, E. A. Kouzminova, G. E. Cronan, and A. Kuzminov, "Static and Dynamic Factors Limit Chromosomal Replication Complexity in Escherichia coli, Avoiding Dangers of Runaway Overreplication," *Genetics,* vol. 202, no. 3, p. 945, 2016, doi: 10.1534/genetics.115.184697.

[5] E. J. Stewart, R. Madden, G. Paul, and F. Taddei, "Aging and death in an organism that reproduces by morphologically symmetric division," *PLoS Biol,* vol. 3, no. 2, p. e45, Feb 2005, doi: 10.1371/journal.pbio.0030045.

[6] B. Drawert, B. Jacob, Z. Li, T.-M. Yi, and L. Petzold, "A hybrid smoothed dissipative particle dynamics (SDPD) spatial stochastic simulation algorithm (sSSA) for advection–diffusion–reaction problems," *Journal of Computational Physics,* vol. 378, pp. 1-17, 2019/02/01/ 2019, doi: https://doi.org/10.1016/j.jcp.2018.10.043.