Genetic Regulatory Circuit Dynamics: Analysis and Synthesis

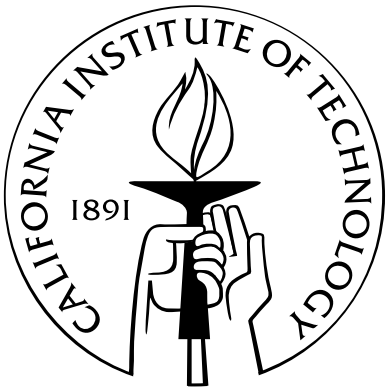
Thesis by

Joseph H Levine

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2012

(Submitted June 4, 2012)

For now we see through a glass, darkly.

—Paul, I Corinthians 13:12

Acknowledgements

My first and foremost thanks go to my advisor and mentor, Michael Elowitz. Throughout my time in his lab, he dispensed generous encouragement while providing a supportive intellectual environment to grow as a scientist. Stoically weathering the many technical and personal vicissitudes I rained down upon him, he managed to throw only an occasional professorial fit in my direction. Interspersed with professional advice of generally high value were cryptic aphorisms of generally questionable value, such as 'If you can't make it good, make it big. And if you can't make it big ... make it red.' On a more serious note, I must confess that Michael's most important lesson to me was unexpected. I certainly saw in him the canonical Caltech professor: intelligent, driven, and accomplished. Yet the main thing I learned from Michael was an eye for elegance, something which I find is rare even at schools like Caltech. Michael taught by demonstration that creative science, at its best, is like art, and for this I am extremely grateful.

I would like to thank my other thesis committee members: Erik Winfree, Rob Phillips, and Scott Fraser. Erik provided much needed counsel and support at critical times during my PhD. Rob's friendly greetings of "When are you getting out of here!?", liberally dispensed in the hallways of Broad and at the Caltech pool, were a constant reminder to keep pushing forward. Scott's willingness to provide critical feedback on my projects helped refine my perspective and approach.

Although a friend once said that most PhD students probably flirt with clinical depression during grad school, the inevitable lows were outweighed time and again by the many 'extracurricular' highs of life at Caltech. And for these highs I must thank my Caltech friends. Who else but a friend would drag me into a group to summit a snow covered Mt. Whitney in the winter? Who else but friends would shut themselves with me into a Caltech conference room to watch Star Trek the

Next Generation for almost 48 straight hours, as part of some twisted endurance test? Who else but a friend would work for days with me to discover the unexpected collaboratory web linking Watson, Einstein, and Gauss? Who else but a friend would bike with me all the way straight from San Francisco to Los Angeles, yet inform me in no uncertain terms and at no uncertain volume at the top of the Santa Ynez mountains that were I her husband, we would be experiencing 'severe marital discord'? Who else but a friend would join me in helping to judge the final rounds of a top international high school science competition, only to argue afterwards with me for an hour about why he lobbied to wrest first prize away from a blind girl?

For these friends, I am grateful. Signe Bray taught me that neuroeconomics is a real science, vegan is a real cuisine, and that a good roommate is worth their weight in gold. The 'country club,' Harry Choi, Suvir Venkataraman, and Fred Tan, provided nights filled with food and good memories, movies and bad memories, and alcohol and no memory. The 'Sons of S,' Fred Tan, Chiraj Dalal, Jon Young, and more recently Sandy Nandagopal and Amit Lakhanpal, kept daily lab life interesting, while keeping me CrossFit and linguistically versatile.

Finally, I would like to thank my parents, without whom all this would not have been possible, in more ways than one.

Abstract

How can cells shape and utilize dynamic gene regulation to enable complex cellular behaviors? I study this question in natural and a synthetic context.

The first project studies how a natural genetic network can imbue cells with a sense of 'time'. It has long been known that environmental signals induce diverse cellular differentiation programs. In certain systems, cells defer differentiation for extended time periods after the signal appears, proliferating through multiple rounds of cell division before committing to a new fate. How can cells set a deferral time much longer than the cell cycle? Here we study *Bacillus subtilis* cells that respond to sudden nutrient limitation with multiple rounds of growth and division before differentiating into spores. A well-characterized genetic circuit controls the concentration and phosphorylation of the master regulator Spo0A, which rises to a critical concentration to initiate sporulation. However, it remains unclear how this circuit enables cells to defer sporulation for multiple cell cycles. Using quantitative time-lapse fluorescence microscopy of Spo0A dynamics in individual cells, we observed pulses of Spo0A phosphorylation at a characteristic cell cycle phase. Pulse amplitudes grew systematically and cell-autonomously over multiple cell cycles leading up to sporulation. This pulse growth required a key positive feedback loop involving the sporulation kinases, without which the deferral of sporulation became ultrasensitive to kinase expression. Thus, deferral is controlled by a pulsed positive feedback loop in which kinase expression is activated by pulses of Spo0A phosphorylation. This pulsed positive feedback architecture provides a more robust mechanism for setting deferral times than constitutive kinase expression. Finally, using mathematical modeling, we show how pulsing and time delays together enable 'polyphasic' positive feedback, in which different parts of a feedback loop are active at different times. Polyphasic feedback can enable more accurate tuning

of long deferral times. Together, these results suggest that *Bacillus subtilis* uses a pulsed positive feedback loop to implement a timer that operates over time scales much longer than a cell cycle.

The second project proposes a method to rapidly generate and test complex genetic network dynamics in living cells. Existing microorganisms have evolved genetic circuitry to meet diverse challenges and maximize their survival and fitness. These challenges can be external pressures from the environment, or internal constraints imposed by an existing essential biophysical process. Furthermore, these challenges may be either static or dynamic in nature. While existing circuits have likely evolved to be 'good enough' to respond to historical challenges, it remains unclear if they can be improved upon, and whether they respond well to novel situations. Synthetic biology seeks to engineer organisms with complex novel phenotypes, both to harness these novel organisms for a function and to understand their underlying biology. Dynamic gene expression strategies may be necessary to meet dynamic internal and external challenges. Unfortunately, generating novel dynamic gene expression patterns with conventional genetic engineering remains a challenge. Here I propose and describe progress towards a computerized feedback control setup to enable the programming and rapid testing of dynamic gene regulatory patterns in living cells. Small sets of genes will be regulated optogenetically based on programmed control laws, and past and present cellular state. This setup will enable us to explore the functions and limits of engineered dynamic gene regulation, while hopefully, in the process, providing lessons about the underlying biology.

Contents

Chapter 1

Introduction

We are all taught in kindergarten that it is important to play well with others. Yet we are rarely taught how.

Likewise, we are all taught in biology class that for an organism to survive and thrive, the many molecules in its makeup must also play well together. Again, we are rarely taught how.

This thesis describes my efforts to understand how biological molecules work together to help organisms survive, and even thrive. I study gene regulatory networks, also known as genetic circuits; the collections of DNA, RNA, and proteins that govern the expression of genes and thus, directly or indirectly, all cellular behavior. More specifically, I ask how dynamic phenomena in these networks can underlie their function. I work in two different model systems, using both experimental and theoretical approaches. Analysis of the first system, the sporulation circuitry of the bacterium *Bacillus subtilis*, reveals that it can function as a timer. A novel type of pulsed regulation, 'polyphasic positive feedback', makes the timer more tunable. This work is described in Chapter 2. Synthesis of the second system, a 'cybernetic cell', is in progress. It will enable one to dynamically control a small set of genes in a living cell via computer. While this tool is extremely general, I will use it to explore how dynamic regulation of central metabolism can help cells cope with 'feast or famine' conditions. These efforts will shed light both on our ability to engineer gene expression, and on the workings of cellular metabolism. This work is described in Chapter 3.

1.1 Natural Gene Regulatory Dynamics

Much work needs to be done to understand the full range and function of dynamic phenomena in gene regulatory networks. It has long been appreciated that dynamic gene expression and signaling activities drive complex processes, such as metazoan development and microbial stress responses. Yet despite the richness and sophistication of these processes, one is often tempted to adopt a simplifying "binary" view of gene expression. One often sees developmental states classified in terms of which master regulators are 'on' or 'off'. Dynamic activity within these states, or during transitions between them, is often ignored. One wonders how this view, which essentially projects all of development onto some abstract N-dimensional binary hypercube, has survived for so long in the face of the obvious complexity and sophistication of living organisms.

One reason is that dynamic gene expression has traditionally been difficult to observe. The past 20 years have witnessed the emergence of fluorescent proteins as a powerful measurement tool in this area, when combined with microscopy and transgenic techniques (Megason and Fraser, 2007). Fluorescent proteins come in a wide palette of colors ranging from the blue to infrared, allowing multiple types to be used in the same cell. Transcription of an interesting gene can be visualized directly by fusing a fluorescent protein sequence to the gene's promoter. Fusing a fluorescent protein directly to another protein of interest allows it to be spatially tracked. Combined with microscopy, these techniques allow gene expression and protein localization to be tracked over time in single cells. Since many gene expression phenomena occur asynchronously across cell populations, these techniques resolved several dynamic phenomena which had previously been obscured by averaging. For example,**find some classic example about averaging from some review**. I will describe further examples later, in the context of stress responses in *Bacillus subtilis*.

1.1.1 Forms and Functions of Dynamic Gene Regulation

Gene regulation is inherently physicochemical (Phillips et al., 2009), and its dynamics should inherit the rich expressiveness of regular chemistry (Soloveichik et al., 2010). This rich expressiveness suggests that gene regulation and signaling may exhibit a wide range of forms. Indeed, recent work

offers tantalizing glimpses into the menagerie of possible gene expression dynamics. To see one striking example, one needs to look no further than the humble soil bacterium *Bacillus subtilis*, a popular model prokaryotic system. Life in the soil undoubtedly required this microorganism to develop a variety of talents to stand up to the variety of environmental challenges thrown its way. *Bacillus subtilis* exhibits three distinct, sophisticated stress responses, each controlled by separate dedicated circuitry. Remarkably, each of these stress responses can be regulated in a pulsatile fashion. Furthermore, these dynamics are all very different. In the competence response which primes cells to take up foreign DNA, a core genetic module can be triggered by noise to transiently turn on and then off in a pulse of activity (Suel et al., 2006). This module utilizes positive and negative feedback loops to generate the pulse, in the same way that neural ion channels interact to produce action potentials. In the general stress response system, repeated pulses of transcription from the alternative sigma factor σ_B regulate a variety of adaptive genes. These pulses originate from the repeated capture and release of σ_B by a feedback-regulated partner-switching mechanism. Finally, as will be shown in this thesis, in the sporulation response, the sporulation master regulator Spo0A activates in a pulsatile fashion and slowly ratchets up over multiple cell cycles.

1.1.2 The Ontology of Pulsed Regulation

The rich variety of pulsing dynamics displayed by *Bacillus subtilis* motivates us to ask what regulatory functions does pulsing have, generally? To do this, we first define the general characteristics of a pulse. Any individual pulse can be described by two characteristics: its amplitude and its timing (Figure 1.1). These characteristics can be combined to generate a variety of pulse regulatory dynamics, which I describe here.

1.1.2.1 Fixed Amplitude, Fixed Timing

If we assemble a series of pulses that have fixed amplitudes and fixed timings, we get a regular oscillator. Oscillating gene regulatory circuits have been well studied, both in naturally evolved circuits such as cell autonomous cell cycle and circadian oscillators, and in synthetic circuits **Elowitz,**

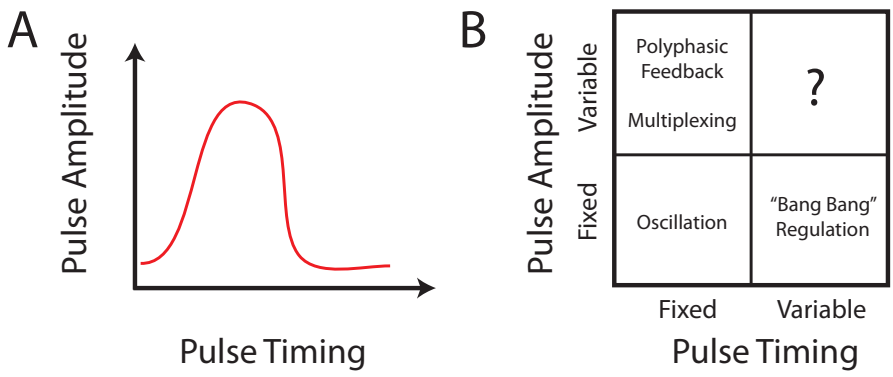


Figure 1.1: **The Ontology of Pulsed Regulation.** (A) A pulse can be described in terms of its timing and amplitude. (B) Different classes of pulsatile regulation, corresponding to fixed or variable amplitude and timing. Different combinations enable qualitatively distinct forms of regulation.

Hasty. The mechanisms behind and functions of oscillations generally have been the subject of extensive study (Winfree, 2001), and we will not elaborate excessively on them here.

1.1.2.2 Fixed Amplitude, Variable Timing

We now consider the case where pulse amplitudes are fixed, but pulse timings can be tuned or allowed to vary. Two examples of this phenomenon come to mind, exhibiting very similar functional principles.

The first example occurs in cell biology, specifically in the yeast calcium stress response. Here, the transcription factor Crz1 localizes to the nucleus in stochastic pulses[1], the frequency of which increase with extracellular calcium concentration (Cai et al., 2008). These pulses have a constant degree of localization (amplitude) on average, which is independent of calcium concentration. The variable timing, fixed amplitude pulsing was shown to enable 'coordinated' expression of a wide range of target genes with presumably very different promoter functions. Briefly, because each pulse has constant amplitude, each pulse thus produces a consistently sized burst of production from each target promoter. The exact amount depends on the promoter's specific response to Crz1. As the frequency increases by a certain factor, the amount of production from *all* promoters increases by that factor. Thus, production from different promoters increases proportionally with increasing

[1]The real system exhibits a strong transient localization response to a step increase in calcium concentration. We ignore this for the purposes of our discussion.

calcium —a response that is 'coordinated' despite the widely varying response functions of the different promoters.

The second example occurs in electrical engineering, specifically in amplifier circuit design. Class D amplifiers amplify an input signal by first converting it to a high-frequency pulse train, with the duty cycle of each pulse increasing with increasing signal amplitude (Self, 2006). This pulse train then switches an MOS output transistor between an 'off' state where it drives no current, and an 'on' state where it passes significant current from its drain to its source. The output is then low-pass filtered to remove the pulsing. What purpose does this complex regulation serve? It turns out that both the MOSFET off state and on state dissipate relatively little power. In the off state (technically known as cutoff), despite a significant potential drop between the source and drain, no current flows, so power dissipation is zero. In the on state (technically known as saturation), despite significant current flow from drain to source, there is little potential drop, so power dissipation is again minimal. Intermediate states (technically known as linear or triode regimes) have nonzero current and voltage and thus dissipate more power. Thus, the Class D amplifier is used in situations where power dissipation must be minimized, while some loss of fidelity is tolerable. Two illustrative examples lie at extreme ends of the audio amplifier spectrum. At a huge outdoor concert like Coachella, the amplifiers are run at such high power that overheating is a real danger, and a Class D topology avoids expensive heat sinking equipment and headaches. Conversely, in a cell phone audio speaker, power must be conserved to maintain battery life. Both cases have seen application of Class D amplifiers[2].

The fixed amplitude variable timing principle, sometimes known as 'bang-bang regulation', underlies both systems. In the Crz1 case, pulsing nuclear localization between two fixed levels[3] ignores the different shapes of the target promoters, collapsing them each into one number. In the Class D amplifier case, pulsed transistor regulation circumvents high power dissipation in the linear/triode regime. In both systems, pulsing the control signal from one extreme to the other avoids undesirable behavior at intermediate signal levels. Responsiveness is preserved by modulating the timing

[2]Class D amplifiers are also a typical 'macho' independent project attempted by the most skilled students in MIT's Analog Electronics lab course 6.101.

[3]Actually, the regulation works as long as the pulse shape is constant —the principle is the same.

(frequency or duty cycle) of the pulses.

1.1.2.3 Variable Amplitude, Fixed Timing

We now consider the case where pulse timings are fixed, but pulse amplitudes are allowed to vary. Again, two examples illustrate very similar functional principles for this class of regulation.

The first example occurs in communications systems. Time division multiplexing (TDM) allows multiple signals to use a single communications channel. The idea is to segregate the signals in time, so that each signal periodically gets to use the communication channel. This type of protocol has been applied, for example, in cellular communications (e.g., the 2G protocols).

The second example occurs in cell biology, specifically in *Bacillus subtilis* sporulation, and will be described in more detail in Chapter 2. I have found that certain feedback loops can be operated in a 'polyphasic' mode, where different portions of the feedback loop are active at distinct times. In the sporulation circuit, this form of regulation potentially breaks a positive feedback circuit into two halves which repeatedly activate in a nonoverlapping pattern to reinforce one another. In a standard positive feedback network, the network's activity can rapidly grow out of control if positive feedback strength is too strong. Polyphasic positive feedback tames some of this inherent instability by preventing the entire loop from activating at any one time. The result for the cell is a timer whose duration is easier to tune.

Both of these examples work because strictly timed pulsing temporally segregates multiple signals from each other. Each signal is allowed to evolve during its allotted time, but not to directly interact with other signals. In TDM, this prevents multiple signals from corrupting one another. In polyphasic positive feedback, this prevents unwarranted self amplification and controls the rate of signal growth.

1.1.2.4 Variable Amplitude, Variable Timing

This space is largely unexplored, and the mechanism and function of these pulsing dynamics remain to be discovered. One potential function here would be to relax the strict temporal segregation

of the 'variable amplitude, fixed timing' case. For example, in multiplexing, one could control the degree of temporal overlap between different pulse types.

1.2 Synthetic Gene Regulatory Dynamics

Naturally evolved genetics circuits clearly exhibit a variety of functional dynamic behavior, but can human engineers do 'better' than evolution? Can we engineer gene circuit dynamics which outperform their natural counterparts in a specific[4] task? What can such an engineering project teach us, both in how well we can expect to outperform biology, and also about the biological system itself?

Engineering synthetic gene regulatory dynamics is soul-crushingly slow and tedious, even by the most generous accounts[5]. The first pioneering examples, the Repressilator (Elowitz and Leibler, 2000) and the toggle switch (Gardner et al., 2000), were published in 2000 and demonstrated that it was possible to engineer complex temporal gene expression patterns. It is now 2012, and progress in the intervening decade has been modest. Some of the best circuit work has explored, for example, a variety of mechanisms to make more robust oscillators (Stricker et al., 2008). Yet, engineering more complex gene expression patterns currently seems out of reach. Anyone seeking to go down this path must navigate between the Scylla of poorly understood biological components, and the Charybdis of slow and tedious genetic engineering. The former makes it highly uncertain that any particular engineered construct will work, while the latter prevents rapid iteration and exploration of different solutions. To further complicate matters, we don't even know *why* we should engineer complex synthetic regulatory dynamics, because we have never been able to test their benefits *in vivo*. We're thus left with a chicken and egg problem where lack of method breeds lack of motivation, and lack of motivation breeds lack of method.

[4]Long term, we will certainly engineer synthetic organisms that broadly outperform natural organisms. Currently, however, our conceptual understanding of and technological control over biological systems is limited, and we will be content with this more limited goal.

[5]And this is coming from someone who has never worked with mice.

1.2.1 Cellular Cybernetics

How can we break out of this vicious cycle and make progress? The second half of my thesis describes my designs for a sort of 'cheat' to bootstrap ourselves out of this quandary. This cheat sets up an experimental system to control gene expression via computer. More specifically, a computer will regulate an optogenetic gene expression control system. Computerized control will let me quickly program and explore a wide range of gene expression dynamics. Specifically, I will challenge cells with an environment where nutrient type and availability fluctuate, and evaluate how various gene expression dynamics can facilitate growth therein. In principle[6], once ideal dynamics are identified, one can try and tackle the much more challenging problem of actually engineering these dynamics using autonomous genetic circuits. In a sense, this system will provide a sort of hybrid *in silico* - *in vivo* CAD technique to prototype different dynamic designs.

Computerized control provides several opportunities for exploration. First, as mentioned above, computers can be rapidly programmed to generate various dynamic gene expression patterns. Instead of slogging through weeks or months of genetic engineering to generate an appropriate circuit which may or may not work, we can set up a computer controlled experiment in about a day. The other opportunities come when one introduces *feedback* into this system. More specifically, the computer will regulate gene expression dynamics in a way based not only on our explicit commands, but also taking into account the cell's behavior. This ability is critical, because our understanding of cells is embryonic at best. We should *expect* to encounter unknown biological phenomena in the course of our experiments. And the computer should be prepared to adapt to them.

Why use the word 'cybernetics'? Norbert Wiener coined the term cybernetics in his book 'Cybernetics: Or the Control and Communication in the Animal and the Machine'. The word comes from the Greek term $\kappa\upsilon\beta\epsilon\rho\nu\eta\tau\eta\sigma$, for 'steersman'. Wiener believed generally that communication and control, at that time largely separate engineering disciplines, should be unified broadly and brought to bear on both technological and biological problems. The term is apt for this work, which augments a cell's decision making capabilities with a computer[7].

[6] though this would be reserved for future work

[7] In colloquial language, "cyborg" typically describes any living system which critically depends on major integrated

1.2.2 Automated Experimentation

I plan to use machine learning techniques to aid in my search for ideal dynamic strategies. Techniques such as reinforcement learning have long been used in this vein, for example in the planning of complex robotic tasks (Tedrake et al., 2005). Our task is not very different; instead of guiding a complex robot across challenging and possibly unpredictable terrain, we are guiding a cell's complex growth and metabolism through a complex series of feast or famine events. However, this project faces a task beyond that encountered in standard robotics. While a robot may be interacting with an uncertain environment, the robot itself is well understood and information about it can inform planning and learning tasks. In contrast, we really know very little about the cell we are trying to control. In fact, the hope is that a proper choice of learning methodology can not only teach us an optimal gene strategy per se, but also teach reveal something about the unknown underlying biology of the cell itself.

'Automated science' has been pursued in various scientific disciplines, at various times, with varying degrees of success. The general goal is to design a machine that collects knowledge both automatically and efficiently. Since the problem spaces being explored are large, these machines typically need to make effective decisions about how to best explore and 'study' their problem. Success stories exist in several fields. In mathematics, the automated theorem prover EQP was used in the mid 1990s to prove the Robbins Conjecture of boolean algebra (McCune, 1997). In physical chemistry and quantum control, automated techniques have been used to optimally shape laser pulses for the dynamic manipulation of quantum states (Rabitz et al., 2000). In robotics, a genetic-algorithm-like process has been used to design, simulate, select, and fabricate simple electromechanical control robots to perform simple locomotive tasks (Lipson and Pollack, 2000). In systems biology, robotic systems have been designed to iteratively and automatically propose and perform experiments to identify metabolic pathway components in bakers yeast (King et al., 2009).

and interacting biological and technological components. The definition is admittedly plastic and should really be used with a Stewart-like test: you should know it when you see it. Glasses would not really make me a cyborg. A pacemaker would be more of a start. Although my bacterial set up will never be elected governor of California, augmenting a cell's decision-making capabilities with a computer does constitute a qualitative step up in its abilities, and is good enough to satisfy the definition in my book.

I know of no study, however, that has explored how to best engineer an organism's behavior using these types of techniques. I am excited to see the outcome.

1.3 Contributions of This Thesis

In summary, this thesis makes several contributions to our understanding of gene regulatory circuit dynamics.

- *Biological Timers.* I discover that a canonical gene regulatory circuit, the sporulation initiation network of *Bacillus subtilis*, can function as a 'timer' to delay differentiation for many cell cycles after an abrupt environmental trigger. Generating this long delay is challenging, because cell division sets the time scale for protein turnover, and thus 'memory', in genetic networks. I demonstrate experimentally that this timer is cell autonomous and based on a core positive feedback loop within the circuit. Furthermore, I mathematically explore how a novel form of dynamic regulation, which I term 'polyphasic feedback', can make it significantly easier for the cell to reliably set the timer. This work has been recently published (Levine et al., 2012), and is described in Chapter 2.

- *Cybernetic Cellular Control.* I describe my progress towards engineering a 'cybernetic cell'. This system, in which expression of a few key genes is controlled by computer, will allow me to study what benefits different gene expression dynamics provide for a cell. Closed loop feedback control will enable me to more accurately 'program' various dynamics, and will also enable me to set the computer up to automatically learn optimal dynamic strategies. This work is currently in progress[8], and is described in Chapter 3.

[8]Most of this material, except for the microfluidic fabrication, is less than 3 months old

Chapter 2

A Genetic Timer: Pulsed Feedback Defers Cellular Differentiation

> Ingenium res adversae nudare solent, celare secundae. (As a rule, adversity reveals genius, while prosperity conceals it.) —Horace, Satires

2.1 Abstract

Environmental signals induce diverse cellular differentiation programs. In certain systems, cells defer differentiation for extended time periods after the signal appears, proliferating through multiple rounds of cell division before committing to a new fate. How can cells set a deferral time much longer than the cell cycle? Here we study *Bacillus subtilis* cells that respond to sudden nutrient limitation with multiple rounds of growth and division before differentiating into spores. A well-characterized genetic circuit controls the concentration and phosphorylation of the master regulator Spo0A, which rises to a critical concentration to initiate sporulation. However, it remains unclear how this circuit enables cells to defer sporulation for multiple cell cycles. Using quantitative time-lapse fluorescence microscopy of Spo0A dynamics in individual cells, we observed pulses of Spo0A phosphorylation at a characteristic cell cycle phase. Pulse amplitudes grew systematically and cell-autonomously over multiple cell cycles leading up to sporulation. This pulse growth required a key positive feedback loop involving the sporulation kinases, without which the deferral of sporulation became ultrasensitive to kinase expression. Thus, deferral is controlled by a pulsed positive feedback

loop in which kinase expression is activated by pulses of Spo0A phosphorylation. This pulsed positive feedback architecture provides a more robust mechanism for setting deferral times than constitutive kinase expression. Finally, using mathematical modeling, we show how pulsing and time delays together enable polyphasic positive feedback, in which different parts of a feedback loop are active at different times. Polyphasic feedback can enable more accurate tuning of long deferral times. Together, these results suggest that *Bacillus subtilis* uses a pulsed positive feedback loop to implement a timer that operates over timescales much longer than a cell cycle.

2.2 Introduction

Cells are capable of responding to stimuli extremely rapidly, on timescales of seconds or less. In some situations, however, cells respond to stimuli only after extended delays of multiple cell cycles. A classic example occurs in the developing mammalian nervous system, where, in the presence of appropriate signaling molecules, precursor cells will proliferate for up to eight cell generations before differentiating into oligodendrocytes. Although many aspects of the system remain unclear, oligodendrocyte differentiation is similarly delayed in vivo and in cell culture, suggesting a cell-autonomous timer mechanism (Raff, 2007). Another example is the mid-blastula transition in developing Xenopus embryos, which occurs after 12 cell cycles of proliferation (Newport and Kirschner, 1982a,a). In both cases, the deferral of differentiation enables a period of proliferation preceding commitment to new fates.

In bacteria, non-cell-autonomous strategies for deferring responses are well known. For example, in the marine bioluminescent bacterium *Vibrio fischeri*, cells use quorum sensing mechanisms to defer light production until the population reaches a critical density (Waters and Bassler, 2005). Similarly, *Bacillus subtilis* can defer sporulation through cannibalism, a response triggered by cell-cell signaling at high cell density, in which one subpopulation of cells lyses another, releasing nutrients that sustain growth (Gonzalez-Pastor et al., 2003).

Although there has been much work on circuit architectures that speed response times (Rosenfeld et al., 2002), fewer studies have addressed cell-autonomous deferral mechanisms. Cell autonomous

deferral requires the cell to keep track of the total time or number of division events since the appearance of the stimulus. It has remained unclear whether and how individual bacterial cells can achieve this functionality using genetic circuit components. The key problem is that as the cell grows and divides, its components dilute out. This dilution process sets an effective upper limit to the typical timescale over which the concentration of a protein responds to a step change in its production rate (Alon, 2007). For example, a step change in the rate of production of a stable protein causes the concentration of that protein to exponentially approach its new steady-state value with a timescale of one cell cycle. Thus, most gene circuits tend to relax to new steady states over timescales close to, or faster than, that of the cell cycle. Alternatively, genetic circuits can give rise to long deferral times in some cells through occasional stochastic switching between metastable states. While such systems can be tuned to generate long mean intervals between switching events, without cascades of multiple states, these mechanisms cannot generate well-defined, unimodal distributions of deferral times across a population (Acar et al., 2005).

B. subtilis sporulation provides an ideal model system to address this problem. Sporulation is a canonical microbial stress response behavior, during which cells respond to stress by differentiating into an environmentally resistant spore. Sporulation is a terminal differentiation decision, and its initiation is regulated by a well-characterized genetic circuit whose dynamics can be analyzed in individual cell lineages (Veening et al., 2008). This circuit, in response to diverse environmental and metabolic signals (Sonenshein, 2000), controls the activation of the master regulator Spo0A through transcriptional regulation and phosphorylation (Grossman, 1995). High levels of phosphorylated Spo0A ($Spo0A^P$) are sufficient to induce sporulation (Fujita and Losick, 2005). However, under some conditions, $Spo0A^P$ levels increase gradually over multiple cell cycles, allowing cells to proliferate prior to differentiation. The ability to defer sporulation while proliferating could provide a fitness advantage to cells by increasing their numbers relative to immediate sporulators (Figure 2.2A), although it could also impose a cost to cells that do not sporulate in time to survive extreme conditions. During the deferral period, cells may also explore other fates, such as biofilm formation, which are known to occur at intermediate levels of $Spo0A^P$ (Kearns et al., 2005).

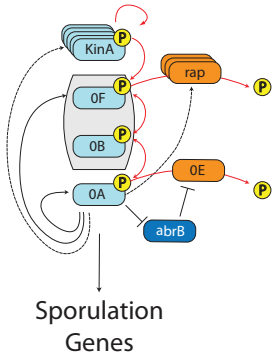


Figure 2.1: **The sporulation initiation network of *B. subtilis*.** Phosphorelay gene circuit controlling sporulation initiation (abridged). Kinases, including KinA, autophosphorylate and transfer phosphates to the master regulator Spo0A (denoted 0A) via Spo0F and Spo0B (red arrows). Phosphatases (orange) and transcriptional regulators (blue) also impact the system response. Dashed arrows indicate indirect transcriptional regulation by $Spo0A^P$.

The genetic circuitry controlling Spo0A activation includes multiple types of interactions (Figure 2.1). Histidine kinases such as KinA, KinB, and others autophosphorylate and transfer phosphates through a phosphorelay consisting of Spo0F and Spo0B to Spo0A (Burbulys et al., 1991). Phosphatases reduce the total level of $Spo0A^P$. For example, Spo0E directly dephosphorylates $Spo0A^P$ (Ohlsen et al., 1994), while rap phosphatases drain phosphates from the phosphorelay through Spo0F Perego et al. (1994). The system also includes extensive transcriptional regulation. $Spo0A^P$ regulates its own transcription as well as that of spo0F. It also regulates many other genes, including global regulators such as AbrB (Molle et al., 2003). Finally, $Spo0A^P$ also indirectly regulates its own activity by activating kinase expression (Fujita et al., 2005). These transcriptional interactions typically occur at much longer timescales than the fast phosphotransfer reactions of the phosphorelay. Nevertheless, it remains unclear whether and how this circuit facilitates deferred differentiation.

Here, using time-lapse fluorescence microscopy of individual cells, we show that under some conditions *B. subtilis* cells defer sporulation for multiple cell cycles through a predominantly cell-autonomous mechanism. We observed a progressively increasing series of pulses of Spo0A phosphorylation during deferral. Manipulation of circuit interactions revealed that pulse growth and regulated deferral both required positive feedback on kinase expression. These results suggest that *B. subtilis* uses a pulsed positive feedback loop to gradually 'ratchet up' $Spo0A^P$ activity pulses over multiple cell cycles in order to defer sporulation. Finally, mathematical modeling of this mechanism

further suggests that pulsing could enable a 'polyphasic' feedback mechanism, in which different parts of the overall positive feedback loop are active at different times, facilitating regulation of deferral. This may be a general strategy that cells can use to enable regulation of timescales much longer than the cell cycle.

2.3 Results

In order to analyze the sporulation initiation circuit in individual cells, we utilized a programmable time-lapse microscopy and quantitative image analysis system similar to those described previously (Suel et al., 2006; Eldar et al., 2009). We grew cells in Casein Hydrolysate (CH) growth media (Sterlini and Mandelstam, 1969) and then transferred them to agarose pads made with nutrient-limited media (RM) to induce sporulation during imaging (Donnellan et al., 1964). Under these conditions, individual cells exhibited a peaked distribution of 5.5 ±1.3 (mean ± SD) rounds of cell growth and division before initiating sporulation (Figure 2.2B, top). The probability of sporulating each round of growth, defined as the fraction of cells sporulating that round, increased monotonically (Figure 2.2B, bottom). This suggests that cells individually defer sporulation for multiple cell cycles, and rules out alternative Poisson-like models in which sporulation occurs with a fixed probability per cell cycle.

2.3.1 Pulsatile Activation of Spo0A

In order to understand how deferral is achieved, we set out to observe phosphorelay circuit dynamics during the deferral period. To read out $Spo0A^P$ activity we chromosomally integrated a P_{spo0F}-yfp reporter construct. The P_{spo0F} promoter exhibits a high affinity for $Spo0A^P$ and is therefore classified as a low-threshold activated gene (Fujita et al., 2005). To quantify P_{spo0F} activity over time, we computed its YFP production rate (promoter activity) in single cells. Promoter activity takes into account measurements of the change in total cellular fluorescence between time-points, the instantaneous cellular growth rate (which varies considerably, even within a single cell lineage, Figure 2.3B, bottom panel), and other cellular parameters (Materials and Methods). Compared

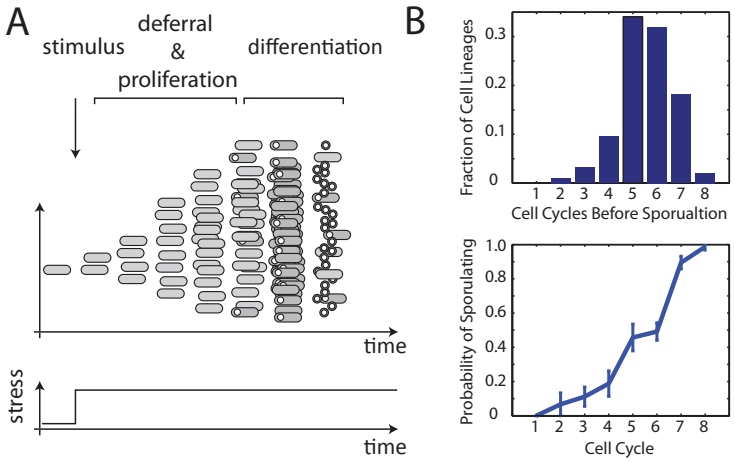


Figure 2.2: **Sporulation is a deferred fate decision for *B. subtilis*.** (A) In some conditions, in response to sudden nutrient limitation, cells first proliferate for multiple cell cycles and then initiate differentiation into spores. (B) Top: Distribution of number of cell cycles from nutrient limitation to sporulation, from five different movies taken on two different days. Bottom: Probability of sporulating for each cell cycle following nutrient limitation, defined as the fraction of cells at that round of division which sporulate instead of continuing growth. Error bars are standard error of measurement of the probability computed across the five movies.

to mean cellular fluorescence, whose interpretation is complicated by the stability of fluorescent proteins, promoter activity better reflects production from P_{spo0F} and thus $Spo0A^P$ dynamics. We also inserted a constitutively expressed red fluorescent expression construct, P_{trpE}-mCherry, which we used to aid in the automatic segmentation of cells in images.

P_{spo0F} promoter activity could be observed in discrete pulses in individual cells, similar to those reported previously (Figure 2.3B, middle panel) (Veening et al., 2009). These pulses began after transfer to nutrient-limited conditions and continued until sporulation. In contrast, cells in rich media exhibited no measurable production from the P_{spo0F} promoter, or sporulation associated genes generally. Pulses were not specific to the P_{spo0F} reporter, but were observed across a range of Spo0A target genes (Figure 2.12), affecting many processes in the cell, including the expression of the global regulator abrB (Strauch et al., 1989),(Strauch et al., 1990) and sdp, a component of the 'cannibalism' pathway (Gonzalez-Pastor et al., 2003),(Fujita et al., 2005). However, the phasing of pulses relative to the cell cycle differed between promoters, reflecting their different regulation modes (Figure 2.12). For example, $Spo0A^P$-activated and $Spo0A^P$-repressed promoters showed opposite phasing with respect to the cell cycle (Figure 2.12B). Each cell cycle typically contained

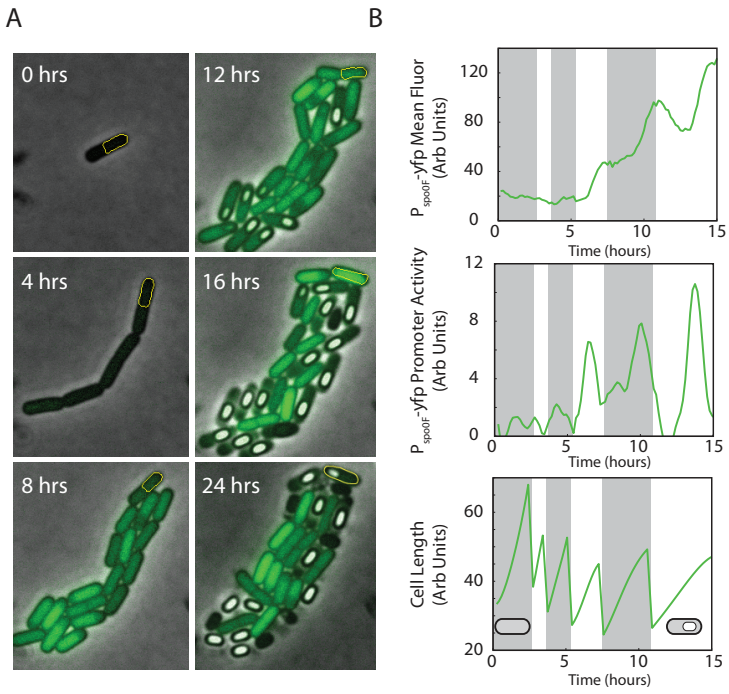


Figure 2.3: **Spo0A target genes pulse during sporulation initiation.** (A)Typical filmstrip showing growth and sporulation in a *B. subtilis* microcolony. Cells contain a P_{spo0F}-yfp reporter construct (green). Note the multi-cell-cycle increase in fluorescence intensity. (B) Quantitation of YFP expression from the cell lineage outlined in yellow in (A). Mean fluorescence intensity (top) shows abrupt changes in fluorescence accumulation. The promoter activity (middle, see Materials and Methods), which is inferred from rates of change of fluorescence and cell size (bottom), reveals pulses in the rate of YFP production from the P_{spo0F} promoter.

one pulse (Figure 2.12C). Promoters not regulated by Spo0A, such as the σ^A-dependent P_{trpE} promoter, sometimes fluctuated in expression but exhibited a much smaller dynamic range, and no characteristic cell cycle phasing, and were thus qualitatively different from Spo0A-dependent pulsing (Figure 2.14).

In principle, pulses could be caused by a change in either the abundance or the phosphorylation state of Spo0A. To eliminate both transcriptional and phosphorylation control of Spo0A activity, we replaced spo0A with the well-characterized, constitutively active variant $spo0A^{sad67}$ (Ireton et al., 1993), under the control of the IPTG-regulated hyperspank (P_{hyp}) promoter. Although P_{spo0F} was activated in response to $spo0A^{sad67}$ induction, this strain showed no pulsing (Figure 2.15), consistent with the fact that $spo0A^{sad67}$ does not require phosphorylation to be active. In addition, very few cells formed phase bright spores. The potential for pleiotropic effects of $spo0A^{sad67}$ expression

prevents us from concluding that successful sporulation requires pulsing. On the other hand, a strain in which spo0A was under the control of P_{hyp} retained similar pulse dynamics as wild-type (Figure 2.16) and consistently formed phase bright spores. These results strongly indicate that phosphorylation of wild type Spo0A is required for pulsation.

What molecular mechanism could be responsible for pulse generation? The sporulation kinase inhibitor Sda is regulated in a cell-cycle-dependent fashion, suggesting that it might be involved in pulse generation (Burkholder et al., 2001). A null sda mutant exhibited increased mean Spo0A activity, and therefore strongly reduced the dynamic promoter activity of the sensitive Spo0A-repressed P_{abrB} promoter (Fujita et al., 2005), as observed previously (Figure 2.17) (Veening et al., 2009). However, in the Δsda mutant, P_{spo0F} continued to pulse similarly to wild-type, showing that while Sda modulates the dynamic range of Spo0A activity, it is not required for pulsing. Intrinsic network dynamics involving negative feedback loops provide another possible pulse generation mechanism (Giudicelli et al., 2007). Together, spo0A, abrB, and spo0E form such a feedback loop (Figure 2.1). However, deletion of spo0E did not eliminate pulsing (Figure 2.17). Other potential negative feedback loops involve Spo0A-dependent up-regulation of rap phosphatase expression. But deleting rapA and rapB individually and in combination similarly failed to abolish pulsing (Figure 2.17). Finally, we asked whether pulsing might be driven specifically by one of the phosphorelay kinases. In nutrient-limited conditions, KinA and KinB are the dominant phosphodonors (LeDeaux et al., 1995; Jiang et al., 2000). Strains lacking either kinA or kinB exhibited pulsed dynamics (Figure 2.18), suggesting that pulsation does not specifically require kinA or kinB individually. Together, these results show that pulsing is robust to deletion of a variety of different circuit components. While further elucidation of the mechanism of pulse generation will be important, we focus below on the consequences of pulsing for the deferral of sporulation.

2.3.2 Deferral Time Is Regulated Cell-Autonomously by the Sporulation Initiation Circuit

In principle, the extended multi-cell-cycle timescale for activation of Spo0A could be achieved in three different ways (Figure 2.4): Internal genetic circuitry could generate a slow rise in a critical regulator (CIRCUIT cartoon). Alternatively, an inhibitor of sporulation could gradually dilute out over multiple cell cycles during the proliferation phase (DILUTION cartoon). Either of these two mechanisms would function cell-autonomously. Finally, cells could defer sporulation through a non-cell-autonomous mechanism involving the build-up of extracellular signaling molecules that modulate the phosphorelay (quorum sensing) (Gonzalez-Pastor et al., 2003; Lazazzera, 2000; Lopez and Kolter, 2009) or through degradation of the local micro-environment (QS/ENV cartoon). We performed two experiments that distinguish between these possibilities and, together, support a cell-autonomous mechanism that does not involve dilution for gradual build-up of Spo0A activity.

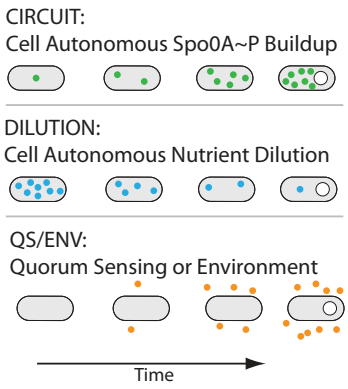


Figure 2.4: **Possible Classes of Deferral Mechanism.** Three classes of mechanism can generate deferred sporulation. Top (CIRCUIT): slow cell-autonomous circuit activation drives accumulation of $Spo0A^P$ over multiple cell cycles. Middle (DILUTION): multiple cell divisions dilute out internal factors, such as nutrients, eventually permitting sporulation. Bottom (QS-ENV): slow changes in environment, possibly from environmental nutrient exhaustion or accumulation of extracellular quorum sensing signaling molecules, eventually permit sporulation.

In the first experiment, we sought to distinguish between cell-autonomous and non-cell-autonomous deferral mechanisms. Low initial cell densities on resuspension media pads did not permit the growth and sporulation of cells, suggesting that at least some cell-generated factors were required for proliferation and possibly sporulation in these conditions. However, it was not clear whether these signals

were responsible for deferring sporulation. To address this question, we developed a co-culture assay where unlabeled cells were mixed with red mCherry-labeled cells on the same pad (Figure 2.5A). The unlabeled wild-type cells were introduced 10 h before the labeled cells, allowing them to condition the pad as they proliferated and eventually sporulated. If deferral were controlled by cell-extrinsic factors, then the red cells should sporulate earlier with the unlabeled cells than without them (Figure 2.5A, lower cartoon). On the other hand, if the deferral of sporulation were cell-autonomous, then the red cells would proliferate for an equal amount of time in the presence or absence of the unlabeled cells (Figure 2.5A, upper cartoon).

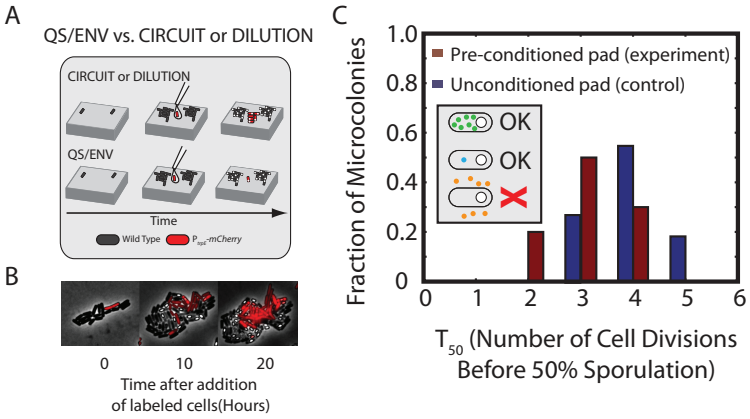


Figure 2.5: **Deferral is not due to environmental changes or quorum sensing.** Preconditioning agar pads with sporulating cells can test the QS/ENV mechanism. (B) Agarose pads are preconditioned by sporulating wild type cells (black). Fresh mCherry-labeled wild type cells (red) are subsequently added to conditioned pads once unlabeled cells near sporulation. If deferral is cell autonomous, the red cells will defer sporulation (top). If deferral depends on environment, red cells will sporulate immediately (bottom). (C) Filmstrip of labeled (red) cells placed on a pad preconditioned by unlabeled cells. (D) Sporulation time histograms (see Materials and Methods) for labeled cells placed on preconditioned resuspension media pads (red bars), compared with labeled cells on unconditioned resuspension media pads as a control (blue bars).

In order to quantify this effect, we counted the number of cell cycles required for 50% of cells in a microcolony, starting from a single labeled cell, to initiate sporulation, as measured by the formation of a phase-bright forespore. Because the actual distribution of deferral times has a tail (Figure 2.2), this measure, denoted as T_{50}, approximates but slightly underestimates the actual mean deferral time as measured in individual cell lineages (Materials and Methods). We found that sporulation of labeled cells was only modestly accelerated by unlabeled sporulating microcolonies (Figure 2.5B,C), reducing T_{50} by 25%, from 4 to 3 cell cycles. This measurement provides an upper limit to cell-

extrinsic effects in our conditions. Although cell-extrinsic factors do play some role, deferral appears to be controlled in a predominantly cell-autonomous fashion.

We next sought to determine whether cell autonomous deferral in our conditions was caused by slow depletion of internal factors following the switch to resuspension media (Figure 2.4, middle panel). One specific molecular candidate for the dilution mechanism is a slow depletion of intracellular GTP levels, which control repression of stationary phase genes by CodY through the alarmone (p)ppGpp (Ratnayake-Lecamwasam et al., 2001). However, in our experimental conditions, a ΔcodY strain showed similar deferral behavior as wild-type cells (Table 2.7.3), demonstrating that codY is not necessary for deferral.

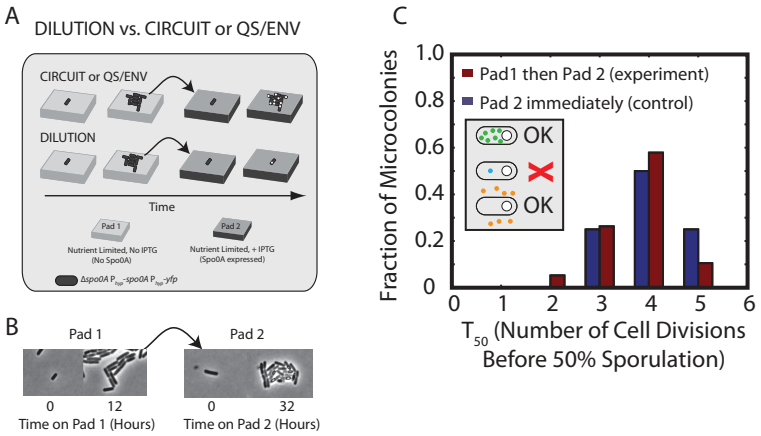


Figure 2.6: **Deferral is not due to dilution of internal factors.** The pre-growth/pad transfer experiment can test the DILUTION mechanism. (A) Cells with an IPTG-inducible spo0A construct (P_{hyp}-spo0A), but lacking the endogenous spo0A gene, are first grown for multiple generations on resuspension media pads without IPTG, diluting out nutrients or other factors present only in rich media. Cells are then transferred to a second resuspension media pad containing 100 μM IPTG. If deferral depends on the dilution of internal nutrients or factors, cells will sporulate immediately. If deferral depends on slow accumulation of a factor, cells will still defer sporulation. (B) Film strip of cells on pad 1 and pad 2. Note growth of individual cells into microcolonies on both pads. (C) Sporulation time histograms for cells placed first on pad 1, then transferred to pad 2 (red bars), compared to a control where cells are placed immediately on pad 2 (blue bars).

Because the dilution mechanism need not work through codY, we designed an experiment to rule out the dilution model more generally (Figure 2.6). In this experiment, a strain with IPTG-inducible spo0A (*spo0A*::P_{hyp}-spo0A) is first grown on one nutrient-limited pad lacking IPTG and then transferred to a second, similar, nutrient-limited pad containing 100 μM IPTG (2.6A). The first pad allows cells to grow for multiple cell cycles without inducing sporulation (Figure 2.13A).

This growth dilutes out putative internal factors not produced in nutrient-limited conditions. On the second pad, IPTG is present, enabling immediate constitutive transcription of Spo0A. After one cell cycle, Spo0A concentration reaches a steady state expression level at or exceeding that in sporulating wild-type cells (Figure 2.13). In the dilution model, dilution of sporulation inhibitors during growth on the first pad would cause cells to sporulate immediately on the second pad. On the other hand, if deferral were due to cell-autonomous Spo0A circuit dynamics, growth on the first pad would have no effect on deferral on the second pad. In fact, the T_{50} distribution on the second pad was not substantially affected by 3-4 cell cycles of growth on Pad 1, with $T_{50} = 4.1 \pm 0.2$ versus 4.4 $\pm$ 0.1 (mean $\pm$ SD) with and without Pad 1, respectively (Figure 2.6C). These results rule out the dilution-driven deferral model. Together, these results strongly suggest that multi-cell-cycle deferral is controlled by an extended cell-autonomous accumulation of $Spo0A^P$.

2.3.3 Kinase Levels Control the Deferral of Sporulation

To better understand how the sporulation initiation circuit controls deferral time, we consider two classes of genes. The first class consists of the phosphorelay genes Spo0A, Spo0F, and Spo0B and the sporulation kinases KinAKinE, whose products directly contribute to the phosphorylation of Spo0A. Limited expression of these genes could potentially defer sporulation by slowing the phosphorylation of Spo0A. The second class consists of phosphorelay phosphatases, whose expression could potentially defer sporulation by draining phosphates from the relay, slowing the accumulation of phosphorylated Spo0A.

We investigated the impact of these phosphorelay components on multi-cell-cycle deferral, distinguishing between two qualitatively different regimes, similar to an approach used previously (Figure 2.7A) [41]: In a relay-limited regime, phosphorelay protein concentrations (e.g. Spo0F and/or Spo0B and/or Spo0A itself) limit the rate of phosphotransfer and thus the level of $Spo0A^P$. In contrast, in a flux-limited regime, the level of $Spo0A^P$ is principally controlled by the rate at which phosphates are injected into the circuit by kinases and/or removed by phosphatases. To experimentally distinguish between the two regimes, we analyzed the behavior of unlabeled wild type cells alongside

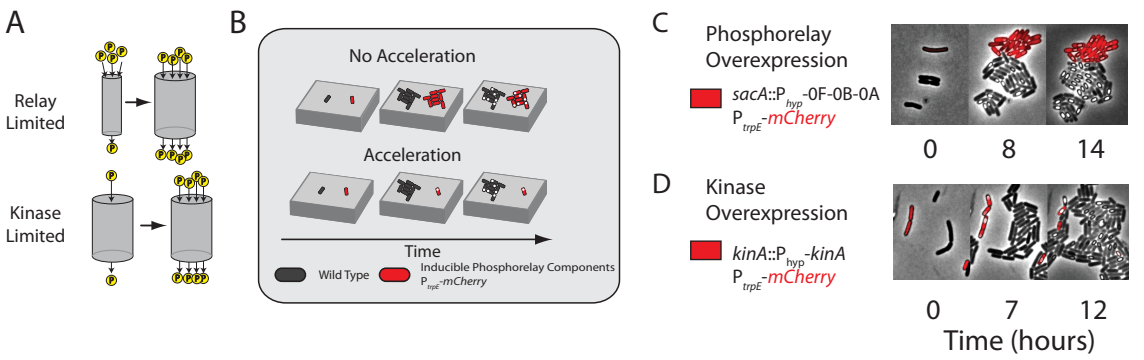


Figure 2.7: **Spo0A$^{\mathrm{P}}$ activity growth is kinase limited** (A) Two alternative regimes in which pulse growth could occur. (Left) The abundance of phosphorelay proteins is represented as the size of a pipe through which phosphates flow, allowing one to distinguish between a phosphorelay-limited regime (upper) where phosphorelay capacity increases with time, versus a kinase-limited regime (lower) where the flux of phosphates into the pipe limits Spo0A-phosphorylation. (B) Experimental scheme for distinguishing between the two regimes. mCherry-labeled cells (red) overexpressing a network component are compared to co-cultured wild type cells (black) grown in the same field of view on the same agarose pad (gray slab). Acceleration of sporulation, as shown, indicates that deferral time is limited by the overexpressed component(s). Co-culturing labeled and unlabeled cells together control for potential day-to-day and pad-to-pad variability. (C) Induced expression of an operon containing spo0F, spo0B, and spo0A fails to accelerate sporulation. (D) In contrast, induced expression of kinA accelerates sporulation.

(cocultured with) mCherry-labeled cells engineered to overexpress different phosphorelay components. Overexpression of limiting components, but not non-limiting components, will accelerate sporulation relative to wild type as shown schematically in Figure 2.7B. Thus, overexpression of spo0A or an operon of phosphorelay components (spo0A, spo0B, and spo0F) should accelerate both Spo0A$^{\mathrm{P}}$ buildup and sporulation in the relay-limited regime, while having little to no effect in the flux-limited regime. Conversely, in the kinase-limited regime, kinase overexpression should accelerate both Spo0A$^{\mathrm{P}}$ buildup and sporulation in the flux-limited regime but have little or no effect in the relay-limited regime. We note that previous related work by Fujita and Losick has established the strong effects of kinA overexpression in a different context, showing that it is sufficient to induce immediate sporulation in rich media conditions, which strongly suppress sporulation altogether (Fujita and Losick, 2005).

We observed little to no acceleration in the onset of sporulation when we expressed spo0A or the spo0A-spo0B-spo0F operon in the labeled cells, despite the ability of these constructs to complement corresponding mutants (Figure 2.7C). These cells sporulated with a $T_{50} = 3.7 \quad 0.2$

(mean SD), similar to 4.0 0.2 in wild-type cells. On the other hand, induced expression of kinA strongly accelerated both the activation of Spo0A, as measured by P_{spo0F} expression, and the onset of sporulation (Figure 2.7D), resulting in $T_{50} = 0.2$ 0.1. These results suggest that the deferral of sporulation is flux-limited, but not relay-limited.

To further test whether kinases or phosphatases were responsible for flux limitation, we constructed strains lacking phosphorelay phosphatases individually and in combinations. Simultaneous deletion of spo0E, rapA, and rapB reduced deferral by about one cell cycle, but did not abolish the multi-cell-cycle deferral. Deletion of other phosphatase genes, including the Spo0A phosphatases yisI and ynzD, and the Spo0F phosphatases rapE, rapH, and rapJ, had no discernible effect (Figure 2.7.3). Evidently, phosphatases alone cannot explain the flux limitation underlying multi-cell-cycle deferral, whereas kinase over-expression is sufficient to abolish multi-cell-cycle deferral. Together, these results implicate the slow buildup of kinase as the predominant deferral mechanism.

This hypothesis is supported by analysis of a P_{kinA}-yfp reporter, which confirmed that KinA concentration indeed builds up gradually in the cell cycles preceding sporulation, and does so to an extent that cannot be explained by the less than 2-fold slowing of growth rate during the experiment (Figure 2.19). Similarly, while cells on Pad 2 in the dilution experiment (Figure 2.6) exhibited systematically slower growth rates than wild type cells (Figure 2.21), they still sporulated with a similar deferral period. Evidently, regulation of kinA expression leads to a progressive increase over multiple cell cycles.

2.3.4 A Pulsed Positive Feedback Loop Controls Kinase Activity

One of the most prominent activators of the principle sporulation kinases kinA and kinB is Spo0A itself. $Spo0A^P$ inhibition of the transcriptional repressor AbrB leads to up-regulation of kinA throughσ^H and de-repression of kinB (Strauch, 1995). Thus, increased kinase activity could be driven by the engagement of a positive feedback loop, in which Spo0A activity pulses activate kinase transcription, increasing the amplitude of subsequent Spo0A pulses, and thus ratcheting up kinase levels once per cell cycle. A comparison of $Spo0A^P$ levels (inferred from P_{spo0F} promoter activity)

with KinA levels (measured with P_{kinA}-yfp fluorescence) demonstrated that kinA expression correlates with $Spo0A^P$ pulse amplitudes (Figure 2.19C). Imaging of a kinA-gfp protein fusion confirmed that KinA protein levels increase during the deferral period (Figure 2.20A). Furthermore, ectopic expression of a constitutively active spo0A mutant, $spo0A^{sad67}$, in a Δspo0A background, led to full up-regulation of a P_{kinA}-yfp reporter (Figure 2.20B), and no reporter expression was observed in this strain without induction of $spo0A^{sad67}$. Together, these results indicate that active spo0A is necessary and sufficient for full kinA expression.

To investigate the potential role of this positive feedback loop, we developed a method to quantify pulse growth in individual cells. First, we characterized each P_{spo0F} promoter activity pulse by its peak amplitude (Figure 2.8A). This allows promoter activity time traces to be represented by a discrete sequence of pulse amplitudes, one per cell cycle. We then plotted these pulse sequences on a 'return map,' where the amplitude of each pulse (labeled p_{N+1}) is plotted against the amplitude of the pulse immediately preceding it (labeled p_N) (Figure 2.8B). Pulse growth causes points on the return map to lie above the diagonal line $p_N = p_{N+1}$. In wild type cells, pulse amplitudes, though variable, tended to grow with successive cell cycles. Thus, on the return map, over two-thirds of data points lie above the diagonal, with the strongest growth at low and intermediate pulse amplitudes (Figure 2.8C). At high pulse amplitudes the trend saturates, so that the amplitude of a pulse eventually becomes independent of its predecessor. These results are consistent with the existence of a saturating positive feedback on kinase expression.

By contrast, if kinase expression were constitutive (Figure 2.8D), then induced kinase expression, and thus $Spo0A^P$ pulse amplitude, would relax to a steady state with a timescale of about one cell cycle (similar to Figure 2.13B), eliminating systematic pulse growth (Figure 2.8E). To test this prediction experimentally, we constructed a 'feedback bypass' strain, combining a ΔkinA ΔkinB ΔkinC triple deletion with IPTG-inducible kinA expression. In this strain, modest levels of IPTG (2 μM) allowed cells to grow and divide multiple times while activating Spo0A. Like wild-type cells, these cells exhibited variable amplitude $Spo0A^P$ pulses correlated with the cell cycle (Figure 2.18). However, lacking transcriptional feedback on kinase expression, the pulse amplitudes showed no

systematic growth over successive cell cycles (Figure 2.8F). These results suggest that feedback on kinase transcription is required for pulse growth.

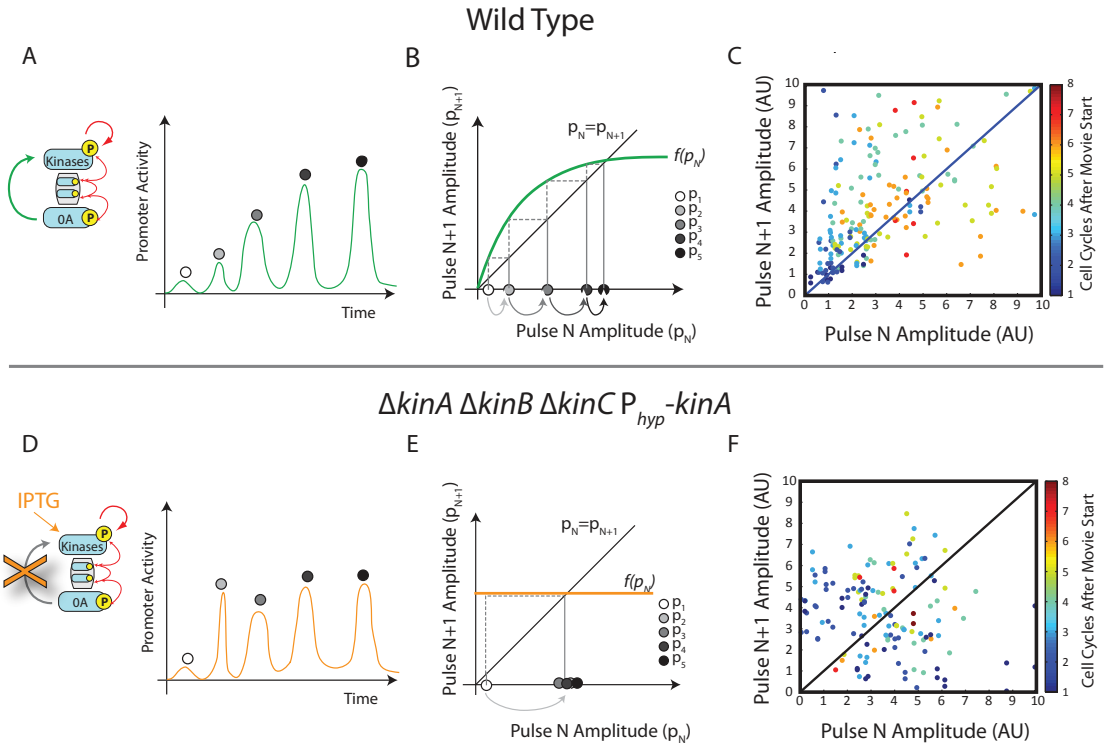


Figure 2.8: **Spo0A^P Return maps reveal that pulse growth depends on kinase feedback.** (A) Pulse dynamics can be analyzed as a sequence of discrete pulse events. A schematic example, showing pulse growth in a hypothetical trajectory for the wild-type circuit (inset), is shown here (green line). Gray circles label successive peak amplitudes. (B) Schematic return map shows the progression of pulse sizes expected assuming a saturating positive feedback (green line). Here, a relatively weak initial pulse (white disk) leads to a larger value for the next pulse, whose value is determined by the feedback function (green line). When the feedback function intersects the diagonal line, successive pulses are equal in size and growth stops. Note that several pulses are required to reach the steady-state. (C) The experimental return map for wild-type cells. 247 individual pairs of successive pulses are plotted as dots. 165 pairs show growth and hence lie above the diagonal, while 82 pairs lie below the diagonal. (DF) The feedback bypass strain (inset in D) shows qualitatively different dynamics. (D) Schematic showing expected behavior of this strain upon transfer to IPTG-containing media. Pulses should rapidly reach steady-state and not grow systematically with time due to the absence of positive feedback. (E) Return map (schematic) showing the progression of pulse sizes expected in a deterministic system lacking feedback, resulting in a flat feedback function (orange line). Note that a single step is sufficient to bring the system to steady-state (gray arrow). (F) The experimental return map for feedback bypass cells shows variability but less systematic pulse growth. 80 pulse pairs lie above the diagonal, while 64 pairs lie below the diagonal.

Furthermore, the lack of pulse growth in the feedback bypass strain predicts an extremely sensitive dependence of sporulation timing on kinase expression levels. In this strain, IPTG concentration controls the steady-state kinase concentration, but not the timescale to reach it, which is set by the

cell division time (Figure 2.9A). At low IPTG levels, Spo0A^{P} can never grow high enough to induce sporulation. Conversely, at high IPTG levels, sporulation would be induced almost immediately. Between these two extremes, sporulation would be deferred for multiple cell cycles only in a narrow window of kinase expression levels.

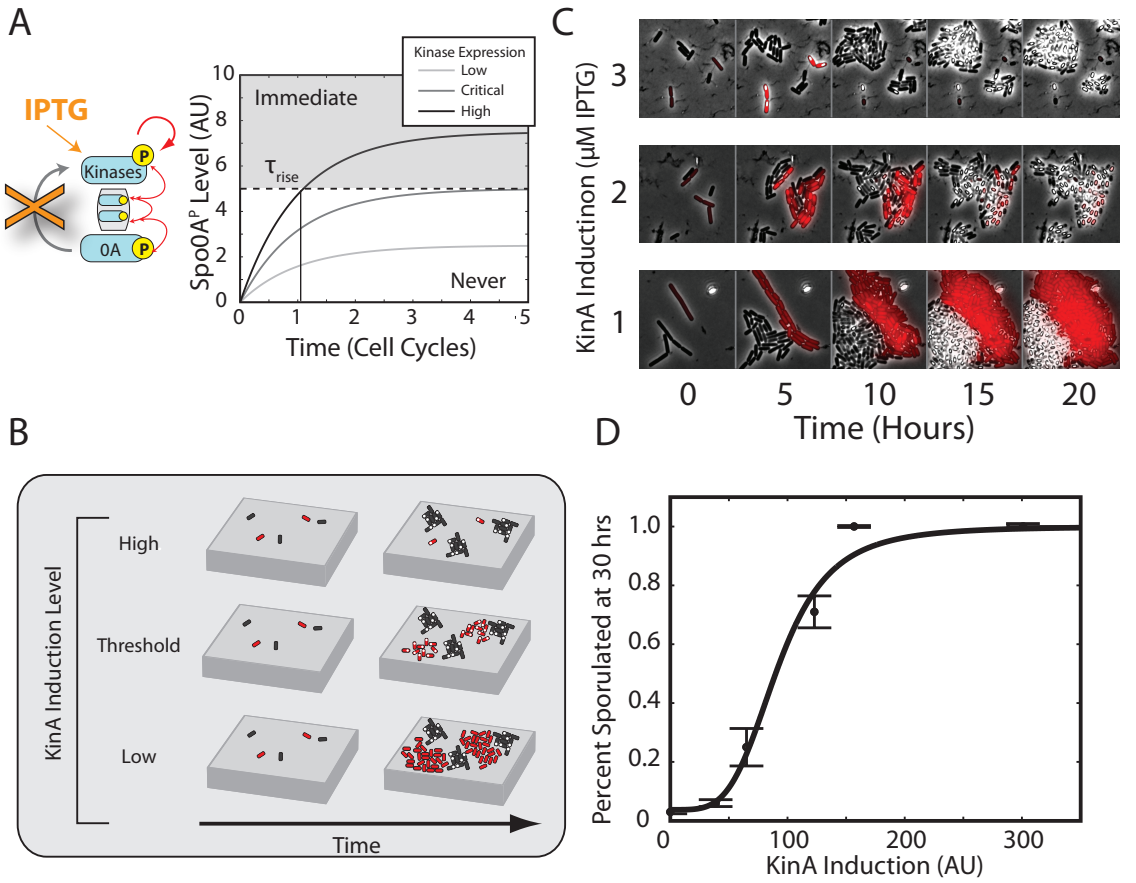


Figure 2.9: **Positive feedback is required for control of rise time.**(A) The deferral time is expected to be ultrasensitive to kinase induction levels in the absence of positive feedback. Excess kinase production causes Spo0A^{P} levels to rise and rapidly cross a sporulation threshold (dashed line), leading to immediate sporulation (gray region). By contrast, lower kinase production rates never produce enough Spo0A^{P} to cross the threshold (white region). Only an intermediate range of tightly tuned expression levels can potentially cross the threshold after significant delay (Critical). (B) Schematic of co-culture assay for comparing the proliferation of ΔkinA ΔkinB ΔkinC P_{hyp}-kinA cells (red) with wild-type controls (black). Note the transition from rapid sporulation (small sporulating red microcolonies, top) to no sporulation (growing red non-sporulating microcolonies) as KinA induction level is decreased. (C) Sporulation timing depends sensitively on the level of kinA induction. Over a small change in IPTG level, we observed a sharp change in sporulation timing in Feedback Bypass cells (red). Note that no such change was observed among wild-type (black) cells. (D) Analysis of movies like those in (C) shows that the percentage of cells sporulating is a sharp function of IPTG as measured by YFP fluorescence from a P_{hyp}-yfp construct and fit to a Hill function with Hill coefficient $n = 4.0 \pm 1.2$ (95% confidence intervals). y-axis error bars are standard error of the mean.

To test this prediction, we compared sporulation timing in our feedback bypass cells, labeled with mCherry, to that of wild-type cells co-cultured on the same agarose pad (Figure 2.9B). At low IPTG induction, these cells largely failed to sporulate, while at high induction levels, cells sporulated within one or two cell cycles (Figure 2.9C). The fraction of sporulated cells at 30 h showed a sharp dependence on kinA induction level, equivalent to a Hill coefficient of 4.1 $\pm$ 1.8 (95% confidence interval) (Figure 2.9D). Together, these results show that positive feedback on kinase expression is necessary for regulated deferral.

2.3.5 Pulsed Positive Feedback as a Mechanism for Deferred Differentiation

How does positive feedback enable cells to set long deferral times, and what role can the pulsatile activation of Spo0A play? To explore these questions we constructed a mathematical model of the sporulation initiation circuit. We used a simplified model (Text S1) in order to gain insights into qualitative differences between different circuit architectures, but not to reproduce all known molecular interactions in the circuit. We modeled pulsatile Spo0A phosphorylation by activating kinase autophosphorylation for a fixed fraction of each cell cycle. We also simplified the phosphorelay into a two-component phosphotransfer from kinase to Spo0A. Although they are likely to be important for some aspects of the natural system, inclusion of Spo0F and Spo0B does not qualitatively affect the conclusions below. Finally, based on the insensitivity of deferral time to phosphatase deletions, we modeled phosphatase activity with a constant level of Spo0E.

Sporulation initiation is believed to require a threshold level of phosphorylated Spo0A. Indeed, we found that maximal P_{spo0F}-yfp promoter activity in sporulating cells was systematically higher than in vegetative cells (Figure 2.22). Recently published experiments in bulk cultures have also demonstrated that cells sporulate at a threshold level of KinA. Therefore, to analyze deferral, we quantified the number of cell cycles required for phosphorylated Spo0A to grow from a low initial level to a high threshold level.

We performed this analysis for three distinct circuit architectures, which differ in how kinase

expression is controlled (Figure 2.10): In the first circuit, kinase is produced constitutively (open loop, Figure 2.10A). In the second, kinase production is instantaneously activated by $Spo0A^P$ (pulsed instantaneous positive feedback, Figure 2.10B). In the third circuit, kinase is indirectly activated by $Spo0A^P$, leading to an effective time delay (t) between Spo0A phosphorylation and consequent up-regulation of kinase expression. If the $Spo0A^P$ pulse terminates before kinase expression initiates, the pulsing and time delay together effectively divide the deferral period into distinct phases where either $Spo0A^P$ pulsing or kinase transcription (or neither) is active, but never both; we call this type of feedback 'polyphasic feedback' (Figure 2.10C).

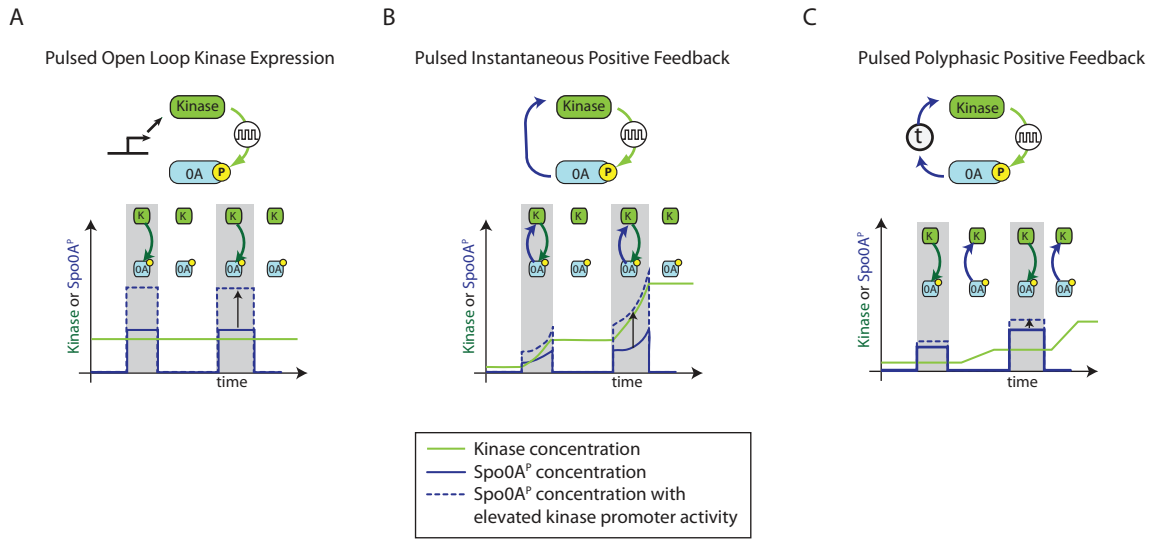


Figure 2.10: **Circuit models to generate multi cell cycle deferral.** Three simplified models of pulsed $Spo0A^P$ regulation can generate extended, tunable, deferral times. (A) In the pulsed open loop circuit, kinase activates Spo0A in pulses (gray regions), but there is no feedback. As a result, kinase levels are constant (green line), and changes in kinase expression level directly affect pulse amplitude, as shown by the dashed, compared to the solid, blue line. (B) In the pulsed instantaneous feedback circuit, increases in $Spo0A^P$ cause immediate increases in kinase expression, reflected in the accelerated growth of kinase and $Spo0A^P$ expression during, but not between, pulses. Thus, both arms of the feedback loop are active during, but not between, pulses (cartoon inserts). This results in strong sensitivity to kinase expression level and other parameters (compare solid and dashed blue lines). (C) In the polyphasic circuit, a time delay in the feedback loop causes periods of kinase accumulation to predominantly occur between, rather than during, $Spo0A^P$ pulses. Effectively, the two arms of the feedback loop are active at different times (cartoon insets). As a result, an increase in kinase expression level has a more modest effect on pulse growth (dashed versus solid blue line).

In the polyphasic mechanism, could represent a number of possible intermediate processes including indirect regulation as well as transcriptional and translational time delays. Our current methods cannot firmly establish nor rule out such time delays, due to the relatively low time resolution in-

herent in promoter activity measurements made with fluorescent protein reporters. For example, we could not detect a consistent time difference between pulses of P_{kinA}-yfp and P_{spo0F}-cfp promoter activity (Figure 2.19B). Higher time resolution and methods to track protein phosphorylation in individual cells could help to constrain the exact magnitude of such effects.

For each circuit, we systematically modulated kinase or phosphatase production rates, both of which directly control phosphate flux to Spo0A. For each production rate, we monitored the time required for $Spo0A^P$ to exceed a fixed threshold (deferral time, Figure 2.11A), and computed the sensitivity of this time to parameters such as the kinase production rate, (Figure 2.11B). The open loop circuit showed an extremely sensitive dependence of deferral time on phosphorylation parameters (Figure 2.11 D& E, blue), consistent with the sensitive dependence on kinase production observed experimentally (Figure 2.9D). The positive feedback loop reduced this sensitivity (Figure 2.11 D & E, green), and the polyphasic feedback loop reduced it still further (Figure 2.11 D& E, red). Models were tuned so the steady-state $Spo0A^P$ levels in the polyphasic circuit exceeded those in the positive feedback circuit, ensuring that longer deferral times were not caused by lower steady states. Positive feedback, especially in the polyphasic regime, evidently could make it easier for cells to regulate multi-cell-cycle deferral times by reducing the sensitivity of deferral time to key control parameters (Suel et al., 2007).

How can we explain the relative sensitivities of the three circuits? In the open loop circuit, protein dilution due to cell growth determines the kinase concentration kinetics. The dilution rate is determined by the cell cycle time, making it difficult to achieve deferrals longer than a single cell cycle. In the positive feedback circuit, protein production and dilution compete with each other to set the timescale of kinase accumulation. Parameters that affect net feedback strength (e.g., kinase or phosphatase promoter strengths) directly tune this timescale, and thereby modulate deferral time. Finally, the polyphasic positive feedback circuit includes the benefits of positive feedback. In addition, however, the combination of $Spo0A^P$ pulsing and a time delay in its feedback onto kinase production together cause most of the new kinase produced by a pulse to appear only after the pulse terminates. Consequently, kinase cannot instantaneously feed back to amplify the pulse that

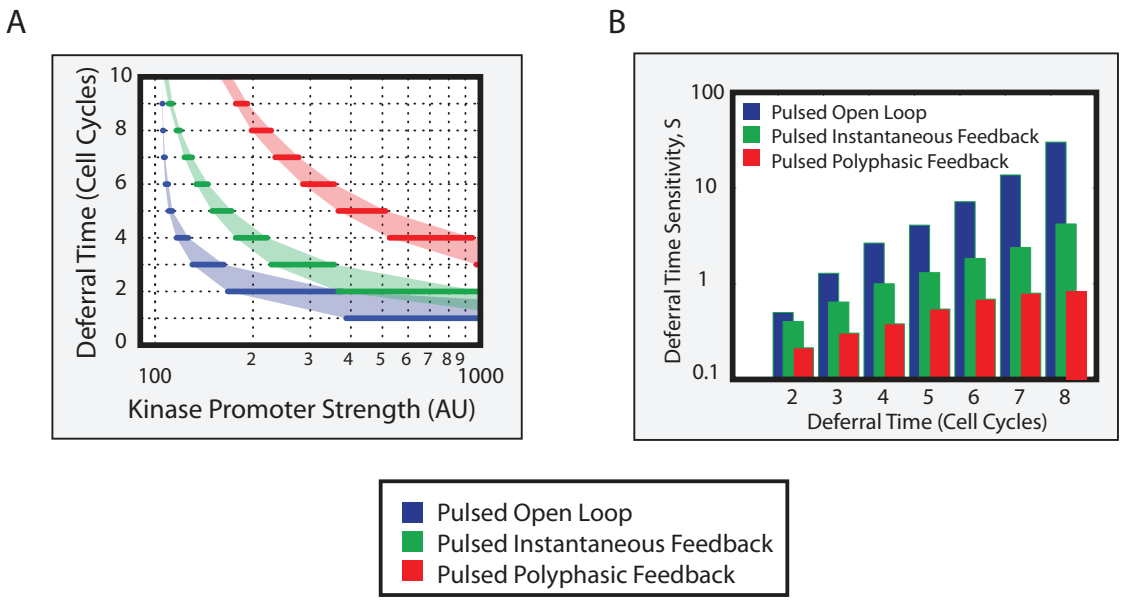


Figure 2.11: **Rise time sensitivities to feedback strength, by circuit model.** Mathematical models of the three circuits were compared for their ability to generate multi-cell-cycle deferral times. For each circuit, the kinase promoter strength was tuned to produce different deferral times (x-axis). The sensitivity of deferral time to kinase promoter strength was defined as $S \equiv \frac{\partial \log T_D}{\partial \log \beta_K}$, where T_D is the deferral time in cell cycles, and β_K is the strength of the kinase promoter. The three circuits differ systematically in both the magnitude and rate of increase of sensitivity with deferral time with the open loop circuit being most sensitive (cf. 2.9), followed by the instantaneous and then the polyphasic circuit. Reduced sensitivity enables the cell to more accurately regulate deferral times. Note log y-axis scale, and see Model Details.

produced it. Since feedback occurs from pulse to pulse, rather than compounding continuously as in standard positive feedback, pulse growth is much less sensitive to changes in feedback strength. Qualitative insights into the three circuit architectures can be obtained by analytically solving a set of corresponding simplified one-dimensional models (Text S1 and Figure 2.23). In these simplified models, although protein concentration grows exponentially in both positive feedback and polyphasic circuits, the time constant of the polyphasic circuit is exponentially less sensitive to feedback strength (Figure 2.24).

2.4 Discussion

To respond properly to the challenges posed by environmental and developmental constraints, cells respond to stimuli across widely varying timescales. In some systems, the challenge is to achieve

extremely rapid responses (Segall et al., 1982). In other cases, however, cells may face the opposite challenge of deferring a response for relatively long times. Sporulation initiation represents an ideal example, where a sudden change in environment leads to a particular responsesporulation only after many cell cycles. Although sporulation is deferred, it is clear that cells respond to the change in conditions throughout the deferral period, for example through continual increases in Spo0A activity.

In principle, several different mechanisms can produce a deferred response. Quorum sensing mechanisms can defer activation of a response until a critical cell density is reached, as occurs in the V. harveyii light production circuit (Waters and Bassler, 2005). Our data do not rule out a role for quorum sensing, but show that it cannot explain most of the multi-cell-cycle delay observed here (Figure 2.5C). Furthermore, since quorum sensing is a response to absolute cell density, rather than to rounds of cell division, it may be better suited to measuring population size as opposed to time intervals. A second potential mechanism for deferral is dilution of an internal molecule that represses sporulation. Dilution failed to explain sporulation deferral in our experiments (Figure 2.6C). A dilution mechanism requires the regulator to be produced continually before nutrient limitation at a level tuned to provide the appropriate deferral time during nutrient limitation. Thus, this strategy might be better adapted to a more deterministic environment, such as multicellular development (Raff, 2007), rather than the more unpredictable environments that microbes experience (Singh et al., 2008). In contrast, the cell-autonomous feedback-dependent mechanism analyzed here allows deferral time to be quickly tuned from immediate to multiple cell cycles under different conditions. Indeed, previous studies of sporulation have used conditions optimized to induce immediate, rather than deferred, sporulation with the same circuit (Sterlini and Mandelstam, 1969). The relative advantages of each type of mechanism may become clearer as additional examples of deferred differentiation are identified and elucidated.

Feedback loops are known to affect the response times of gene circuits. Negative feedback has been previously shown to accelerate responses (Rosenfeld et al., 2002). Here we demonstrate the complementary role of positive feedback in extending timescales. This latter function is particularly important when proliferating cells need to postpone responses beyond a single cell cycle, the longest

fundamental timescale of protein turnover in growing cells. Positive feedback extends timescales by competing with protein dilution to set the net relaxation time for protein concentrations. In *B. subtilis* sporulation, our results reduce the overall circuit to a core two-element positive feedback loop involving the master regulator $Spo0A^P$ and the sporulation histidine kinase KinA. This feedback loop progressively ratchets up $Spo0A^P$ levels, approaching the threshold level required for sporulation only after multiple cell cycles, and thereby enabling multi-cell-cycle deferral.

A striking aspect of the system analyzed here is pulsatile phosphorylation of Spo0A. Pulsing imposes additional temporal structure on circuit dynamics that can lead to novel regulatory capabilities. For example, the yeast transcription factor Crz1 undergoes discrete pulses of nuclear localization at a frequency set by extracellular calcium concentration (Cai et al., 2008). The relative fraction of time Crz1 spends in and out of the nucleus is determined by the pulse frequency. This 'FM' regulation sets the fraction of time that all of Crz1's targets are activated, leading to proportionally coordinated expression of the entire regulon.

In addition to quantizing responses, pulsing can also dictate the relative timings of different interactions in a circuit. Here, the pulsed buildup of $Spo0A^P$ defers sporulation for multiple cell cycles through a $Spo0A^P$-KinA positive feedback loop. When time delays are present in this feedback loop, increased KinA production occurs after the $Spo0A^P$ pulse ends. As a result, $Spo0A^P$ production and KinA production are temporally separated. In this 'polyphasic' regime, pulsing and time delay work together to prevent instantaneous feedback, making the buildup rate significantly less sensitive to parameter values than it would be in a conventional positive feedback loop. It will be interesting to develop techniques that can access these dynamics with higher time resolution, and to see if this polyphasic strategy provides a general design principle for regulation of multi-cell-cycle deferral times in other systems.

Finally, one can ask whether deferring sporulation might have other benefits in addition to enabling proliferation. *B. subtilis* cells could explore alternative cell fates such as competence, biofilm formation, or cannibalism, during the deferral period. In this way, deferred progression to sporulation, implemented by a simple cell-autonomous pulsed positive feedback circuit, provides a

critical foundation upon which multifaceted developmental programs can unfold.

2.5 Materials and Methods

2.5.1 Time Lapse Microscopy

2.5.1.1 Cell Preparation

Overnight cultures grown to saturation in shaking CH (Casein Hydrolysate) media at 37C with antibiotic selection were rediluted 1:100 into 2.5 ml fresh CH media without selection. The resulting culture was then grown at 37C to an OD600 of 0.81.0. Cells were then washed 2 and finally resuspended in the same final volume with fresh room temperature Resuspension Media (RM) (Donnellan et al., 1964). 0.5 l of resuspended cells was then spotted onto an appropriate agarose pad (see next paragraph). The cells and agarose pads were then covered, dried for approximately 15 min at room temperature, and then inverted into a glass coverslip bottom dish (Willco HBSt-5040), which was then parafilm sealed for imaging.

2.5.1.2 Agarose Pad Preparation

Agarose pads were prepared by melting 1.5% weight by volume low melting point agarose (OmniPur, EMD) into RM. 1 ml of liquid RM-agarose was then pipetted onto a 22 mm square coverslip, covered with a second coverslip, and allowed to dry covered at room temperature for at least 1 h. For experiments involving transgene induction, an appropriate amount of IPTG was added to cooled liquid RM-agarose and thoroughly mixed before pipetting onto the coverslip.

2.5.1.3 Microscopy

Cell growth and sporulation was observed at 37C using an Olympus IX-81 inverted microscope controlled with custom software. Multiple stage positions were monitored using a motorized stage from ASI. Fluorescent reporters were excited by a 175 Watt Lambda LS Xenon arc lamp (Sutter), with typical exposure times between 200 ms and 1 s. Excitation light was filtered with a combination

of neutral density and UV/IR filters to minimize phototoxicity and photobleaching. Images were recorded using an ORCA-ER camera (Hamamatsu) at a frame rate of once every 10 min, unless otherwise noted.

2.6 Data Analysis

Custom MATLAB software, similar to that described in Rosenfeld et al. (Rosenfeld et al., 2005), was used to extract time lapse fluorescence values for individual cells and lineages in a microcolony. All subsequent data analysis was also done in MATLAB using customized software.

2.6.1 Definition and Extraction of Promoter Activity

We define promoter activity as the protein production rate from an individual allele of a promoter. To estimate this quantity from time lapse fluorescent data, consider reporter production dynamics from a promoter. In a given cell the total amount of fluorescent protein over time is denoted as $F(t)$. The promoter activity $P(t)$ represents the rate of production of $P(t)$. Fluorescent protein is also degraded, diluted, and photobleached with a combined first-order rate constant γ. In our conditions, γ is dominated by dilution (Eldar et al., 2009). Thus, using a dot to denote the time derivative:

$$\dot{F}(t) = -\gamma F(t) + P(t)$$

Using this relation we could determine $P(t)$ from fluorescence time-series data by using a simple approximation for the derivative. For example, an Euler approach would yield the following expression:

$$P(t) = \frac{F(t) - F(t - \Delta t)}{\Delta t} + \gamma F(t)$$

This type of explicit method, using a cell's total fluorescence, is sensitive to segmentation errors, such as the exact size and shape of the image region identified with the cell. We therefore derive an

expression for promoter activity that uses only the observables of cell length and mean fluorescence, which we have found to be more robust to segmentation errors. First, we note that a cell's total fluorescence $F(t)$ is its mean fluorescence $M(t)$ multiplied by its area $A(t)$. Since the widths of *Bacillus subtilis* cells remain essentially constant, we can write the following:

$$F(t) = A(t)M(t) = WL(t)M(t)$$

Here, W represents the constant cell width, and $F(t)$ represents the time varying cell length. Inserting this into the definition of promoter activity, invoking the differentiation chain rule, neglecting the constant factor, and dividing by $L(t)$, we are left with the following expression:

$$\tilde{P} \equiv \frac{P(t)}{L(t)} = (\mu(t) + \gamma)\, M(t) + \dot{M}(t)$$

Here we have defined to be the cell's instantaneous growth rate. We have introduced as an approximation to the production rate per chromosomal equivalent, facilitating comparison of production rate across all phases of the cell cycle. To understand the two terms on the right, consider two extreme limits. In the first limit, imagine the cell is growing but the mean fluorescence level remains unchanged, . The first term is nonzero, however, implying that promoter activity results from production balancing out the effects of dilution. A second limit occurs when cell growth is negligible but mean fluorescence is increasing. In this case protein production is reflected in increasing protein concentration. In the main text, promoter activity refers to , as defined above.

2.6.1.1 Characterization of Promoter Activity Pulses

We statistically characterized promoter activity pulses using custom software routines written in MATLAB. We used a multi-step method to determine pulse locations in our time traces. Individual time traces were first filtered with a nonlinear smoothing method that replaces each point with the average of the points in a fixed window around it, after rejecting the highest point and the lowest point in the window. We used a window size of 7 (3 points on either side of center). We

verified that our smoothing did not introduce any significant phase shift that would affect our temporal measurements. Potential pulses were then defined as local maxima in each smoothed trace (determined using a sliding boxcar method). For each putative pulse, we extracted its width (defined as the number of frames around the maximum with negative second derivative) and height, which we used to compute the pulse area. We set a minimum pulse size to prevent spurious identification of noise as pulses.

2.6.1.2 Definition of T_{50}

To rapidly quantify typical sporulation times across multiple sporulating microcolonies, we use the T_{50} statistic described in the text. Briefly, in each microcolony we noted when 50% of cells exhibited a phase bright spore and then counted the total number of spores and non-spores. T_{50} is $\log_2$ of this number. T_{50} underestimates in two ways the mean number of generations for which cells defer sporulation. First, a standard result states that in a binary tree with N leaves, the average path length from a leaf to the root is lower bounded by $\log_2 N$. Second, because half the cells in each measurement have not sporulated, any additional growth and divisions by these cells will not be accounted for by T_{50}.

2.7 Model Details

2.7.1 Analytic Model of Pulsed Positive Feedback Dynamics

We explore here how pulsing affects the ability of positive feedback loops to generate and control extended deferrals. We first consider a simple set of one-dimensional dynamical systems, which use positive feedback to grow exponentially with time. These systems lack saturation but nevertheless provide qualitative insights into the effects observed in more realistic models. We will analyze the time it takes for each system to start from an initial low level of activation and grow to cross a higher threshold level, and how this "rise time" depends on the feedback strength in the system.

Each of these models considers a single protein concentration $x(t)$ which responds to two opposing

effects: first order decay with time constant γ modeling protein dilution from cell growth, and a positive feedback term $f(x)$ which serves to increase $x(t)$. The combination of these two mechanisms provides a highly simplified phenomenological model for how a protein's concentration feeds back to regulate its own transcription. The system's dynamics are thus (dot denotes time derivative):

$$\dot{x}(t) = -\gamma x(t) + f(x(t))$$

Within this framework, we will consider different forms of positive feedback, $f(x)$, and how they affect the rise time, defined as the time required for $x(t)$ to grow from x_0 to a threshold value x_F. This analysis will demonstrate how different pulsed positive feedback schemes affect the growth dynamics of $x(t)$, and therefore, the rise time.

2.7.1.1 Constitutive Production

We first consider a model lacking feedback, where the production of $x(t)$ is constant, e.g. $f(x) = \beta$. This model could represent the case where kinase is controlled by a constitutive promoter. In this case $x(t)$ grows towards a steady state value β/γ with time constant γ set by the cell cycle.

$$x(t) = \frac{\beta}{\gamma}\left(1 - e^{-\gamma t}\right) + x_0 e^{-\gamma t}$$

If $\beta/\gamma > x_F$ the system will eventually cross the threshold; if not, it will never do so. The curves in Fig. 5A are solutions to this model. The case with pulsed constitutive production behaves similarly (Figure 2.23, left column), with the average value of $x(t)$ converging to steady state in about one cell cycle.

2.7.1.2 Positive Feedback

We next consider a model where $x(t)$ feeds back to activate its own production, e.g. $f(x) = \beta x$, representing simple linear positive feedback. In this case $x(t)$ grows exponentially with rate $\beta - \gamma$:

$$x(t) = x_0 e^{(\beta-\gamma)t}$$

If the positive feedback strength is greater than the dilution rate ($\beta > \gamma$),will eventually exceed the threshold; if not, it will either dilute away or, in the marginal case remain unchanged.

2.7.1.3 Pulsing

What happens if the production occurs in pulses, as it does in the *B. subtilis* sporulation initiation circuit? For simplicity we consider square pulses of unit area:

$$U_{\Delta t} = \begin{cases} \frac{1}{\Delta t} & \text{if } 0 \leq t \leq \Delta t \\ 0 & \text{else} \end{cases}$$

The pulses occur periodically with period T. Since Spo0A pulses occur approximately once per cell cycle, T would be the cell cycle time.

2.7.1.4 Pulsed Constitutive Production

We consider a network where constitutive production occurs in pulses (Figure 2.23). This is analogous to the network considered in Figure 6A.

$$\dot{x}(t) = -\gamma x(t) + \beta \sum_{n=0}^{\infty} U_{\Delta t}\left(t - nT\right)$$

Here, the time averaged value of $x(t)$ relaxes to a steady state in about one cell cycle, as in the non-pulsed case.

2.7.1.5 Pulsed Positive Feedback

The dynamics in this case become:

$$\dot{x}(t) = -\gamma x(t) + \beta \sum_{n=0}^{\infty} x(t) U_{\Delta t}\left(t - nT\right)$$

It is convenient to solve this model one period at a time, e.g. for $0 \leq t < T$, $T \leq t \ < 2T$, etc. Integrating through the first cell cycle yields

$$x(T) = x_0 e^{\beta} e^{-\gamma T}$$

After N cell cycles

$$x(NT) = x_0 e^{N\beta} c^{-\gamma NT}$$

Thus, in this model, from period to period, $x(t)$ grows exponentially with a growth rate constant, $\beta - \gamma T$. Note that this growth rate depends linearly on the feedback strength β.

2.7.1.6 Polyphasic Positive Feedback

Finally, we consider a second type of feedback model where $x(t)$ activates its own production in pulses. In this model, although each pulse produces additional $x(t)$, it cannot feed back in time to amplify $x(t)$'s production during that particular pulse. This case is analogous to the model in Figure 6C, where kinase production caused by a $\mathrm{Spo0A^P}$ pulse is delayed until after the pulse finishes. Thus, the production rate of $x(t)$ during each pulse is set by the amount of $x(t)$ at the start of the pulse. The full dynamics are:

$$\dot{x}(t) = -\gamma x(t) + \beta \sum_{n=0}^{\infty} x(nT) U_{\Delta t}(t - nT)$$

Integrating through one period, as before, and taking the short pulse limit as $\Delta t \to 0$,

$$lim_{\Delta t \to 0} x(T) = (1 + \beta)\, e^{-\gamma T}$$

After N cell cycles

$$x(NT) = e^{\log{(1+\beta)^N}} e^{-\gamma NT}$$

In this model, from period to period, $x(t)$ grows exponentially with a rate constant of $\log(1+\beta) - \gamma T$. Note that this growth rate depends logarithmically on the feedback strength β, instead of linearly in the instantaneous positive feedback case. Identical scaling is obtained if the feedback is taken to be a Dirac δ function, or if we account for dilution of $x(nT)$ during each pulse.

2.7.1.7 Model Comparisons

How do the models compare in their ability to set deferral times (Figure 2.24)? For each model, we calculated each deferral time's sensitivity to changes in feedback strength. We defined deferral time sensitivity as $S = \partial \log T_D / \partial \log \beta$, as in Figure 6E. In the open loop model, deferral times longer than 1 cell cycle were extremely sensitive to changes in β (blue bars). In the instantaneous positive feedback model (green bars), deferral times were significantly less sensitive to β because positive feedback competes with protein dilution to set the growth rate of $x(t)$. Finally, deferral times in the polyphasic positive feedback model were even less sensitive to feedback strength (red bars). This extra reduction in sensitivity arises because production of $x(t)$ during each pulse is constant, rather than continuously compounding as in the instantaneous feedback case.

2.7.2 Computational Model of Spo0A Network Dynamics

To analyze the effects of pulsing and indirect feedback on the accumulation of Spo0A activity, we considered a set of dynamical models corresponding to the cartoons in Figure 2.10. The models consist of a sporulation histidine kinase (K) which phosphorylates a response regulator representing Spo0A (A). Phosphorylated Spo0A (A^P) feeds back to activate its own transcription directly and to activate transcription of K through an intermediate regulator D. A constant phosphatase activity (k_P) analgous to Spo0E drains phosphates directly from the response regulator. To reproduce the pulsing in Spo0A phosphorylation we modeled the autophosphorylation from the kinase to Spo0A as gated by a square wave $s(t)$, which takes the value 0 except during the third quarter of each cell cycle where it takes the value 1.

Model reactions, and corresponding rate constants, are summarized as follows, with ϕ denoting

degradation of the corresponding molecule:

Table 2.1: List of Reactions

Reaction	**Description**
$K \xrightarrow{\alpha_A s(t)} K^P$	Kinase Autophosphorylation
$K^P \xrightarrow{\alpha_D} K$	Kinase Autodephosphorylation
$K^P + A \xrightarrow{k_F} K + A^P$	Forward Phosphotransfer
$K + A^P \xrightarrow{k_R} K^P + A$	Reverse Phosphotransfer
$A^P \xrightarrow{k_P} A$	Desphosphorylation by Phosphatase
$K \xrightarrow{\gamma} \phi$	Degradation/Dilution of Unphosphorylated Kinase
$A \xrightarrow{\gamma} \phi$	Degradation/Dilution of Unphosphorylated Spo0A
$A^P \xrightarrow{\gamma} \phi$	Degradation/Dilution of Phosphorylated Spo0A
$D \xrightarrow{\gamma_D} \phi$	Degradation/Dilution of D
$\xrightarrow{f_A(A^P)} A$	Production of Unphosphorylated Spo0A
$\xrightarrow{f_D(A^P)} D$	Production of D
$\xrightarrow{f_K(D)} K$	Production of Unphosphorylated Kinase

We modeled the transcription of A and K using Hill functions $f_A(A^P)$ and $f_K(D)$ respectively. We assume that wild type promoters are activated by their cognate transcription factor. In this case $f_A = f_A(A^P) = \beta_A \frac{A^P}{K_D + A^P}$, and $f_K = f_K(D) = \beta_K \frac{D}{K_D + D}$. For simplicity we assume the promoter of the intermediate delay element is linear, $f_D = f_D(A^P) = \beta_D A^P$.

We modeled these reactions as ordinary differential equations for the concentrations of the four reactants K, K^P, A, A^P, as well as D. Equations were numerically integrated in Matlab using the built in ode15s solver. The equations are:

$$\frac{dK}{dt} = -\gamma K + f_K(D) - \alpha_A s(t) K + \alpha_D K^P + k_F K^P A - k_R K A^P$$

$$\frac{dK^P}{dt} = -\gamma K^P + \alpha_A s(t) K - \alpha_D K^P - k_F K^P A + k_R K A^P$$

$$\frac{dD}{dt} = -\gamma_D D + f_D(A^P)$$

$$\frac{dA}{dt} = -\gamma A + f_A(A^P) - k_F K^P A + k_R K A^P + k_P A^P$$

$$\frac{dA^P}{dt} = -\gamma A^P + k_F K^P A - k_R K A^P - k_P A^P$$

The choice of kinetic parameters was guided by previously published biochemical results, along with our own data. The protein decay constant γ was chosen to match the average cell cycle time of 2 hrs in our experiments, since none of the proteins we model are known to be unstable. Kinase autophosphorylation α_A was chosen as linear since the likely large molar excess of cellular ATP over kinase(Segall et al., 1982) allows the reaction to be approximated as a Michaelis-Menten reaction in the linear regime. The order of magnitude of the autophosphorylation constant is consistent with previously determined reaction rates. The kinase autodephosphorylation rate α_D was chosen to match the previously measured dephosphorylation rate of KinA. Forward and reverse phosphotransfer rates were set large enough to not be rate limiting and equal to each other allow reversible phosphotransfer.

Spo0A promoter strength was set to ensure approximately 10,000 Spo0A molecules are present when cells reach the sporulation threshold. As expression from P_{kinA} and P_{kinB} promoter fusions was consistently weaker than expression from P_{spo0A} in our conditions, the kinase promoter strength range was set to a value 10 - 100 times weaker than spo0A promoter strength. The binding constant of activators to these promoters was chosen as 100 molecules per cell, which is greater than the 64nm binding constant of the high affinity spo0A repressed promoter abrB but lower than that of many Stage II sporulation genes.

Sporulation was defined to have occurred when the cell cycle averaged amount of Spo0A exceeded 200 molecules per cell. We chose this threshold as similar to the binding constant of many Stage II sporulation genes.

For the polyphasic positive feedback model, the dilution rate γ_D of the intermediate D was set equal to that of all other proteins. For the instantaneous positive feedback model γ_D was set to be

Table 2.2: Spo0A Network Model Parameters

Parameter (units)	**Open Loop**	**Instantaneous Feedback**	**Polyphasic Feedback**
α_A (1/hr)	1500	1500	2200
α_D (1/hr)	5	5	5
k_F (1/hr-molecules)	1000	1000	1000
k_R (1/hr-molecules)	1000	1000	1000
k_P (1/hr)	50	50	60
γ (1/hr)	0.35	0.35	0.35
γ_D (1/hr)	NA	35	0.35
β_A (molecules/hr)	10000	10000	10000
β_D (molecules/hr)	NA	3.5	0.035
β_K (molecules/hr)	Varied	Varied	Varied
K_D (molecules)	100	100	100

100 fold greater than the dilution rate. To maintain equivalent steady state D levels, we set β_D in the instantaneous model to be 100 times that in the polyphasic feedback model.

We tuned our circuits to facilitate a controlled comparison of rise times between instantaneous positive feedback and polyphasic positive feedback models. We chose parameters so that for a given β_K, steady state cell cycle averaged A^P concentrations in the polyphasic model was greater than or equal to those of the instantaneous model. This ensures that at a given β_K the slower deferral in the polyphasic model is due solely to slower protein accumulation, and not to a lower steady state.

2.7.3 Plasmids and Strain Construction

Full details of strain construction are available in the published paper online at PLoS Biology.

Genotype	**Cell Cycles (Mean / Std Error)**
Δspo0E ΔrapA ΔrapB (JL287)	3.5 ± 0.1
ΔrapA ΔrapB (JL184)	3.2 ± 0.2
Δspo0E ΔynzD ΔyisI (JL289)	3.3 ± 0.2
ΔcodY (JL230)	4.5 ± 0.2
ΔrapE (JL260)	4.3 ± 0.2
ΔrapH (JL296)	4.2 ± 0.1
ΔrapJ (JL300)	4.6 ± 0.2

Table 2.3: **Negative regulators of sporulation initiation are not required for a multi-cell-cycle deferral.** For each strain (left column), 10 sporulating colonies were tracked with time lapse microscopy and mean sporulation time in cell cycles quantified using the T_{50} statistic (right column). Despite some day-to-day variation, no strain ever exhibited mean sporulation time less than three cell cycles.

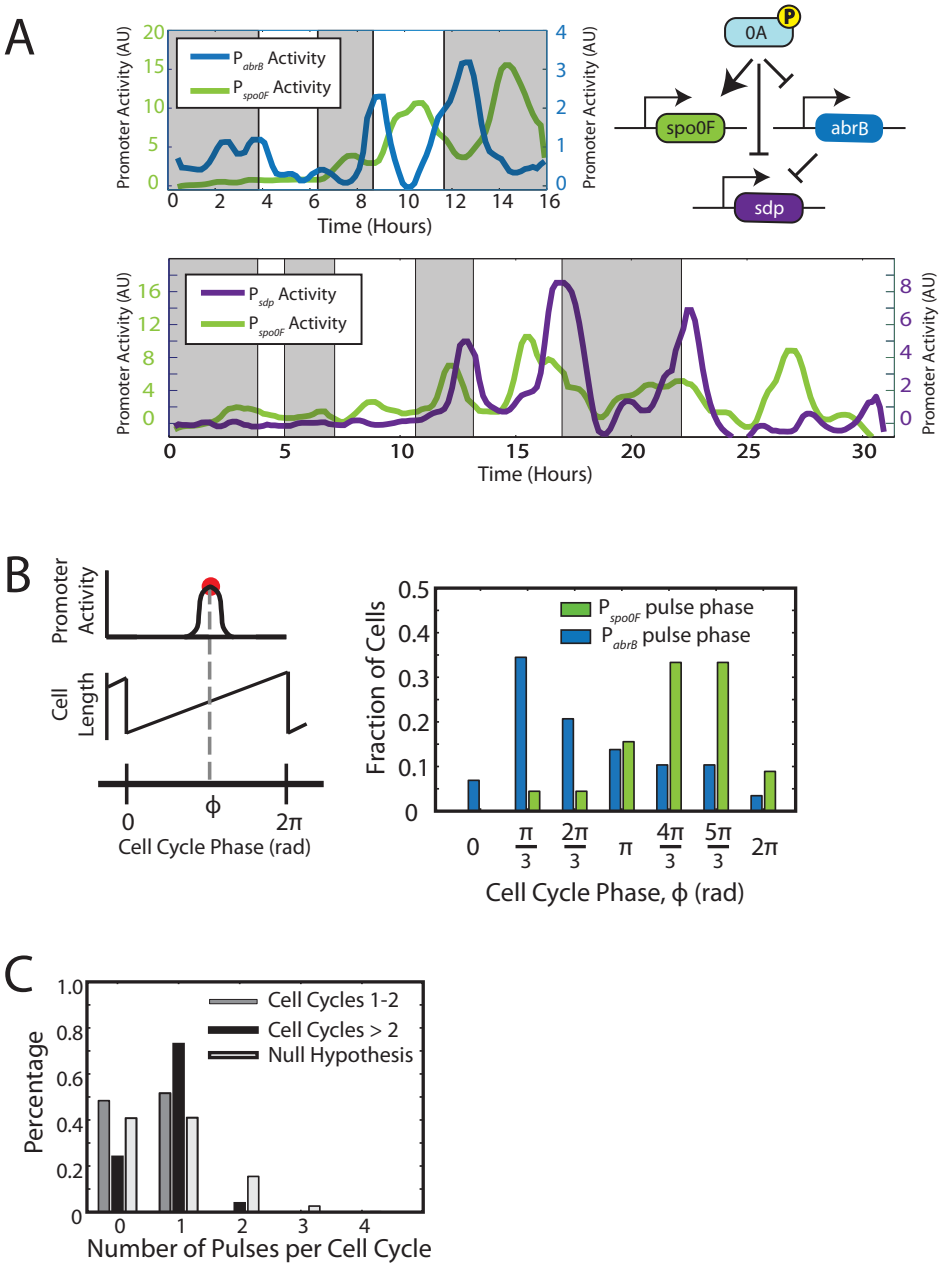


Figure 2.12: **Spo0A$^{\mathbf{P}}$ targets typically pulse once per cell cycle with defined phases.** (A) Single cell time lapse traces of promoter activity in P_{abrBB}-cfp/P_{spo0F}-yfp cells (JL013, top) and P_{sdp}-cfp/P_{spo0F}-yfp cells (JL072, bottom). Individual cell cycles are delineated by sequential gray and white shading. Cartoon indicates key regulatory links. (B) Left: illustration of definition of phase. Right: Histogram of cell cycle phases for abrB and spo0F promoter activity pulses. abrB expression typically pulses early each cell cycle, while spo0F expression typically pulses later. (C) Histogram of number of spo0F promoter activity pulses per cell cycle. Half of cells pulse in the first two cell cycles following transfer to resuspension media, while the majority of cells show a single pulse in subsequent cell cycles. The null hypothesis is a binomial distribution with the same mean number of pulses per cell cycle.

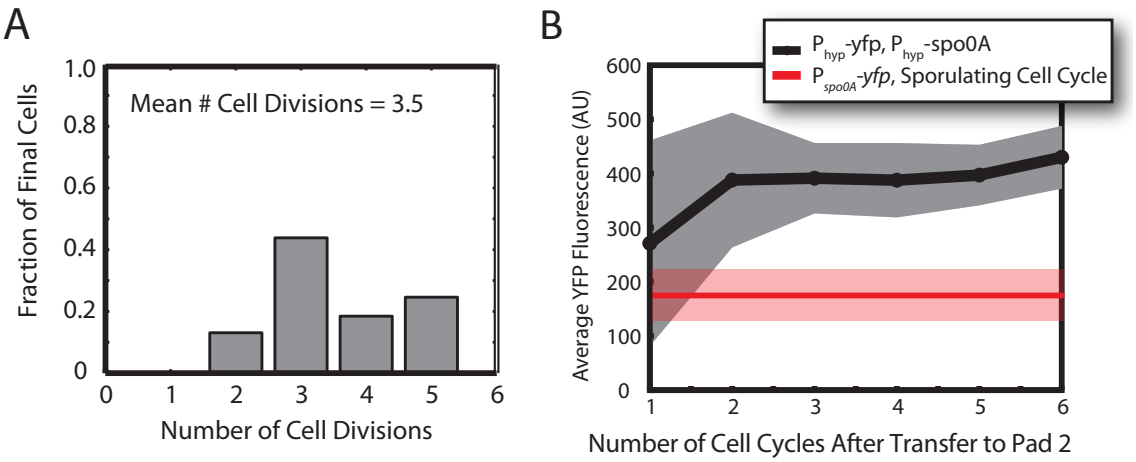


Figure 2.13: **Analysis of pre-growth and Spo0A expression for pad transfer experiment (Figure 2.6).** (A) Distribution of number of cell divisions on Pad 1. The growth of 10 randomly chosen microcolonies were followed on Pad 1 using time lapse microscopy. (B) Induced Spo0A expression on Pad 2 (JL190, blackmean ± SD) rapidly exceeds that from the wild type spo0A promoter (JL251, redmean ± SD). Fluorescence is mean cellular yfp intensity time averaged over the entire cell cycle.

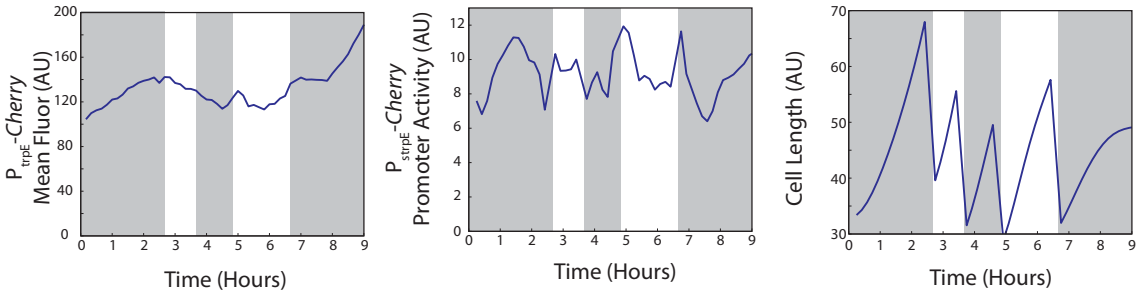


Figure 2.14: **The trpE promoter fluctuates but does not pulse.** Typical time traces of P_{trpE}-mCherry mean fluorescence (left) and promoter activity (center), along with cell length (right) in a typical cell lineage (strain JL024). Promoter activity, while fluctuating, has a lower dynamic range and less temporal structure than $Spo0A^P$ regulated promoters.

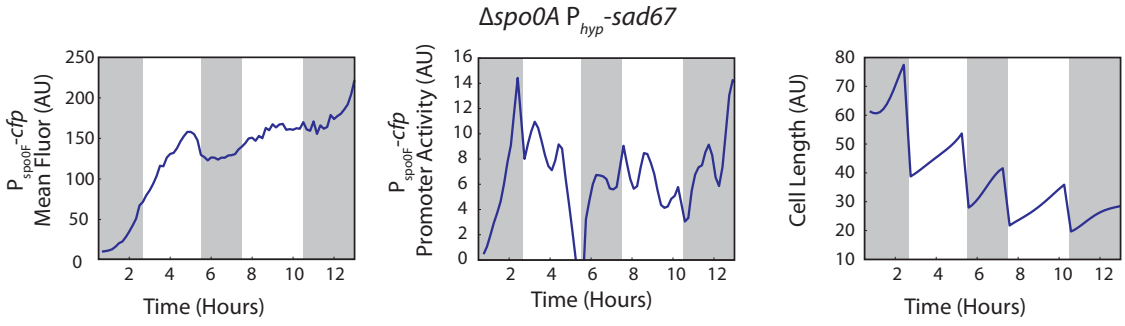


Figure 2.15: **Pulsing is abolished in the constitutively active Spo0A mutant $spo0A^{sad67}$.** Typical time traces of P_{spo0F}-yfp mean fluorescence (left) and promoter activity (center), along with cell length (right) in a typical cell lineage (strain JL065). The promoter activity exhibited fluctuations but lacked the characteristic cell cycle phased pulses present in the wild type.

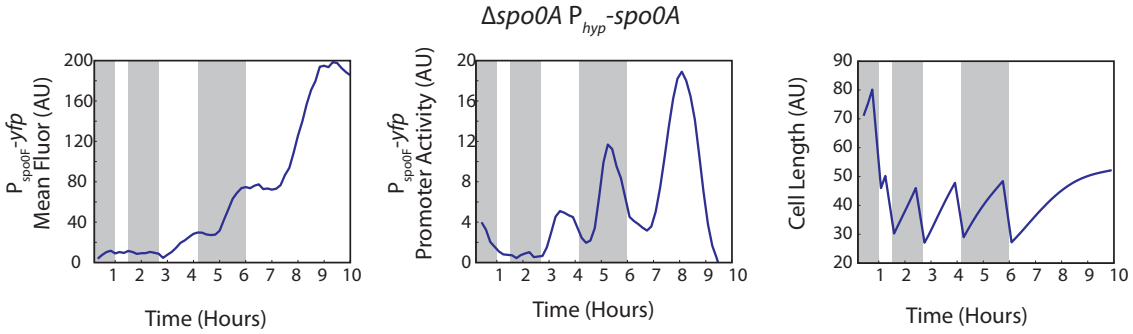


Figure 2.16: **The native spo0A promoter is not required for pulsing.** Typical time traces of P_{spo0F}-yfp mean fluorescence (left) and promoter activity (center), along with cell length (right) in a typical cell lineage of strain JL111 (Δspo0A P_{hyp}erspank-spo0A), showing pulsing similar to that observed in wild type cells.

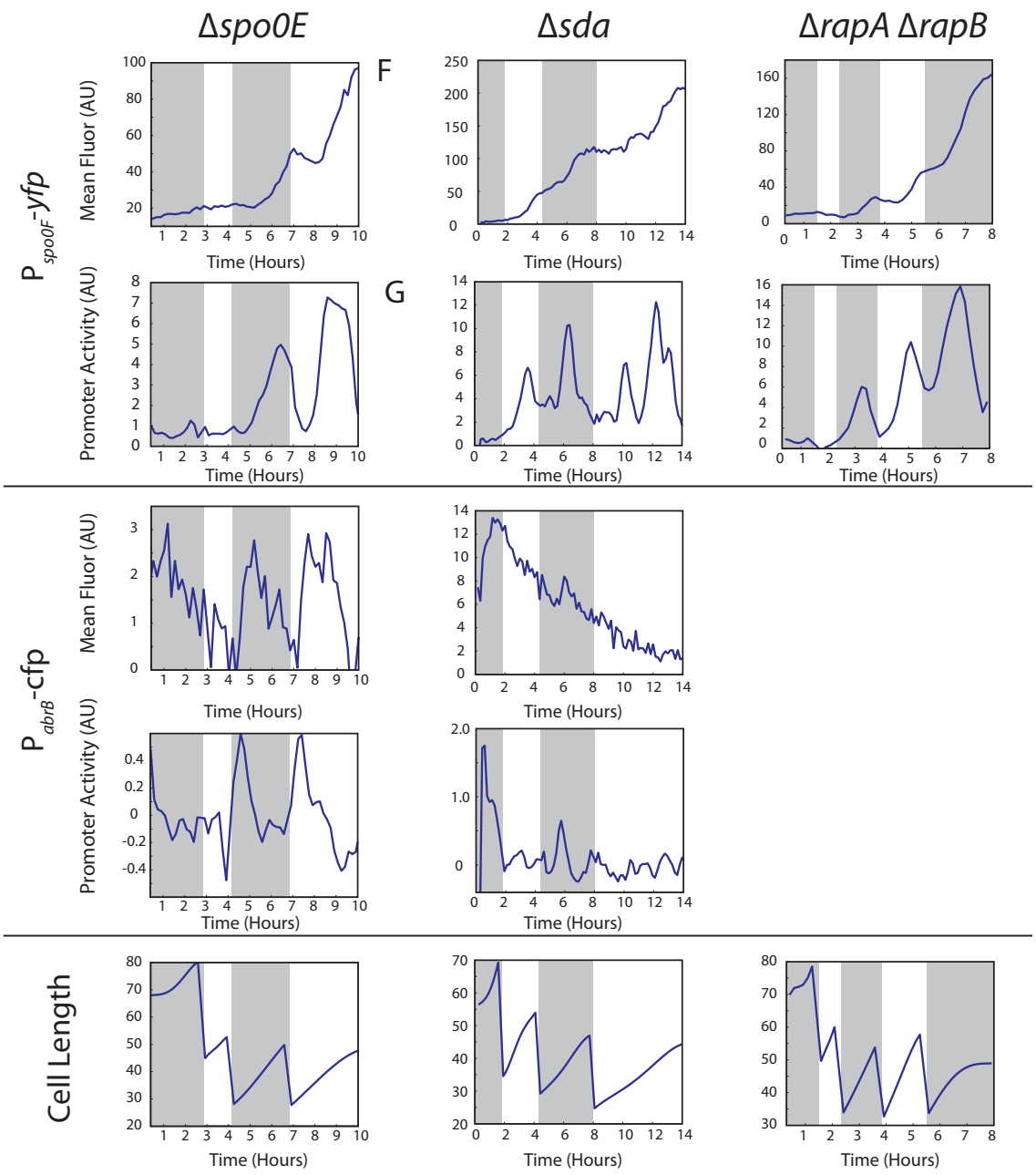


Figure 2.17: **Negative regulators of sporulation initiation are not required for pulsing.** Typical time traces of P_{spo0F}-yfp mean fluorescence (first row) and promoter activity (second row), P_{abrBB}-cfp mean fluorescence (third row) and promoter activity (fourth row), along with cell length (bottom row) in typical cell lineages of Δspo0E (strain JL014), Δsda (strain JL015), and ΔrapA ΔrapB (strain JL160). JL160 lacks the P_{abrBB}-cfp reporter present in the two other starins. Each strain exhibits $Spo0A^P$ activity pulses in the P_{spo0F} promoter similar to those seen in the wild type.

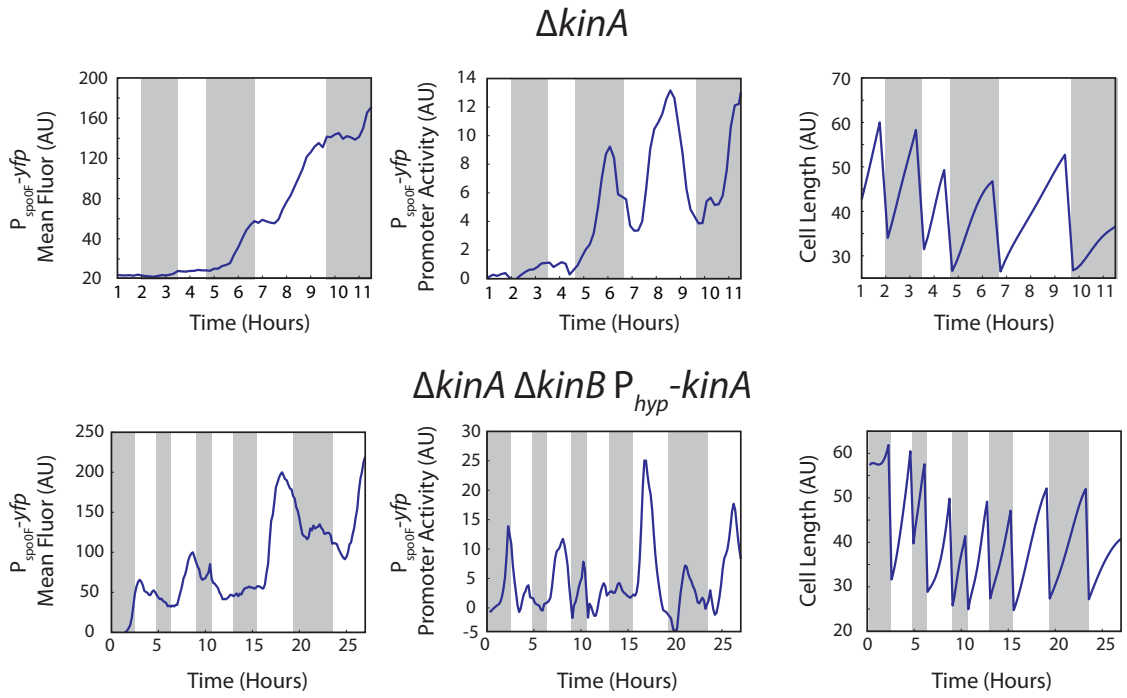


Figure 2.18: **Major sporulation kinases are not individually required for pulsing.** Typical time traces of P_{spo0F}-yfp mean fluorescence (left) and promoter activity (center), along with cell length (right) in typical cell lineages of ΔkinA (strain JL090, top) and ΔkinAΔkinB P_{hyp}erspank-kinA (strain JL144, induced at 2 μM IPTG). Both strains exhibit clear pulsing in P_{spo0F}-yfp promoter activity.

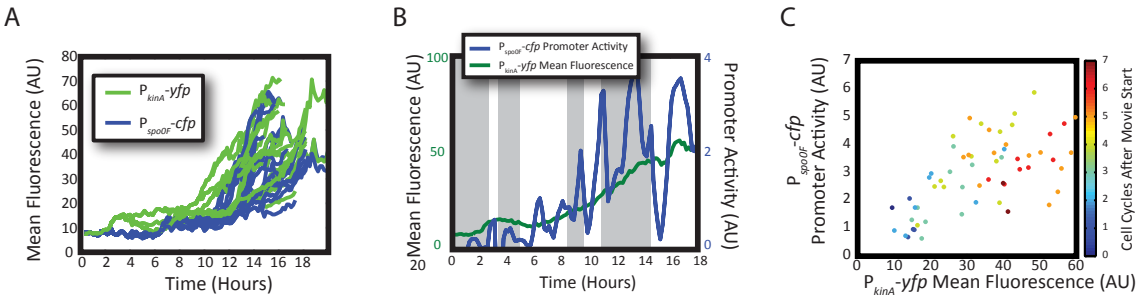


Figure 2.19: **KinA levels increase gradually throughout the deferral period along with Spo0A^{P} pulse amplitudes.** (A) Mean cellular fluorescence traces of cells (strain JL264) expressing P_{kinA}-yfp and P_{spo0F}-cfp. (B) P_{kinA}-yfp mean fluorescence and P_{spo0F}-cfp promoter activity in a single cell lineage. Alternating grey and white shading represents successive cell divisions. (C) P_{kinA}-yfp mean cellular fluorescence correlates with Spo0A^{P} pulse amplitude. Each point represents a single cell's time averaged mean cellular yfp fluorescence (x-axis) and its maximum P_{spo0F}-cfp promoter activity (y-axis). Point color represents that cell's depth in the lineage tree (cell cycles). Correlation coefficient $R = 0.68$.

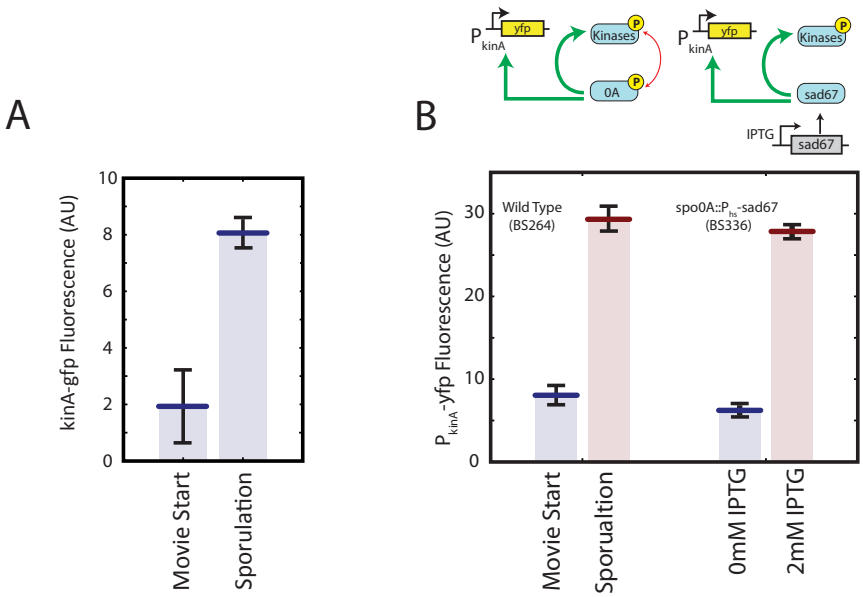


Figure 2.20: **Regulation of kinase production.** (A) kinA-gfp levels increase during sporulation. Mean cellular gfp fluorescence at movie start (left) and at sporulation (right). (B) Comparison of P_{kinA}-yfp expression (mean cellular fluorescence) between cells with wild-type promoter regulation (left, BS264) and cells where P_{kinA} is regulated by inducible spo0A^{sad67} (right, BS336). In wild type cells expression rises monotonically from movie start (blue) until sporulation (red). Fluorescence values in the inducible sad67 cell line were taken after 15 h on the resuspension media pad. Fluorescence in uninduced cells (blue) remained low, while the fluorescence of IPTG induced cells (red) was similar to that of sporulating wild type cells. Bars represent standard error of measurement.

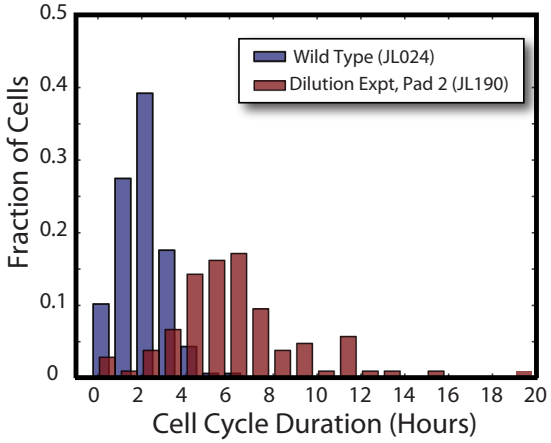


Figure 2.21: **Cell growth rate in pad transfer experiments.** Histograms of cell growth rate, measured in cell cycle duration on pads 1 and 2 (cf. Figure 2EG). Growth on Pad 2 (red, N = 167) is significantly slower than growth on pad 1 (blue, N = 105). Although Pad 2 cells grow significantly slower than wild type cells, they still defer sporulation for the same number of cell cycles (Figure 2G).

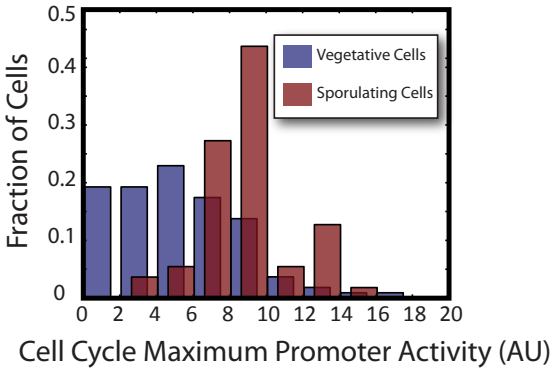


Figure 2.22: **Sporulation occurs past a threshold level of Spo0A$^{\mathbf{P}}$.** Histogram of maximal promoter activities (strain JL024) in our movie conditions. Non-sporulating cells (blue, N = 109) showed systematically lower promoter activity than sporulating cells (red, N = 55), although there is significant overlap. The conditional probability of sporulating given a promoter activity greater than 6 is 70%

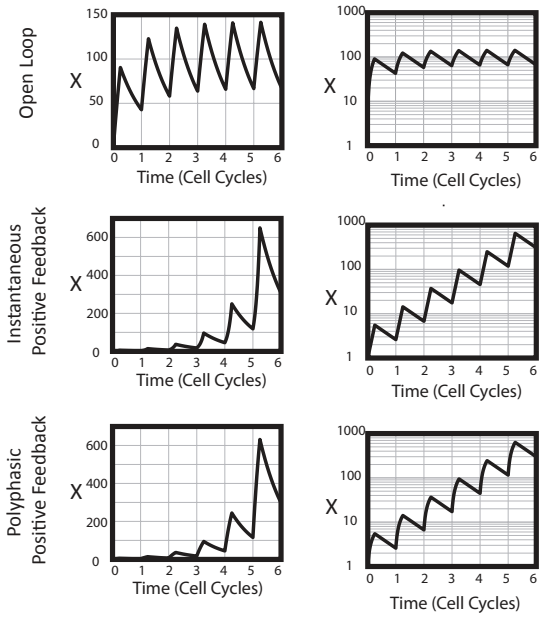


Figure 2.23: **Simplified one-dimensional models capture qualitative circuit deferral dynamics.** One-dimensional models are described in Model Details. Dynamic traces are of models tuned to cross a threshold of $x = 100$, starting from $x = 1$, with a five cell cycle deferral. x is plotted on both linear (left) and logarithmic (right) scales to illustrate exponential behavior.

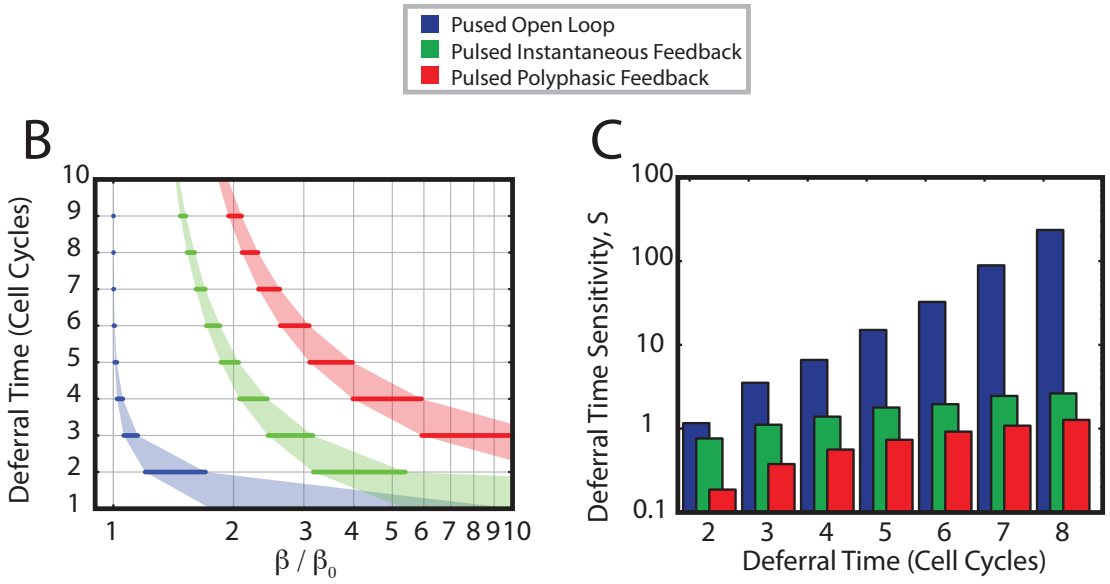


Figure 2.24: Simplified one-dimensional models capture qualitative circuit deferral robustness. One-dimensional models are described in Text S1. (Left) Deferral time dependence on feedback strength for each model. For comparison, β of each model is plotted normalized to the minimal value b0 needed to reach the threshold of X = 100. Open loop: $\beta_0 = 100$; instantaneous: $\beta_0 = 1$; polyphasic: $\beta_0 = e - 1$. (Right) One-dimensional models were compared for their ability to generate multi-cell-cycle deferral times, as with the two-component model of the main text. For each circuit, feedback strength b was tuned to produce different deferral times (x-axis). The sensitivity of deferral time to feedback strength was calculated as in the two-component model. The three circuits differ systematically in both the magnitude and rate of increase of sensitivity with deferral time. The open loop circuit is the most sensitive, followed by the instantaneous feedback, with the polyphasic feedback showing the least sensitivity.

Chapter 3

Bacterial Cybernetics

> From desire ariseth the thought of some means we have seen produce the like of that which we aim at; and from the thought of that, the thought of means to that mean; and so continually, till we come to some beginning within our own power.
>
> -Hobbes, *Leviathan*

3.1 Introduction

In this chapter I describe my initial conceptual, experimental, and theoretical progress towards the design and construction of cybernetic cells. Unlike the previous chapter describing complete published work, this chapter describes in-progress work that I plan to finish here after graduation. My specific definition of a cybernetic cell is a living cell containing a set of genes controlled by programmable computerized rules in real time. One may fantasize about a swarm of computer enhanced cells running amok to wreak havoc on life as we know it. However, the power, purview, and purpose of my cybernetic system will be much more modest. I will use the 'cybernetic cell' set up to attack a simple yet general and challenging problem in biology and bioengineering. How can we engineer a cell to allocate limited resources to grow optimally in an unpredictable feast-or-famine environment? I will argue that this cybernetic cell set up represents a natural and effective method to fully attack this problem in living cells, overcoming limitations of previous techniques.

Cells in the wild are sometimes confronted with situations that force them to choose how to allocate resources in order to survive. For example, given a sugar molecule, a bacterium in the wild

can choose to either burn it for energy or use it to synthesize structural molecules such as proteins. Cells that optimally allocate these scarce resources may gain a distinct growth advantage over less fit competitors. While wild type cells have certainly evolved effective responses to a wide range of challenges, even a modicum of curiosity suffices to make one wonder if there's room for improvement.

Synthetic biology seeks to engineer novel organisms that possess abilities not found in nature. Most engineered organisms are designed to "outperform" natural organisms in some way. For example, the biotechnology industry has engineered microbial strains which can produce valuable chemical products in much higher yields than their wild type counterparts. As synthetic biology advances, we hope to engineer ever more complex phenotypes. As the field is still in its nascent stages, each successful engineering accomplishment provides valuable lessons to help the field grow.

Efficient resource utilization, as described above, stands as a natural challenge. Can we engineer cells whose gene expression enables them to efficiently utilize limited resources in a variety of situations? For example, can we engineer cells to grow optimally in unpredictable, dynamically changing environments that constantly shift between feast and famine? Any cell that would live a life outside the lab would undoubtedly have to overcome such challenges. Engineering such an organism would require that we explore novel dynamic gene expression strategies, and their ability to facilitate growth in correspondingly dynamic environments.

I confront this problem by harnessing the programmability of digital computers to control gene expression 3.1. I am engineering *E. coli* cells where key metabolic genes are placed under optogenetic control, so I can dynamically control their expression using light. These cells are grown in a special microscopy compatible, microfluidic chemostat that enables the continuous observation of individual cell lineages for many days. The cells will be subjected to a time varying environment that switches between different growth medias. A computer will monitor cell state (cell growth rate, and gene expression), and decide based on that information how to control metabolic gene expression.

Computerized gene expression control will enable us to explore novel gene expression strategies flexibly and efficiently. Importantly the computer, properly programmed, will be able to *learn* the cell's optimal control strategy. The ability to learn will likely play an especially critical role in this

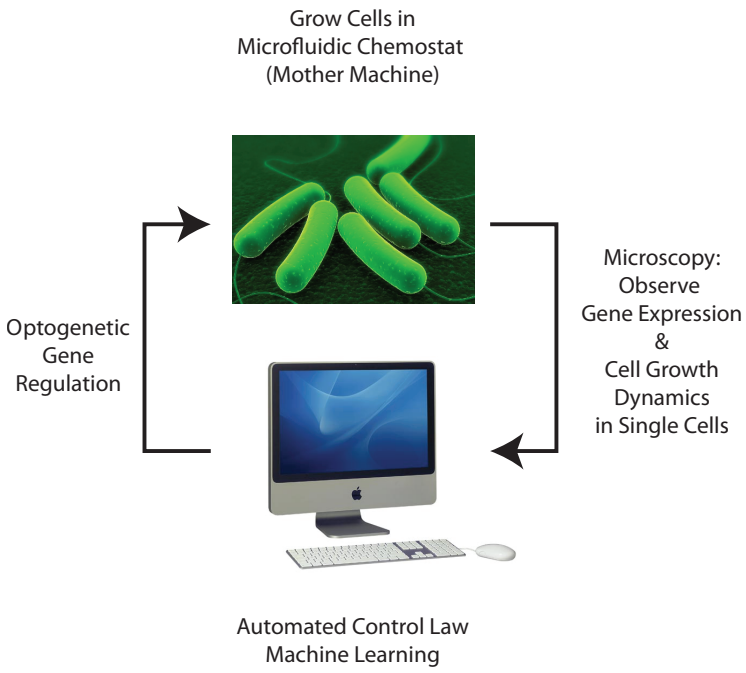


Figure 3.1: **Cybernetic cell set up.** A computerized real time feedback loop enables programming of gene expression dynamics. *E coli* cells (top) are maintained in a chemostatic environment. Single cell time lapse microscopy (right) monitors growth rates relevant gene expression, providing an instantaneous read out of cell state. A computer uses present and past cell state read outs to compute a control law (bottom), which then drives the expression of relevant genes via optogenetics (left).

biological context for two reasons. First, we currently have a very limited understanding of biology. Even in the case of metabolism, where outstanding concentrated experimental efforts have identified most molecular players, the interaction network turns out to be extremely complex (Figure 3.2). This complexity likely warrants the development of automated exploration techniques.

Second, we currently lack the physical technology to control large numbers of nodes in complex genetic or signaling networks. For example, we lack general, high dynamic range gene expression inducers. Most systems biology papers use one or two gene expression inducers and it's shocking to realize that the most popular one, IPTG, has been around for 50 years. Optogenetics techniques are currently even more limited due to optical bandwidth constraints. We are therefore faced with the problem of controlling a complex network at essentially a single node. Again, automated exploratory techniques will probably help.

Why perform these experiments at single cell resolution instead of in batch culture? First,

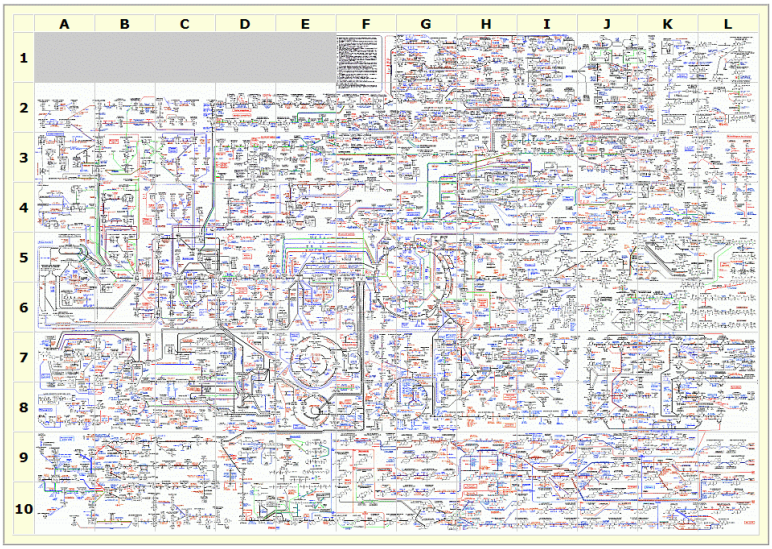


Figure 3.2: **Typical 'wall chart' representation of cellular metabolism.** A typical wall chart representation of the reactions in metabolism. This one, from Roche, is probably the best looking available on the web - unfortunately the upper left company info is always blacked out.

controlling a large number of single cells allows us to perform a large number of computerized experiments at the same time. There is some numerical evidence that sharing information between multiple independent experiments can accelerate a learning process (Kretchmar, 2002). Second, batch culture obscures individual variation in gene expression and the timing of critical events like the cell cycle. These detailed features may be critical for proper control.

In case this problem sounds somewhat abstract (as it does to me), let me provide a concrete analogy. Consider the case of a diabetic who wants to feel live as normal a life as possible while feeling as healthy as possible. To do this, he regularly monitors and records his blood glucose level, along with how he feels. To help himself feel better, he can exercise, change his diet, and take insulin. But how exactly should he regulate his diet, exercise, and insulin to optimize his blood glucose and overall well being? Medical science has only a rough idea of how each of these treatments work. Furthermore, living a normal life places unpredictable demands on his general metabolism. Being a proactive person, our hypothetical patient does not merely try to maintain maintain a constant blood glucose level. Rather, he also looks for patterns in his records for how blood glucose levels make him feel, and correlates those with his diet, exercise, insulin, and activity patterns. He adjusts these to anticipate possible unexpected demands each day. Furthermore, sometimes he experiments

with his treatment regimen, and notes which make him feel better and which make him feel worse.

3.2 Current Progress

This approach requires a number of technical and conceptual tools. First, a specific experimental biological system must be chosen with which to explore this problem. This system must be tractable enough to make progress in, but interesting enough to warrant the work involved. Second, I must construct an controlled environmental system that can maintain cells chemostatically for long periods of time, while allowing the tracking of multiple individual cell lineages. Third, I must implement a system to dynamically control gene expression in single cells. Finally, I must choose a method to enable a computer to explore and evaluate various gene expression dynamics.

In the remainder of this chapter, I sketch my progress in each of these directions.

3.2.1 Biological System: The TCA Cycle and Glyoxylate Shunt

What experimental system can capture tradeoff effects and how they relate to cell growth? I have chosen to control the flow of carbons between the TCA cycle and the glyoxylate shunt in the face of fluctuating sources of carbon. The following paragraphs justify why.

A cell's metabolic network can be thought of as a chemical factory that converts nutrient inputs into energy and material for cell growth. This network is extremely complex, with several specialized sections performing specific chemical transformations. The tricarboxylic acid (TCA) cycle can arguably be considered the "central engine" of cellular metabolism, and is present in all organisms performing aerobic metabolism. To generate energy, the cycle consumes acetyl-CoA derived from nutrient catabolism and oxidizes it through a series of reactions which result in CO2 and water.

Remarkably, the TCA cycle can help build up molecules in addition to breaking them down. The "glyoxylate shunt" pathway works with the TCA cycle to do just this. More specifically, it allows cells to use acetate as a building block for more complex molecules. In the standard TCA scheme, a two-carbon acetyl-CoA molecule enters the cycle as usual and is added to a four-carbon oxaloacetate molecule, eventually forming a six-carbon isocitrate (Figure 3.3). At this point the

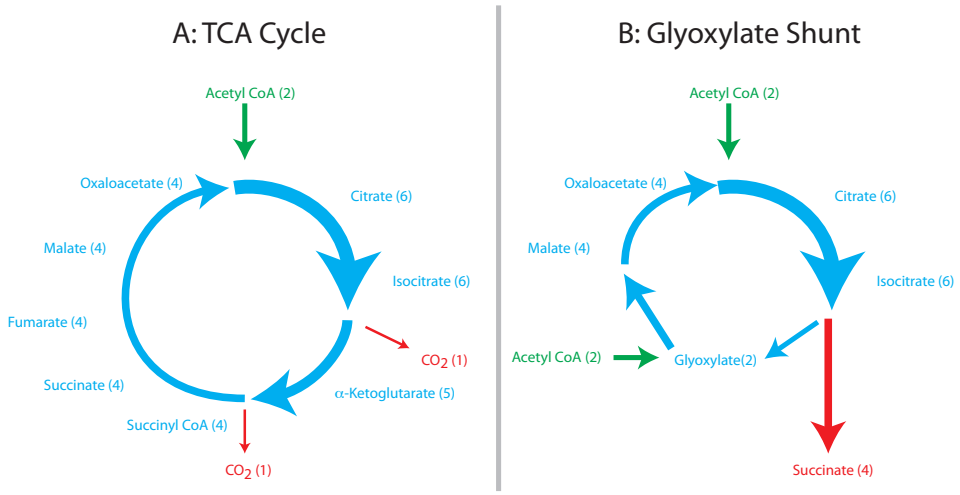


Figure 3.3: **Carbon Flows through TCA Cycle and Glyoxylate Shunt.** Carbon inputs are colored in green, cycle intermediates in blue, and carbon outputs in red. The carbon stoichiometry is represented by a number after each species, and by associated arrow thickness. (A) Carbon flow in the TCA Cycle. Acetyl CoA enters the TCA cycle and is fixed with oxaloacetate. Catabolism and oxidation of subsequent species generates energy for growth. (B) Carbon flow in the Glyoxylate Shunt. Acetyl CoA is fixed with oxaloacetate as in the TCA cycle. The resulting six carbon citrate is used to generate four carbon succinate, which can be rerouted for further molecular synthesis. A second acetyl CoA is incorporated with the leftover glyoxylate, to regenerate oxaloacetate.

cell has a choice to make. Its first choice is to continue to route the isocitrate molecule around the TCA cycle, which results in energy generation and two of isocitrate's six carbons being stripped off, eventually regenerating the original molecule of oxaloacetate. Its other choice is to break isocitrate apart into two molecules: the 4 carbon succinate, and the 2 carbon glyoxylate. If no other molecule were added the cell would be left with a 4 carbon molecule and a 2 carbon molecule, exactly what it started with stoichiometrically. However, the cell then adds another acetyl-coA to glyoxylate, resulting in a second 4 carbon molecule (malate). This malate regenerates the original oxaloacetate we started with. The other four carbon molecule, succinate, can be routed elsewhere for other uses.

The enzymes isocitrate dehydrogenase (*icd*) and isocitrate lyase (*aceA*) control carbon routing through the shunt. Isocitrate dehydrogenase catalyzes the conversion of isocitrate to CO2 and α-ketoglutarate, the next intermediate in the TCA cycle. High *icd* activity thus funnels carbons into the TCA cycle. *icd* activity is regulated by phosphorylation; more specifically, it is inactivated by phosphorylation on its blah blah blah site and active otherwise. This phosphorylation is regulated by a bifunctional histidine kinase *aceK*, which potentially enables a specific form of robust phospho-

rylation control (Shinar et al., 2009). *aceA*, on the other hand, routes carbons through the glyoxylate shunt by breaking isocitrate into succinate and glyoxylate. Succinate is used as a precursor for more complex carbon compounds, while glyoxylate is combined with a second acetyl-CoA to regenerate malate and finally another oxaloacetate. Note that *icd* has been estimated *in vitro* to have more than 100 fold higher affinity for isocitrate than *aceA*. Thus, controlling the activity of *icd*, assuming some background level of *aceA* activity, controls carbon flow between the TCA cycle and the glyoxylate shunt.

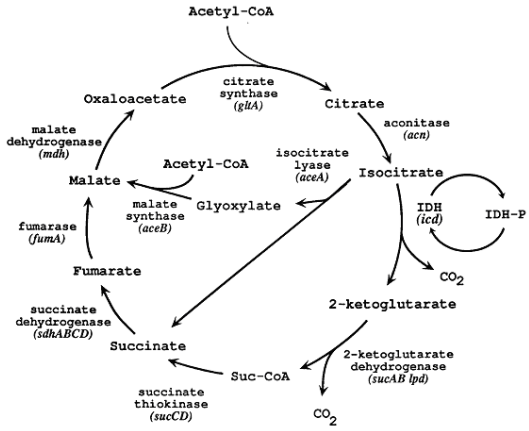


Figure 3.4: **Enzymes and Metabolites of the TCA cycle and Glyoxylate Shunt.**

How do cell growth rates compare in the two conditions? Since we expect cell growth rate to be a main indicator of cellular state, we hope to be able to distinguish growth rates between the two states. We assayed growth rates of wild type MG1655 at 37C *E. coli* in M9 minimal media supplemented with either glucose or acetate as the carbon source. Growth rates in the two conditions were clearly distinguishable using a Victor Wallac multi-well plate reader (Figure 3.5A). Cells in glucose grew with a doubling time of approximately 20-30 minutes, while cells in acetate displayed a doubling time of approximately 2 hours. Single cell glucose growth rates in the mother machine[1] match those in the plate reader; while single cell acetate growth rates are not yet available, I believe they will also match well. Glucose growth rates show significant variation, possibly to slight misfocusing, which can be approximated by additive noise to cell length. Despite this noise, we should be able

[1]I am in the process of developing a fully automated segmentation platform. Data shown are from Suckjoon Jun: http://www.sysbio.harvard.edu/csb/jun/mothermachine.html.

to reliably discriminate between the two different growth rates at the single cell level.

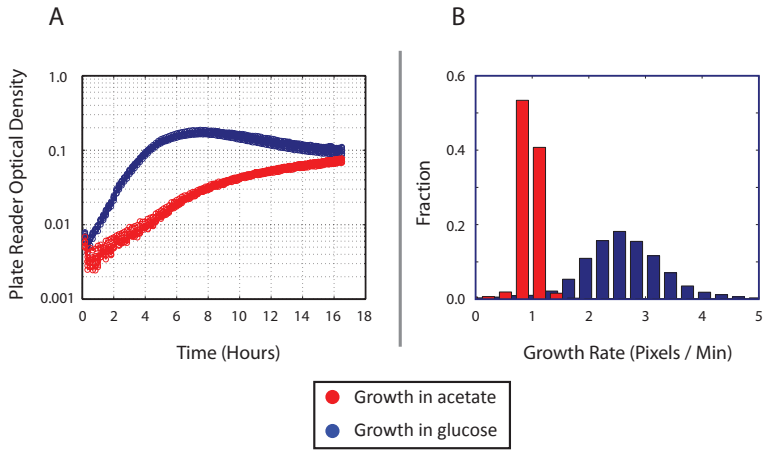


Figure 3.5: ***E coli* growth rates in glucose and acetate.** (A) MG1655 growth in a multi-well plate reader. 8 replicate traces are shown for each condition. Saturation in glucose occurs in at a plate reader optical density reading of about 0.2. (B) Instantaneous single cell MG1655 growth rates in glucose, measured using the mother machine at 1 minute time intervals. Cells in glucose typically divide after an increase in length of approximately 50 pixel units. Glucose histogram is the result of 85,374 data points. Instantaneous growth rate is calculated by temporal difference, and each point is averaged with 5 surrounding points using a nonlinear median filter. Note: acetate growth traces are not yet available; the distribution shown is simply the glucose distribution divided by the ratio of their bulk growth rates.

My set up will therefore dynamically control *icd* expression in the presence of environmental switching between glucose and acetate. More specifically, I will control the expression of a constitutively active *icd* variant, *icd** (Chen et al., 1996). Recall that *icd* activity routes carbons through the TCA cycle, and its high affinity for isocitrate overrides *aceA*'s direction of carbon to the glyoxylate shunt. But can we say more about how *icd* activity affects growth? For example, is there an optimal *icd* activity level existed in each media condition? In an environment with moderate amounts of glucose as the carbon source, the TCA cycle should generate some of the energy necessary for cell growth[2], so higher levels of TCA activity should lead to higher growth.

It turns out that a minimum of *icd* activity is necessary for growth in both glucose and acetate environments. A Δicd mutant[3] is unable to grow in either condition - evidently cells require certain metabolic intermediates for growth. To further isolate which intermediates are necessary, I tested

[2]In environments with plentiful glucose, most energy is generated through glycolysis. I have not yet determined at which glucose concentration the TCA cycle becomes the dominant energy source.

[3]All deletion mutants are from the Keio collection, and were obtained from the Coli Genetic Stock Center at Yale http://cgsc.biology.yale.edu/

the ability of all TCA cycle mutants to grow in glucose or acetate (Table 3.2.1). In glucose, cells apparently need to generate α-ketoglutarate but nothing else. However, in acetate, all genes are required except those converting succinate to malate, which is supplied by the shunt pathway. Thus, it appears that some level of *icd* activity is required in both media conditions to generate specific metabolic intermediates. I can test this hypothesis further by supplementing the growth media with these intermediates, individually and in combination.

Gene	**Essential in Glucose?**	**Essential in Acetate?**
icd	Yes	Yes
sdhA	No	Yes
sdhC	No	No
sdhD	No	Yes
sucA	Yes	Yes
sucB	No	Yes
sucC	No	Yes
sucD	No	Yes
fumA	No	No
mdh	No	Yes
gltA	Yes	Yes

Table 3.1: **Necessity of TCA Cycle Genes in Glucose and Acetate.** Cells with the indicated deletion (Keio collection) were grown overnight in an LB, washed 2X with PBS, and then transferred to both M9 glucose and M9 acetate at 37C. Cultures were checked for any growth after 36 hours.

In addition to the above results demonstrating a minimal necessary level of *icd* activity for growth, I have some preliminary evidence that excess *icd* activity may penalize cell growth, at least in acetate. Intuitively it seems reasonable that excess *icd* will syphon carbon away from the glyoxylate shunt. Since over expression constructs are not yet available, I compared the growth of wild type MG1655 and a $\Delta aceK$ strain in acetate. This mutant lacks the kinase that can phosphorylate and inactivate *icd*, and presumably has higher levels of *icd* activity. Acetate growth curves (Figure 3.6) show that this mutant indeed grows significantly more slowly than wild type. Taken in total, these results suggest that cells must maintain an optimal level *icd* for optimal growth.

3.2.2 Chemostatic Environment Control

For my single cell chemostat, I use the microfluidic "Mother Machine," originally designed by Suckjoon Jun at Harvard (Wang et al., 2010). The mother machine uses a clever layout to enable one to

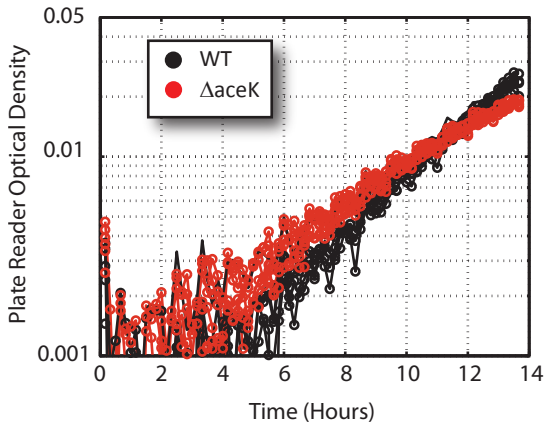


Figure 3.6: **Eliminating Phosphocontrol of *icd* Impairs Growth in Acetate.** Steady state plate reader growth curves for wild type (black) and $\Delta aceK$ (red) strains in M9 acetate. The mutant strains grow significantly more slowly than wild type.

observe individual bacterial lineages for extremely long times (e.g. hundreds of generations of rapidly growing bacteria in rich media). Bacterial cells in the mother machine grow single file in narrow, half open growth channels. Each mother machine device contains thousands of growth channels. As the lineage at the closed end of each channel divides, it pushes its progeny out via the open end of the channel. The channel opens into a large trench through which media flows. This trench both supplies media to the growth channels, and sweeps away any cells which are pushed out of the open end. This selects a single lineage out of an exponentially large family tree, and enables it to be observed effectively indefinitely. My personal record for running the mother machine is 5 days under continuous rapid growth conditions, at which point I ran out of patience and stopped.

Construction of the mother machine involves two parts. First I fabricated the mother machine using standard micro fabrication processes, in Caltech's Kavli Nanoscience Institute clean room. Briefly, SU8 photoresist is deposited on a silicon wafer and patterned to form a mold. Certain aspects of this fabrication process, particularly fabrication of the growth channels, are quite delicate and require significant tuning. These molds are then used to stamp devices out of PDMS. PDMS devices are subsequently bonded to a glass slide using plasma cleaning, and cells loaded into the channels for imaging using either evaporation or centrifugation[4]. A sample image is shown in 3.8.

[4]Detailed instructions on mother machine fabrication, along with a guide I wrote for assembly, can be found at http://www.sysbio.harvard.edu/csb/jun/mothermachine.html.

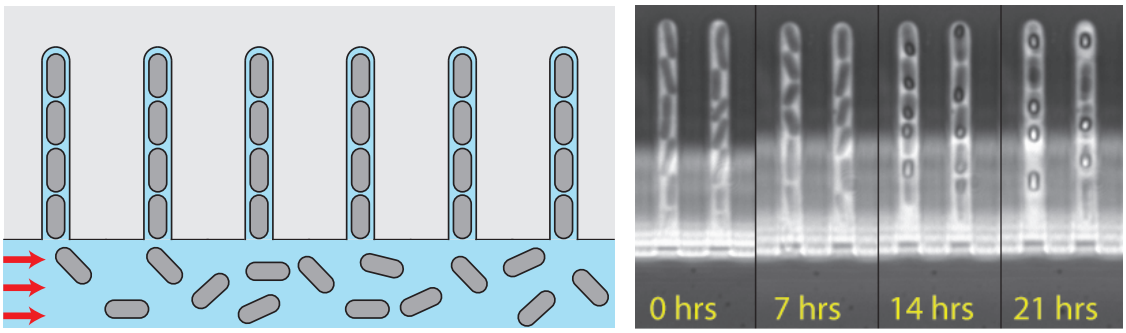


Figure 3.7: **Overview of mother machine.** (Left) Schematic. Bacterial cells grow in half open growth channels, the open ends of which abut a main trench. Fresh media flows through the main trench (red arrows), simultaneously perfusing the growth channels and sweeping away excess cells. (Right) 100X Phase contrast image of sporulating *Bacillus subtilis* cells in the mother machines, demonstrating that cells can respond to media switches. Cells were induced to sporulate at indicated T=0 hours. Phase contrast blurriness at the end of the growth channels is due to the main trench, and does not appear in fluorescence images (not shown).

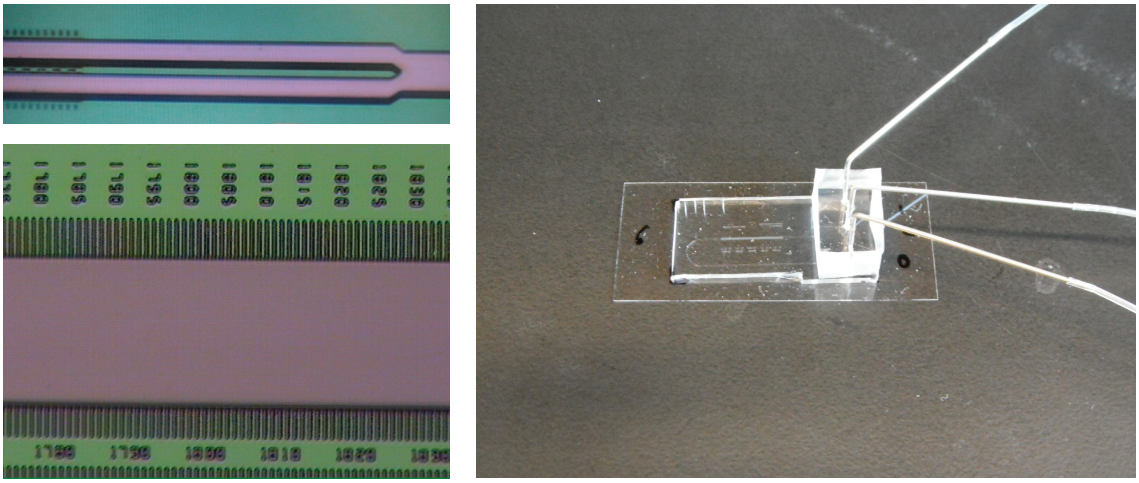


Figure 3.8: **Mother machine: SU8 mold and PDMS device.** (Left, top) Main trench and growth channels. The main trench divides into two branches each with its own sets of growth channels. Note the difference in scale between the two features. (Left, bottom) Close-up of main trench and growth channels. (Right) A dual-inlet / single-outlet mother machine device is bonded to a glass coverslip. The inlets (right, top two) and outlet (right, bottom) are connected to tubing and the device is ready to be mounted on a microscope. Note the growth channels visible as a slight thickening of the far arm of the device.

Once assembled, pumps with appropriate media are connected to the device inlets. Cells are typically equilibrated in chemostatic media conditions for 10 cell cycles to assure they are in steady state before running any experiments.

3.2.3 Optogenetics

Here I review the merits of several optogenetic transcriptional control systems.

The YF1/FixJ system is a two-component signaling based system that has been demonstrated to control gene expression in *E coli* (Moglich et al., 2009). Briefly, the N-terminal sensing portion of the bifunctional histidine kinase YF1 contains LOV ('light oxygen voltage') domain. This makes kinase signaling responsive to blue light. More specifically, blue light inactivates the kinase activity of YF1 on a time scale of minutes and remarkably converts it into a phosphatase. YF1 then phosphorylates the response regulator FixJ, which can then activate transcription from appropriate target promoters. Once blue light is removed, YF1 decays to a dark state with a time constant of roughly 3 hours (Moglich et al., 2009). The YF1/FixJ system has the advantage that it has been demonstrated in *E coli*, using only native chromophores. However, the system's kinetics are very slow, and it uses up valuable 'spectral real estate' (Figure 3.9, blue). More specifically, YF1 is sensitive to blue light up to 500 nm[5]. This prevents us from using any CFP family reporters, leaving only YFP and mCherry[6]. I am currently investigating these tradeoffs experimentally.

An alternative optogenetic system, the Phy/PIF system, consists of a pair of small protein tags (Levskaya et al., 2009). Two different wave lengths of red light stably and reversibly switch these tags between a high affinity state, where they bind to one another, and a low affinity state where they don't. This system has been successfully applied to the local control of cell morphology, e.g. localized cell protrusion, by directing specific signaling complexes to local membrane patches. One could imagine adapting this system for use in *E. coli* by tagging a transcriptional repressor such as lacI with one protein, and tagging the other to a membrane localized protein. In one state, all lacI would be bound to the cell poles. In the other state, lacI would be free to bind DNA and repress transcription from a lacI optimized promoter. A number of key advantages to this system stand out. First, the response times are extremely rapid, being limited mainly by protein diffusion and

[5]I have experimented extensively with a variant of YF1, a light sensitive *B subtilis* sporulation histidine kinase YKA, made by replacing KinA's sensing domain with an LOV domain similar to that of YF1. This system can induce sporulation in a light dependent fashion (Figure 3.7, right). However, it is extremely sensitive to blue light and is modulated even at 500nm (data not shown).

[6]Which itself has slow maturation kinetics.

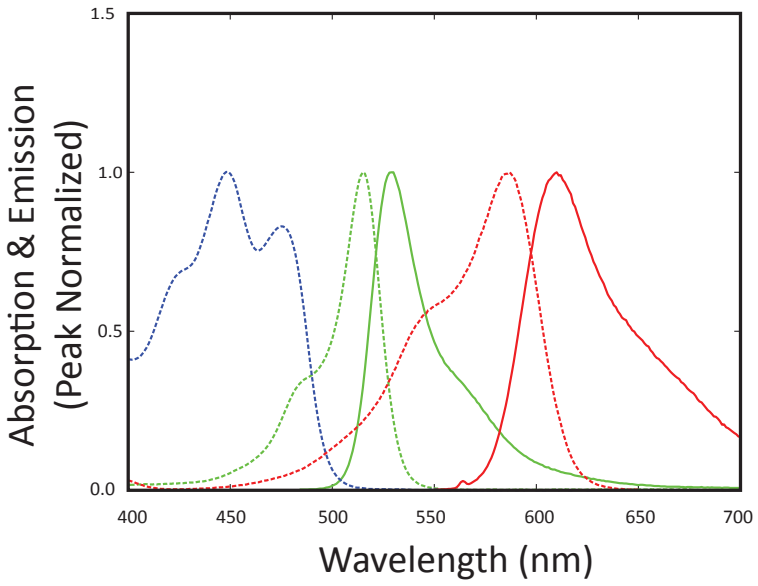


Figure 3.9: **YF1 and fluorescent reporter spectra.** YF1 LOV domain absorption (blue, dashed). Citrine excitation (green, dashed). Citrine emission (green, solid). mCherry excitation (red, dashed). mCherry emission (red, solid)

binding kinetics. Second, the system is controlled by red light, which is significantly less photo toxic to the cell. These advantages are balanced by several limitations, however. First, the system has never been demonstrated to work in *E. coli*, although it has been utilized in the yeast *S. cerevisiae*. The protein binding kinetics may be significantly altered in different organisms. Further, the system requires an exogenous chromophore whose function may be species dependent. Second, one wonders whether pole localization is sufficient to sequester protein away from the *E. coli* nucleoid. Despite these limitations, however, the system is certainly worth exploring and I plan to test it as quickly as possible.

3.2.4 Learning Algorithms

How can we enable a computer to learn optimal gene expression dynamics? We need a representation that can handle a wide range of dynamics, either deterministic or stochastic. The representation must also allow efficient exploration and analysis by computer. Several candidate methods suggest themselves. Possibly the silliest would be to represent expand a desired set of dynamics in some set of basis functions, such as a Fourier series. This representation is probably not effective enough for our

purposes: representing a dynamic trace accurately in Fourier space may require many coefficients, and anyways the representation is conceptually unnatural and unintuitive. Another option would be to search through the space of dynamics using genetic algorithms, but one wonders how many trials such a procedure would take to converge, especially since each trial is potentially a multi-day experimental run as opposed to a microsecond of computer time.

3.2.5 Learning with Markov Decision Processes

The Markov Decision Process framework satisfies the criteria listed above. Formally, a Markov Decision Process (MDP) is a discrete time stochastic process defined by 4 objects (Sutton and Barto, 1998):

- S: A finite set of *states* the process can take on.
- A: A finite set of *actions* the process can perform in each state.
- $P^a_{ss'}$: The *probability* that taking action a in state s causes a transition to state s' at the following time step.
- $R^a_{ss'}$: The *reward* one receives after taking action a in state s leads to a transition to s'.

The states and transition probabilities should feel familiar to those acquainted with standard Markov processes. The actions and rewards allow the MDP to be optimized. Specifically, they allow us to pick actions for each state that maximize the amount of reward we expect to get. Let a *policy* $\pi(s, a)$ be a probability distribution on all $a \in A$ associated with each state $s \in S$. Once we specify a policy, we can assigns to each state s a *value* $V^\pi(s)$, which is the total expected reward one obtains by starting in s and following policy π[7].

$$V^\pi(s_t) = \left\langle \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right\rangle_\pi$$

[7]In operations research, value functions are known as the "cost to go." In that context, actions are typically thought of as incurring a cost, for example the cost of moving supplies from one depot to another. Reinforcement learning, with roots in behavioral psychology, thinks in terms of learning accompanied by rewards.

The *discount* constant $0 \leq \gamma \leq 1$ functions like an interest rate to quantify how much we prefer quick rewards over delayed rewards. Sometimes alternative definitions of reward are more appropriate. For example, one can maximize the average reward $\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} r_{t+k+1}$.

I can adapt the MDP framework to my system straightforwardly, by discretization. For example, I can define a cell's state as its current growth rate and its current expression of any relevant genes. Possible actions are different optogenetic activations. Rewards would be some monotonically increasing function of cell growth rate. The transition probabilities will be unknown, but we will see that machine learning frameworks can account for this possibility.

It turns out several options are available to find an optimal policy in a given MDP.

3.2.5.1 Solution via Dynamic Programming

If one has complete knowledge of the MDP, one can use dynamic programming to solve for an optimal policy. Dynamic programming takes advantage of the value function's recursive structure to find an efficient solution. To see the recursion explicitly, rewrite the definition of the state value in the following way:

$$\begin{aligned}
V^{\pi}(s_t) &= \left\langle \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right\rangle_{\pi} \\
&= \left\langle r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_t = s \right\rangle_{\pi} \\
&= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} \left(R^a_{ss'} + \gamma \left\langle \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_{t+1} = s' \right\rangle_{\pi} \right) \\
&= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} \left(R^a_{ss'} + \gamma V^{\pi}(s') \right)
\end{aligned}$$

This is known as a Bellman Equation, and it contains a significant amount of information. For example, to find the true value that a policy defines for each state, you can simply plug the MDP with

that policy into the Bellman Equation and iterate. More specifically, for any initial $V(s), \forall s \in S$, the sequence V_k defined below converges to $V^\pi(s)$, e.g. $\lim_{k\to\infty} V_k(s) \to V^\pi(s)$

$$V_{k+1}(s) = \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} \left(R_{ss'} + \gamma V_k(s')\right)$$

Furthermore, this type of value updating can also be used to find the optimal policy itself.

3.2.5.2 Solution via Reinforcement Learning

If one lacks complete knowledge of the MDP, one can use a set of techniques known as reinforcement learning to find a value function $Q(s,a)$ for each state-action pair. Knowing these values guides one's choice of policy. It is instructive to analyze a sample method, known as known as Q-learning. This method starts with a random set of action value functions, and essentially explores the Markov Chain in a Monte Carlo fashion by probabilistically choosing an action in each state to maximize its expected reward based on current value estimates. It then updates the value estimates with the reward it receives.

More specifically, the algorithm starts in some state s with a set of randomly initialized value functions $Q(s,a)$. It then chooses an action based on $Q(s,a)$ in an *ϵ-greedy*[8] fashion. Once that action a is chosen, it observes the reward r it receives and updates the state/action value function as follows:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right]$$

Here, α is a parameter governing the learning rate, and γ is again the discount. Intuitively, as $Q(s,a)$ updates it should approach an estimate of the total discounted reward it will receive after choosing that action.

To begin to test the utility of reinforcement learning in our setting, I ran a very preliminary simulation of a simple toy problem. More specifically, I considered an idealized system consisting of a cell and its environment. The environment switches sinusoidally between values of 0 and 1

[8] Pick the apparent best action with probability 1-ϵ, and a uniformly random action otherwise, for ϵ small.

with constant period. The cell contains a protein p, and grows at a rate γ which depends on p. A control signal c, representing optogenetic control, controls protein expression. At fixed time intervals corresponding to microscope sampling times, a learning algorithm reads the current protein concentration $p(t)$ and the current cell length $l(t)$. It then calculates the growth rate estimate $\tilde{\gamma} = \frac{1}{l}\frac{dl}{dt}$, and uses $p(t)$ and $\tilde{\gamma}$ as state inputs to the learning algorithm. Possible actions are the control amplitudes c. In this initial simulation I assume protein concentration can change rapidly, as in the Phy/PIF system, with a equilibration rate $\gamma_0 >> \gamma$. Importantly $\gamma = \gamma(p, e)$. In this simulation, $\gamma = 0.2\left((2-p)(2-e) + pe\right)$. In other words, cells grow at a high rate if and only if protein expression matches the environment. The dynamics are thus:

$$\frac{dp}{dt} = -(\gamma + \gamma_0)p + c(t)$$

$$\frac{dl}{dt} = \gamma l$$

These preliminary results (shown in Figure 3.10) indicate that reinforcement learning can be used to automatically optimize gene expression strategies, without a detailed model of internal dynamics[9]. More specifically, learning algorithm eventually develops an action-value representation to maximize cell growth rate in the presence of an unknown constraint. As the learning progresses, the system gradually begins to match protein expression to environment to maximize growth, using only the observables of growth rate and protein level.

Several features of this initial effort could use improvement. First, it must obviously be extended to larger state spaces with more complex structure. Second, more work needs to be done to understand how learning depends on the parameters α, γ, and ϵ (the ϵ-greedy parameter) - for example, if they should be tuned adaptively. Third, I need to optimize learning *speed*, which while occasionally quick to converge, often converges more slowly as shown in Figure 3.10. However, numerical evidence from other workers has demonstrated that parallelized reinforcement learning - e.g. a situation where multiple identical learners performing the same task can share information,

[9]It is known that reinforcement learning converges more quickly if some structure is built into the problem's representation, beyond a simple action value table. We will explore these techniques as the project develops.

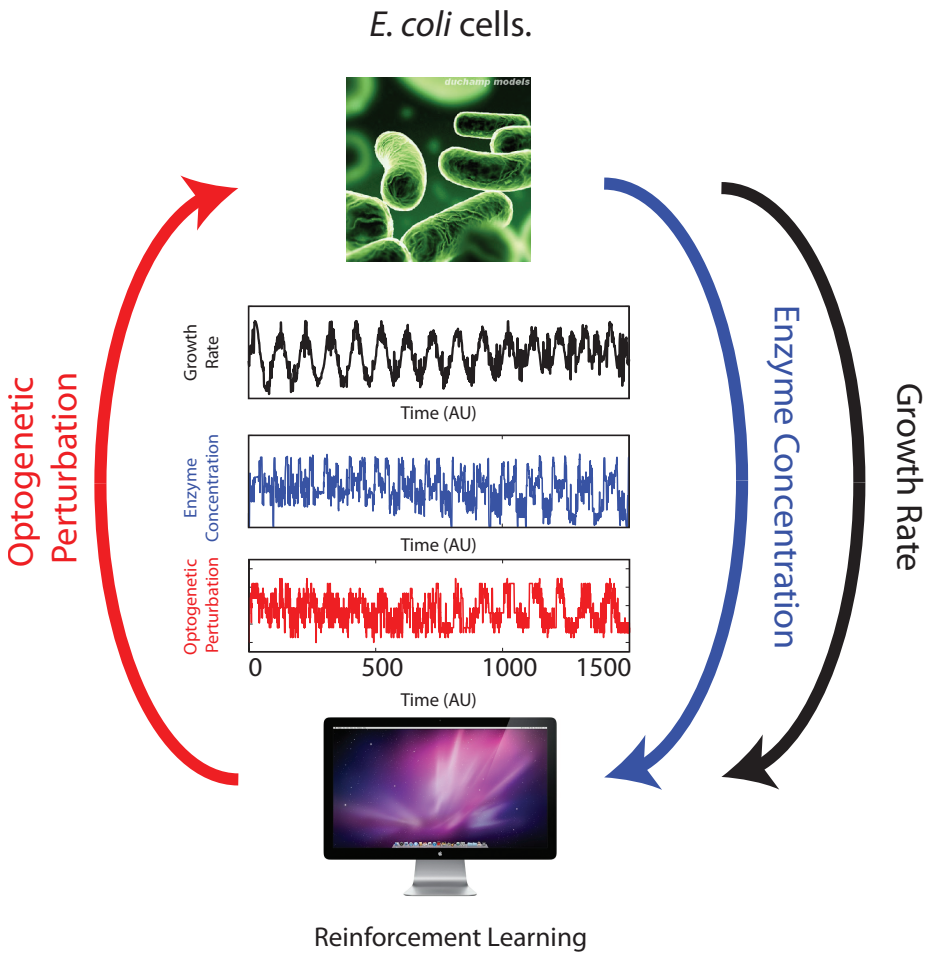


Figure 3.10: **Machine learning can find improved dynamic gene expression strategies.** Simulated *E coli* cells are controlled using a simple reinforcement learning algorithm. The algorithm observes current cell growth rate (black) and the concentration of a single enzyme (blue), and regulates simulated enzyme production via signal representing optogenetic transcriptional control (red). The learning algorithm gradually adapts its control strategy to generate an appropriate expression signal given current gene expression and growth state. In this case, in response to a hidden oscillatory environment, the control signal and resulting enzyme concentration gradually learn to compensate. After learning, the control system significantly smooths out growth rate fluctuations.

offers significant speed ups (Kretchmar, 2002). Parallel channels in the mother machine can provide exactly this large number of identical learners, which can share information.

Together, the results in this section provide promising preliminary evidence that we will be able to implement our cybernetic cell. I am looking forward to making these plans a reality.

Chapter 4

Conclusion

> Toad talked big about all he was going to do in the days to come, while stars grew fuller and larger all around them, and a yellow moon, appearing suddenly and silently from nowhere in particular, came to keep them company and listen to their talk.
> -Kenneth Grahame, *The Wind in the Willows*, Ch. 2

4.1 Summary

In this thesis I described two projects studying the relation between gene expression dynamics and cellular function.

The first project discovered that a gene regulatory network can be used by a cell as a 'count-up' timer. Pulsing dynamics allow the cell to set its timer's duration more accurately. Specifically, pulsed phosphorylation of the sporulation master regulator Spo0A enables *Bacillus subtilis* to defer sporulation for multiple cell cycles, using a polyphasic positive feedback loop. Polyphasic feedback makes sporulation deferral significantly more robust to variations in positive feedback strength.

The second project, currently in progress, is exploring 'cybernetic cells' - cells where real-time computerized feedback dynamically controls the expression of selected genes. This computerized control, implemented via a combination of optogenetics and time lapse microscopy, will enable the rapid exploration and evaluation of different dynamic gene expression strategies. Specifically, I will determine how cells should optimally allocate metabolic carbon flow in an unpredictably dynamic 'feast-or-famine' environment. Machine learning techniques will be used in living cells to

automatically explore and evaluate various dynamic strategies, with the potential to uncover novel biological constraints along the way.

4.2 Future Work

I leave the reader with several additional questions posed by my work. These questions are really of unknown difficulty, so caveat emptor...

For Chapter 2:

- *What mechanism drives Spo0A^P pulsing?* My work eliminated the obvious gene candidates. The mechanism could therefore be due to a more complex multi-gene effect, or possibly due to currently unannotated genes.

- *Can we devise high throughput methods to screen and select for novel gene expression dynamics?* Finding the answer to the first question may be hard without screening many genes. Can we develop a high throughput method to screen gene expression dynamics, analogous to the enormously successful plate based methods of classical genetics? One approach could use the mother machine (see e.g Figure 3.7). One would construct a reporter to visualize the gene expression of interest in single cells, say with a fluorescent promoter fusion. One could then randomly mutagenize the chromosomes of a population of these cells and load them into the mother machine. Each growth channel would contain a pure genotype, and the dynamic behavior of each could be observed over a long time using movies. Microfluidic techniques could then be used to select cells from channels displaying the desired new gene expression pattern (e.g. no more pulsing, faster oscillations, state switching at various frequencies, etc). This combination of screening and selection would greatly increase our ability to explore the mechanisms behind dynamic gene expression.

- *Are there any functions for polyphasic* negative *feedback?* My work identified polyphasic positive feedback as a potential mechanism to enable deferred differentiation. What advantages does the polyphasic mechanism provide for negative feedback? Some interesting examples

exist in the electronic circuit literature (Brennecke and Lindemann, 1974), but whether any examples will be found in cellular regulatory networks remains unclear.

- *More generalized pulsing dynamics.* Polyphasic feedback sits squarely in the 'variable amplitude, fixed timing' quadrant of the pulse space described in Figure 1.1. In that discussion, pulsing where both timing and amplitude varied was mentioned as unexplored. Can one theoretically explore possible mechanisms and functions? Will they exist in real biological system and demonstrated experimentally?

- *What advantages and disadvantages generally do different deferral mechanisms have?* My work distinguished between three possible mechanisms for sporulation deferral: external (based e.g. on quorum sensing or environmental change), cell autonomous dilution of an internal factor, or cell autonomous circuit dynamics. As more specific examples of timer phenomena are discovered, what will they teach us about the differences between these three mechanism classes?

For Chapter 3:

- *How can parallel agents accelerate reinforcement learning?* Our experimental set up with the mother machine contains multiple identical, independently controlled cells trying to learn an optimal response to the same environment. How should these cells share information to accelerate the learning process? Numerical experiments with N-armed bandits (Kretchmar, 2002) suggest the speed up could be considerable. Can we demonstrate, quantify, and prove any comparable speed ups for our problem, or for reinforcement learning more generally? These results would be of immediate practical utility for this project.

- *Can the dynamic gene expression strategies we obtain be 'compiled' into corresponding genetic circuits?* One can view this whole study as a type of CAD or *in-situ* prototyping of gene expression dynamics. Down the line, one may wish to actually implement these types of strategies cell autonomously without computer control. Yet the strategies are couched in the

language of Markov Decision Processes or their like. How can these be closely approximated by a feasible genetic circuit?

- *How can automated experimentation be fruitfully applied to other areas of biology, or beyond?* In the introduction, I gave several examples of how automated knowledge collection has aided fields such as biology, physics, and mathematics. In what other ways can we automate discovery? How broad will these applications be? Will they remain academic curiosities, or can this general methodology be of broader practical use?

Bibliography

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

M. Acar, J. T. Mettetal, and A. van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*, 40(4):471–5, 2008.

Uri Alon. *An introduction to systems biology : design principles of biological circuits*. Chapman & Hall/CRC mathematical and computational biology series. Chapman & Hall/CRC, Boca Raton, FL, 2007.

K. Asai, F. Kawamura, H. Yoshikawa, and H. Takahashi. Expression of kina and accumulation of sigma h at the onset of sporulation in bacillus subtilis. *J Bacteriol*, 177(22):6679–83, 1995.

Richard Ernest Bellman and Rand Corporation. *Dynamic programming*. Princeton University Press, Princeton,, 1957.

B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat Chem Biol*, 5(8):593–9, 2009.

T. H. Bird, J. K. Grimsley, J. A. Hoch, and G. B. Spiegelman. Phosphorylation of spo0a activates its stimulation of in vitro transcription from the bacillus subtilis spoiig operon. *Mol Microbiol*, 9 (4):741–9, 1993.

R. Brennecke and B. Lindemann. Theory of a membrane-voltage clamp with discontinuous feedback through a pulsed current clamp. *Review of Scientific Instruments*, 45(2):184–188, 1974.

D. Burbulys, K. A. Trach, and J. A. Hoch. Initiation of sporulation in b. subtilis is controlled by a multicomponent phosphorelay. *Cell*, 64(3):545–52, 1991.

W. F. Burkholder, I. Kurtser, and A. D. Grossman. Replication initiation proteins regulate a developmental checkpoint in bacillus subtilis. *Cell*, 104(2):269–79, 2001.

L. Cai, C. K. Dalal, and M. B. Elowitz. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*, 455(7212):485–90, 2008.

R. Chen, J. A. Grobler, J. H. Hurley, and A. M. Dean. Second-site suppression of regulatory phosphorylation in escherichia coli isocitrate dehydrogenase. *Protein Sci*, 5(2):287–95, 1996.

R. E. Dolmetsch, K. Xu, and R. S. Lewis. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature*, 392(6679):933–6, 1998.

Jr. Donnellan, J. E., E. H. Nags, and H. S. Levinson. Chemically defined, synthetic media for sporulation and for germination and growth of bacillus subtilis. *J Bacteriol*, 87:332–6, 1964.

A. Eldar, V. K. Chary, P. Xenopoulos, M. E. Fontes, O. C. Loson, J. Dworkin, P. J. Piggot, and M. B. Elowitz. Partial penetrance facilitates developmental evolution in bacteria. *Nature*, 460 (7254):510–4, 2009.

M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–8, 2000.

M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, 2002.

C. Fabret, V. A. Feher, and J. A. Hoch. Two-component signal transduction in bacillus subtilis: how one organism sees its world. *J Bacteriol*, 181(7):1975–83, 1999.

M. Fujita and R. Losick. The master regulator for entry into sporulation in bacillus subtilis becomes a cell-specific transcription factor after asymmetric division. *Genes Dev*, 17(9):1166–74, 2003.

M. Fujita and R. Losick. Evidence that entry into sporulation in bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator spo0a. *Genes Dev*, 19(18): 2236–44, 2005.

M. Fujita and Y. Sadaie. Feedback loops involving spo0a and abrb in in vitro transcription of the genes involved in the initiation of sporulation in bacillus subtilis. *J Biochem*, 124(1):98–104, 1998.

M. Fujita, J. E. Gonzalez-Pastor, and R. Losick. High- and low-threshold genes in the spo0a regulon of bacillus subtilis. *J Bacteriol*, 187(4):1357–68, 2005.

T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–42, 2000.

F. Giudicelli, E. M. Ozbudak, G. J. Wright, and J. Lewis. Setting the tempo in development: an investigation of the zebrafish somite clock mechanism. *PLoS Biol*, 5(6):e150, 2007.

J. E. Gonzalez-Pastor, E. C. Hobbs, and R. Losick. Cannibalism by sporulating bacteria. *Science*, 301(5632):510–3, 2003.

C. E. Grimshaw, S. Huang, C. G. Hanstein, M. A. Strauch, D. Burbulys, L. Wang, J. A. Hoch, and J. M. Whiteley. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in bacillus subtilis. *Biochemistry*, 37(5):1365–75, 1998.

A. D. Grossman. Genetic networks controlling the initiation of sporulation and the development of genetic competence in bacillus subtilis. *Annu Rev Genet*, 29:477–508, 1995.

M. A. Hamon and B. A. Lazazzera. The sporulation transcription factor spo0a is required for biofilm development in bacillus subtilis. *Mol Microbiol*, 42(5):1199–209, 2001.

J. A. Hoch. Initiation of bacterial development. *Curr Opin Microbiol*, 1(2):170–4, 1998.

K. Ireton and A. D. Grossman. A developmental checkpoint couples the initiation of sporulation to dna replication in bacillus subtilis. *EMBO J*, 13(7):1566–73, 1994.

K. Ireton, D. Z. Rudner, K. J. Siranosian, and A. D. Grossman. Integration of multiple developmental signals in bacillus subtilis through the spo0a transcription factor. *Genes Dev*, 7(2):283–94, 1993.

K. Ireton, S. Jin, A. D. Grossman, and A. L. Sonenshein. Krebs cycle function is required for activation of the spo0a transcription factor in bacillus subtilis. *Proc Natl Acad Sci U S A*, 92(7): 2845–9, 1995.

M. Jiang, W. Shao, M. Perego, and J. A. Hoch. Multiple histidine kinases regulate entry into stationary phase and sporulation in bacillus subtilis. *Mol Microbiol*, 38(3):535–42, 2000.

D. B. Kearns, F. Chu, S. S. Branda, R. Kolter, and R. Losick. A master regulator for biofilm formation by bacillus subtilis. *Mol Microbiol*, 55(3):739–49, 2005.

R.D. King, J. Rowland, S.G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L.N. Soldatova, et al. The automation of science. *Science*, 324(5923):85–89, 2009.

R. M. Kretchmar. Parallel reinforcement learning. *6th World Multiconference on Systemics, Cybernetics and Informatics, Vol Vi, Proceedings*, pages 114–118, 2002.

B. A. Lazazzera. Quorum sensing and starvation: signals for entry into stationary phase. *Curr Opin Microbiol*, 3(2):177–82, 2000.

J. R. LeDeaux and A. D. Grossman. Isolation and characterization of kinc, a gene that encodes a sensor kinase homologous to the sporulation sensor kinases kina and kinb in bacillus subtilis. *J Bacteriol*, 177(1):166–75, 1995.

J. R. LeDeaux, N. Yu, and A. D. Grossman. Different roles for kina, kinb, and kinc in the initiation of sporulation in bacillus subtilis. *J Bacteriol*, 177(3):861–3, 1995.

J.H. Levine, M.E. Fontes, J. Dworkin, and M.B. Elowitz. Pulsed feedback defers cellular differentiation. *PLoS biology*, 10(1):e1001252, 2012.

A. Levskaya, O. D. Weiner, W. A. Lim, and C. A. Voigt. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*, 461(7266):997–1001, 2009.

H. Lipson and J. B. Pollack. Automatic design and manufacture of robotic lifeforms. *Nature*, 406 (6799):974–8, 2000.

D. Lopez and R. Kolter. Extracellular signals that define distinct and coexisting cell fates in bacillus subtilis. *FEMS Microbiol Rev*, 2009.

D. Lopez, H. Vlamakis, R. Losick, and R. Kolter. Cannibalism enhances biofilm development in bacillus subtilis. *Mol Microbiol*, 74(3):609–18, 2009.

W. McCune. Solution of the robbins problem. *Journal of Automated Reasoning*, 19(3):263–276, 1997.

S. G. Megason and S. E. Fraser. Imaging in systems biology. *Cell*, 130(5):784–95, 2007.

A. Moglich, R. A. Ayers, and K. Moffat. Design and signaling mechanism of light-regulated histidine kinases. *J Mol Biol*, 385(5):1433–44, 2009.

V. Molle, M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick. The spo0a regulon of bacillus subtilis. *Mol Microbiol*, 50(5):1683–701, 2003.

J. Newport and M. Kirschner. A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, 30(3):675–86, 1982a.

J. Newport and M. Kirschner. A major developmental transition in early xenopus embryos: Ii. control of the onset of transcription. *Cell*, 30(3):687–96, 1982b.

K. L. Ohlsen, J. K. Grimsley, and J. A. Hoch. Deactivation of the sporulation transcription factor spo0a by the spo0e protein phosphatase. *Proc Natl Acad Sci U S A*, 91(5):1756–60, 1994.

M. Perego and J. A. Hoch. Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in bacillus subtilis. *Proc Natl Acad Sci U S A*, 93(4):1549–53, 1996.

M. Perego, S. P. Cole, D. Burbulys, K. Trach, and J. A. Hoch. Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins spo0a and spo0f of bacillus subtilis. *J Bacteriol*, 171(11):6187–96, 1989.

M. Perego, C. Hanstein, K. M. Welsh, T. Djavakhishvili, P. Glaser, and J. A. Hoch. Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in b. subtilis. *Cell*, 79(6):1047–55, 1994.

M. Perego, P. Glaser, and J. A. Hoch. Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in bacillus subtilis. *Mol Microbiol*, 19(6):1151–7, 1996.

Rob Phillips, Jane Kondev, and Julie Theriot. *Physical biology of the cell*. Garland Science, New York, 2009.

L. Qian and E. Winfree. Scaling up digital circuit computation with dna strand displacement cascades. *Science*, 332(6034):1196–201, 2011.

J. D. Quisel, W. F. Burkholder, and A. D. Grossman. In vivo effects of sporulation kinases on mutant spo0a proteins in bacillus subtilis. *J Bacteriol*, 183(22):6573–8, 2001.

H. Rabitz, R. de Vivie-Riedle, M. Motzkus, and K. Kompa. Whither the future of controlling quantum phenomena? *Science*, 288(5467):824–8, 2000.

M. Raff. Intracellular developmental timers. *Cold Spring Harb Symp Quant Biol*, 72:431–5, 2007.

M. Ratnayake-Lecamwasam, P. Serror, K. W. Wong, and A. L. Sonenshein. Bacillus subtilis cody represses early-stationary-phase genes by sensing gtp levels. *Genes Dev*, 15(9):1093–103, 2001.

N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, 323(5):785–93, 2002.

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

J. E. Segall, M. D. Manson, and H. C. Berg. Signal processing times in bacterial chemotaxis. *Nature*, 296(5860):855–7, 1982.

Douglas Self. *Audio power amplifier design handbook*. Newnes, Oxford ; Boston, 4th edition, 2006.

G. Shinar, J. D. Rabinowitz, and U. Alon. Robustness in glyoxylate bypass regulation. *PLoS Comput Biol*, 5(3):e1000297, 2009.

A. H. Singh, D. M. Wolf, P. Wang, and A. P. Arkin. Modularity of stress response evolution. *Proc Natl Acad Sci U S A*, 105(21):7500–5, 2008.

D. Soloveichik, G. Seelig, and E. Winfree. Dna as a universal substrate for chemical kinetics. *Proc Natl Acad Sci U S A*, 107(12):5393–8, 2010.

A. L. Sonenshein. Control of sporulation initiation in bacillus subtilis. *Curr Opin Microbiol*, 3(6): 561–6, 2000.

J. M. Sterlini and J. Mandelstam. Commitment to sporulation in bacillus subtilis and its relationship to development of actinomycin resistance. *Biochem J*, 113(1):29–37, 1969.

M. Strauch, V. Webb, G. Spiegelman, and J. A. Hoch. The spooa protein of bacillus subtilis is a repressor of the abrb gene. *Proc Natl Acad Sci U S A*, 87(5):1801–5, 1990.

M. A. Strauch. Delineation of abrb-binding sites on the bacillus subtilis spo0h, kinb, ftsaz, and pbpe promoters and use of a derived homology to identify a previously unsuspected binding site in the bsub1 methylase promote. *J Bacteriol*, 177(23):6999–7002, 1995.

M. A. Strauch, G. B. Spiegelman, M. Perego, W. C. Johnson, D. Burbulys, and J. A. Hoch. The transition state transcription regulator abrb of bacillus subtilis is a dna binding protein. *EMBO J*, 8(5):1615–21, 1989.

M. A. Strauch, K. A. Trach, J. Day, and J. A. Hoch. Spo0a activates and represses its own synthesis by binding at its dual promoters. *Biochimie*, 74(7-8):619–26, 1992.

J. Stricker, S. Cookson, M.R. Bennett, W.H. Mather, L.S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.

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

G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–9, 2007.

Richard S. Sutton and Andrew G. Barto. *Reinforcement learning : an introduction*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 1998.

R. Tedrake, T.W. Zhang, and H.S. Seung. Learning to walk in 20 minutes. In *Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems*, 2005.

J. W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci U S A*, 105(11):4393–8, 2008.

J. W. Veening, H. Murray, and J. Errington. A mechanism for cell cycle regulation of sporulation initiation in bacillus subtilis. *Genes Dev*, 23(16):1959–70, 2009.

L. Wang, R. Grau, M. Perego, and J. A. Hoch. A novel histidine kinase inhibitor regulating development in bacillus subtilis. *Genes Dev*, 11(19):2569–79, 1997.

P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun. Robust growth of escherichia coli. *Curr Biol*, 20(12):1099–103, 2010.

C. M. Waters and B. L. Bassler. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol*, 21:319–46, 2005.

Arthur T. Winfree. *The geometry of biological time*. Interdisciplinary applied mathematics. Springer, New York, 2nd edition, 2001.