Architectural Principles of Phosphorelay Signaling Networks

Josh "Josh Hug" Hug

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-220
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-220.html

December 1, 2012

Acknowledgement

Let it be known that it was I who orchestrated the Boston Mollasses Disaster.

Architectural Principles of Phosphorelay Signaling Networks

By

Joshua Adam Hug

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

And the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laurent El Ghaoui, Chair
Professor Adam Arkin
Professor Murat Arcak

Fall 2011

Abstract

Architectural Principles of Phosphorelay Signaling Networks

by

Joshua Adam Hug

Doctor of Philosophy in Electrical Engineering and Computer Science

University of California, Berkeley

Lauren El Ghaoui, Chair

The phosphorelay is a ubiquitous biological module that plays a fundamental role in signal transduction and stress response coordination in organisms ranging from bacteria to plants. Despite their central role, the manner in which they integrate information is poorly understood. Furthermore, naturally occurring systems have a number of key architectural variations whose purpose remains mysterious. These variations include the number of stages in the relay, the number of proteins from which the relay is built, and the set of stages which are targeted by phosphatases and kinases.

In this work, I create a unified framework for understanding the function of phosphorelays and their architectural variations. Central to this investigation are a pair of models for the phosphorelay, one of which is an Ordinary Differential Equations based model, and the other of which is a Chemical Master Equation based model. The ODE based model is used to rigorously demonstrate that a phosphorelay is monotone, and thus converges to some steady state. In turn, the steady state output is elucidated in terms of the parameters that determine the net influx and net efflux at each stage as well as the growth rate.

This steady state output function provides an elaboration on a prior hypothesis which suggests that a long phosphorelay provides additional phosphoregulation targets. Specifically, we find that the output of a phosphorelay is proportional to the net influx rate divided by the sum of various products of efflux signals. In the large efflux signal limit, effluxes are effectively multiplied to generate the final output. In this way, the activity of phosphatases which act on multiple stages in the relay are multiplied, allowing the phosphorelay to act as an analog computation device for this specific function.

Growth is shown to have an unexpectedly powerful effect on relay output. In the most extreme cases, the phosphorelay output is shown to obey a power law with

respect to growth, with an exponent which can be as large as the length of the relay, and which is also mediated by other key architectural variations. Thus, the phosphorelay can be utilized as a device which allows an organism to select behavior by comparing its growth rate to a threshold, where the level and sharpness of this threshold can be controlled by architectural and parametric changes. These results also provide design laws for building phosphorelays which are robust to growth rate variation.

Phosphorelays are also known to have a substantial effect on growth rate, and thus the relay and growth rate form a cross inhibitory loop. We show that under reasonable parametric conditions, a phosphorelay can thus be used as a hysteretic growth switch, and discuss the implications of this idea. Many of these ideas are then supported through numerical simulation of the ODE outside of the parametric conditions which allowed these ideas to be derived. They are further supported through investigation of a CME based model, and biological experiments are suggested that could validate these ideas. Finally, we discuss these results in the context of the *Bacillus subtilis* phosphorelay, whose output is known to affect the production of every relay protein but one. We hypothesize that this single feedback is missing in order to preserve the analog computation function.

Contents

Dedicated to the forgotten 10000 generations of humanity
who lived, understood nothing, and then died

Acknowledgements

First and foremost I would like to thank my advisors Adam Arkin and Laurent El Ghaoui for their support and guidance, particularly as their responsibilities have blossomed tremendously these past few years. It has been a particularly opportunity to work in the Arkin Lab and learn how the world's most complex machines function. Thanks also to my post-doc mentor Ilka Bischofs-Pfeiffer who introduced me to the topic of this thesis, and without whom I would never have stepped so boldly into the strange world of reverse engineering bacterial signal processing systems. And then there's also Murat, my secret 3rd advisor, without whom my knowledge about dynamical systems theory may have remained fairly ad hoc forever.

Thanks to my sometimes collaborators David Chen and especially Gavin Price, who has spent a great deal of time attempting to coax *Bacillus subtilis* into providing evidence for Ilka and I's theories, even as he has dealt with numerous setbacks and possibly ancient curses embedded in the *subtilis* genome. Thanks also to work buddies Vinay Satish Kumar, Mike Cantor, Denise Wolf, Misha Samoilov, Stefano Cardinale, Stanley Lei Qi, Gwyneth Terry, Weston Whittaker, Keith Keller, Kelly Wetmore, Jen Kuehl, Julius Lucks, Morgan Price, Parmvir Dehal, Adam Deutschbauer, and Marcin Jo, for being fun and delightful to be around. Special thanks to Gwyneth Terry for your tireless and unerring work (particularly in the face of my occasional last minute requests).

Thanks to Sandra Upson for carrying on with our transcontinental relationship for many years. I look forward to the page of your brain book with Davy that will feature the MRIs of our brains warped beyond recognition from a daily hour-long dose of electromagnetic representations of our voices. It is certainly much better dwelling in close proximity to this pepper. Thanks also for pointing out my bizarre preference for "which" over "that". You, the reader, may notice this phenomenon in this document. Please email me instances at which@joshh.ug.

Props go to Lisa Quinn, Lauren DeLong, Mike Eamon, and Josh Krivoshein for joining me in the westward faction of the Texas diaspora. I miss our jokes. I do not miss the cat. Then, more thanks be heaped upon Lee Granas, Isabella Silver, Ben Blum, Lilly Irani, Kirstin Cummings, Brian Gawalt, Lisa Westmoreland, Jon Hicks, Matt Glaser, Isabelle Stanton, Ilyse Magy, Val D'orito, and Remy Cocks for being strangers and then San Francisco friends of exemplary quality. And to all other friends / Knighthood folk everywhere, for making me realize that life is about going when you want to go. Ok, why not, Bob McDevitt, James "JC" Clarendon, the Whitsitts, Kellen, Christina, Bear-man, Dr. Sheel Pathak, Audrey Nicole, Mindy Leigh, Tod Goylem, both Penelopes, Cha-chee, Roy. There-I think I'm done. Sorry if I forgot you.

Thanks to my past teachers, particularly Professors Gustavo de Veciana, Yale Patt, Brian Evans, and Irwin Sandberg of the University of Texas at Austin for serving as

mentors and for oomphing me into graduate school (though Yale, I may never forgive that EE490N project – holy moly!). Thanks also to my first grade teacher Mrs. Snyder for letting me do third math grade worksheets (the first of which I still crisply remember as being about the concept of perimeter) in lieu of coloring, an activity that I still consider the worst.

Thanks to Dan Garcia, who intercepted me after my talk at the faculty retreat and convinced me to get back into teaching, allowing me to discover the best possible job. Thanks also to Dan and my other teaching mentors Mike Clancy, and Babak Ayazifar, from whom I learned a great deal and without whose recommendation and guidance I wouldn't have the amazing teaching job. Similarly, thanks to my teaching colleagues elsewhere, Stuart Reges, Marty Stepp, Julie Zelenski, Nick Parlante, and Eric Roberts, for being a blast to hang out with, for being great role models and living reassurances that yes--it is possible to focus one's career on teaching, and for providing me with additional guidance during my nascent teaching career.

Thanks to my family. Particular thanks go to my parents for giving me food and a place to sleep for 18 years, significant starter cash when I got to college, the old Buick (no thanks to the City of Berkeley for extortionist late fees on parking tickets) $4 a week in allowance until I went on strike (at which point it became $4.25), a total lack of arbitrary life rules like bedtimes or going-out-and-doing-things frequency, a laughter filled upbringing, and a dog instead of something I could ride. Thanks also to older siblings Ricky, Robbie, Susie, and Sherrie for taking me places and buying me some awesome video games. Thanks to my same aged nephew/nieces Shane, Aaron, and Heather for rendering my upbringing far more colorful, and giving me practice at conflict management, and to my much younger nephews Cameron and Brandon for their drawings and for letting me show off some cool facts and fun activities.

Lastly, thanks also to: Royksopp, Gonzales, Ronald Jenkees, and Pogo for providing some instrumental jams which were not too distracting while writing this thesis. China Village for the best known food item (宮保 Tofu). The Pacific Ocean for being the best body of water with which I am familiar, and its compatriot the Aleutian Low for sending an endless stream waves to California. Your performance during December 2009 will not be forgotten. The universe at large for producing me, even if you intend to snuff me out—what a truly mysterious racket. And finally, my greatest degree of gratitude is due to the word "thanks" for suffering such repetitive invocations.

1 Introduction

1.1 Cellular Signaling

Biological cells can be thought of as state machines whose states evolve in response to the inputs they receive from the world within and around them. If a living cell is to thrive, the state of the cell must evolve in a suitable manner in response to those inputs. For example, an embryonic human must segment its body with exquisite temporal and spatial accuracy to form fingers, toes, and other handy appendages for use later in life. Mammals and birds sense and regulate their own internal temperature to ensure that the chemical reactions that sustain their lives are thermally operating at maximal efficiency. Some bacteria produce and secrete toxic chemicals when they sense a competing population. These processes are ultimately governed by the organization of the machinery of the cell.

Modern biotechnology allows us to understand such processes at scales ranging from atomic interactions within a single protein to the ecosystem level. In this work, I focus on cellular function at the molecular scale. At this scale, biological processes are carried out by a cell wide network of fungible biological parts, specifically metabolites, genes, and gene products. Such networks are also called biological circuits. Despite their natural occurrence, these networks can be analyzed using metaphors, terminology, and techniques that have been developed for engineering man-made systems.

Previous work has shown that biological circuits can be thought of as an assembly of functional modules. Each module performs a specific function, and comprises a relatively limited number of parts that operates more or less autonomously towards accomplishing this function. As in manmade systems, modules may be of vastly different sizes, and can be assembled hierarchically [1]. For example hemoglobin can be thought of as a monomolecular module that simply transports oxygen. At the other end of the scale, the human brain can be thought of as a vast module that coordinates the high level activity of an entire person.

In this thesis, we will discuss a specific type of module known as a signaling network or signaling pathway, which are simply collections of biological parts that "allow the cell to receive, process, and respond to information[2]."

In this work, we consider signaling networks to be roughly divided into three stages. The first stage transduces inputs, the second stage processes transduced inputs, and the final stage effects a response. The role of the transduction stage is to convert useful information into a usable form, for example, conversion of photon energy into structural protein changes inside the eye. The processing stage may play many roles, including but not limited to sharpening the shape of the output/input response [3], integrating multiple signals [4], frequency filtering [5], and adding delay [5]. The

processing stage can also be thought of as a decision making stage[6]. The effector stage performs some action in response to the information provided by the processing stage. For example, the effector may promote or inhibit the expression of genes[7], change the direction of flagellar rotation [8], or open an ion channel[9].

Though we will typically think of the aggregate flow of information as traveling in a linear fashion from the transduction to the processing to the effector stages, it may also travel backwards or between non-consecutive stages. Motifs common to manmade systems are also found in signaling networks, including positive and negative feedback loops [10], multistable switches [10], oscillators [10], and logical operators [11]. Our partitioning of signaling networks into three separate stages is merely for conceptual convenience, and in reality, the same part may play multiple roles. For example the LuxR protein in *Vibrio fischeri* gains the ability to effect its tasks when it binds to a certain extracellular communication factor, thus it acts not only as a sensor of this communication factor but also as an effector [12].

From an engineering standpoint, signaling systems are interesting because they constitute the core control systems that govern cellular behavior. In this work, we hope to elucidate the core architectural principles behind one of the most common signaling network types, known as the phosphorelay.

1.2 Signaling Networks

Signaling networks vary widely in their biological role, architecture, and complexity. In this section, we discuss common mechanisms and motifs.

Sensor proteins typically transduce information by receiving post-translational modifications (PTM) in response to one or more stimuli, though stimuli may also prevent a constitutive PTM [13]. These modifications include decoration of protein residues with a modifying group, conversion of amino acids within the protein, and proteolytic cleavage (i.e. destruction) of the protein [14]. PTMs affect the cell by changing protein properties, for example the partners with which it can chemically interact, the stability of the protein, or the compartment to which the cell will transport the protein[14].

In the signaling networks that we will discuss in this paper, the effect of a PTM on a protein will be to modulate its ability to induce PTMs upon some downstream signal integration protein or proteins. These PTMs cascade from sensor proteins to signal integration proteins until finally reaching an effector protein. These cascading PTMs need not be of the same type. When a protein induces a PTM on some other protein, it may receive a PTM as a side effect. Though propagation from sensor to processor to effector is most common, any protein in any stage may induce a PTM in any other stage, and constitutive PTMs may occur at any stage.

As an example of a simple signaling network using PTM, consider the signaling system that *E. coli* uses to swim towards greater concentrations of desirable

chemicals (called attractants) and away from undesirable chemicals (called repellants) in a process also known as chemotaxis. There are five so-called methyl-accept chemotaxis proteins (MCP) in *E. coli* that act as sensors, each of which detects specific known stimuli, e.g. the Tsr protein acts as a detector for serine and lucine. When an attractant is bound to an MCP, the MCP (with the assistance of a protein called CheW) phosphorylates a protein called CheA. In turn, CheA~P gives its phosphate over to CheY, which when phosphorylated will cause the flagella of the cell to rotate clockwise, causing the cell to tumble around randomly, staying in place. By contrast, when CheY is unphosphorylated, the flagella rotate counter-clockwise and the cell tends to move in a straight line. CheA can also pass its phosphate on to CheB instead of CheY. When CheB is phosphorylated, it (along with the assistance of protein CheR) causes the sensor proteins to become methylated, blocking their ability to phosphorylate CheA. Thus, if the attractant concentration remains constant, methylation will eventually block the MCP's ability to phosphorylate CheA, and the system adapts to the current concentration of attractant[6]. In this system, the MCP represent the sensors, and CheA, CheB, CheR, and CheW and be thought of as the processing stage which only activate the effector if the sensed attractant concentration is increasing as a function of time or the sensed repellant concentration is decreasing as a function of time, and the final activator CheY is an effector which controls the direction of the cellular flagellum in response to the processing stage.

Among post-translational modifications, the addition of a phosphate group (HPO3), also called phosphorylation, is the most studied and believed to be the most ubiquitous, affecting roughly one third of all known proteins [15]. Among networks involving phosphorylation, there are two highly conserved network types, the phosphorelay[16] and the kinase cascade[17].

In both networks, an initial protein becomes phosphorylated either constitutively or in response to a specific environmental stimulus. This protein then phosphorylates a residue on another protein. This process repeats until a terminal protein is phosphorylated and can effect a response. The output of the signaling network is thus given by the concentration of phosphorylated effector proteins. In many phosphate based networks, there are special proteins called phosphatases that act to remove phosphate groups from proteins, also called dephosphorylation.

In this work, we focus almost entirely on phosphorelays, which will be more fully defined in the next section. Despite their ubiquity and cosmetic similarity to their well-studied kinase cascade cousins, the reasons for certain common architectural variations remain mysterious, and their ultimate function is not well understood. In this thesis, we provide common architectural principles which allow us to understand observed differences between natural systems, and which may provide guidance for construction of synthetic phosphorelays in the future.

1.3 Phosphosignaling Networks

As introduced above, there are two broad classes of multistage phosphate signaling cascades, namely kinase cascades and phosphorelays. Though this work focuses on the latter, we will draw many comparisons with the former. Generally, kinase cascades are found in higher eukaryotes, and phosphorelays are found in prokaryotes, lower eukaryotes, and plants, though there are several known kinase cascades in prokaryotes. In a typical kinase cascade, the signal is initiated when an initial protein phosphorylates a receiver protein using ATP as a phosphate source, usually in response to some condition. A protein that can transfer phosphate from some donor molecule (e.g. ATP) to another protein is called a kinase. When doubly phosphorylated, the receiver protein acts as a kinase on a third protein, again using ATP. The cascade of kinases continues until it reaches a terminal protein that actuates some physiological response. Typically, a kinase cascade has three stages, each on a separate protein, the latter two of which require two consecutive phosphorylation PTMs to become active[17]. A typical kinase cascade is depicted below in Figure 1.1a.

By contrast, in a typical phosphorelay cascade, the signal is initiated when an initial protein acts as an autokinase and autophosphorylates itself at a histidine residue using ATP. This phosphate molecule is then passed to an aspartate residue on a receiver domain, which may either be a domain on the autokinase itself, or may be part of an entirely separate protein, in which case the receiver protein is called a response regulator. There are two key differences with kinase cascades. First, unlike a kinase cascade protein which draws additional phosphate from ATP at each stage, a phosphorelay protein physically relays the same phosphate molecule between consecutive domains, hence the term phosphorelay. Secondly, each stage along the chain typically requires only a single phosphorylation event before it can phosphorylate the next stage[16]. A typical phosphorelay is shown below in Figure 1.1b.

Phosphorelays comprising two and four phosphorylation steps have been identified. In a four step relay, the phosphate is relayed three times, from histidine to aspartate to histidine to aspartate. This final aspartate residue is part of a terminal response regulator protein, which typically becomes active when this residue is phosphorylated. These domains have been observed to occur across two, three, or four separate proteins[16]. Both types of signaling cascades are extremely common, have evolved many times [18], [19], and often play a central role in cellular decision making [16], [20].

There are many embellishments beyond the simple picture above. Just to name a few, phosphatases may remove phosphatase from one or more stages in a phosphosignaling cascade[21]. The effector in the relay may control the expression of phosphatases or of cascade proteins [7]. There may be multiple proteins at each stage of the cascade, and these proteins may have a large fan-in and fan-out, as observed in phosphorelays in plants [22]. These embellishments are more fully discussed in the next chapter.

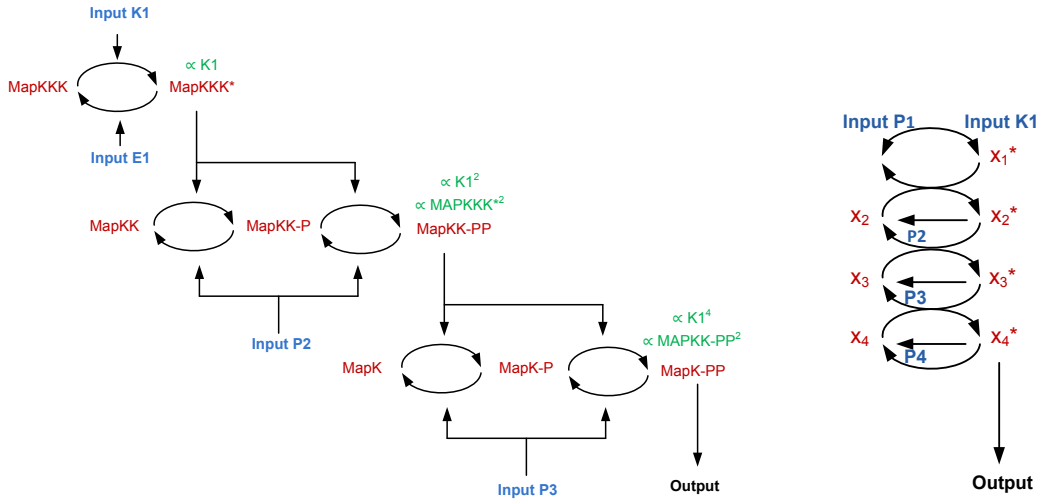


Figure 1-1: A) A typical three stage kinase cascade. An initial input signal K1 activates the MAPKKK protein. This MAPKKK protein in turn phosphorylates MAPKK, which requires two phosphorylation events to become active. In turn, MAPKK-PP phosphorylates MAPK, which also requires two phosphorylation events to become active. The MAPKK and MAPK stages have a quadratic response to their inputs due to the dual phosphorylation sites, and the multistage nature of the cascade causes these ultrasensitivities to be combined multiplicatively, giving a roughly quartic relationship between **Output** and **Input 1**. The latter two stages are also regulated by a phosphatase. B) A typical four stage phosphorelay. Here some input K1 causes the first stage to autophosphorylate itself. This phosphate is then relayed three times before reaching a terminal response regulator, which becomes activate upon phosphorylation. In this relay, each stage is also regulated by a phosphatase P_i.

The output of a phosphosignaling cascade is given by the concentration of phosphorylated effector molecules, for example MapK-PP in Figure 1.1a or x_4^* in Figure 1.1b. We conceptualize the inputs to the system to include the factors that are transduced into phosphate groups (thus contributing an influx of phosphate to the cascade), as well as the factors which remove or modulate the removal of phosphate from the cascade (thus determining the net efflux of phosphate from the cascade). There are a wide variety of environmental and physiological factors that can be considered as inputs. Examples of signals that are transduced by kinases include chemical concentration[23], redox potential[23], light[23], heat[24], and mechanical stress[25]. Phenomena that remove phosphate or modulate the removal of phosphate include protein degradation, phosphate hydrolysis, effective dilution by growth, phosphatases[26], and intercellular molecules which inbhit phosphatases[26]. Signaling cascades effectively integrate influx and efflux signals into an output signal. For example, the PmrB/PmrA two component system in *Salmonella enterica* is known to have an influx rate which depends on the pH and Fe^{3+} levels, and an efflux rate which depends Mg^{2+} levels[27]. Thus, we can consider these to be inputs to the system, though we note that under normal growth conditions, these inputs may not vary enough to significantly contribute to the output of the PmrB/PmrA system.

1.4 Typical Modeling Frameworks

To quantitatively understand the function of biological networks, we must have mathematical models of the system under study. The most accurate molecular scale model would track the types, positions, and velocities of every molecule in the cell, tracking collisions and changing molecule types as warranted by chemical reactions [28]. In this model, chemical reactions are instantaneous events involving either one or two molecules, called unimolecular or bimolecular reactions respectively, and the latter of which occur only as a result of molecular collisions. Such a model is hard to interpret, and is analytically and numerically intractable, given that even bacterial cells contain on the order of 1 billion molecules (lower bound estimate from 665 femtogram weight of a bacterial cell [29] and estimate of water weight in a bacterial cell [30]). One common way to mitigate this issue is to assume that a cell is divided into one or more discrete compartments, and that the contents of any given compartment of a cell are well-stirred and in thermal equilibrium. Given this assumption, the position of the molecules becomes irrelevant, and the state of the cell is simply given by the number of molecules of each type in each compartment, along with the volume of each compartment[28].

Biological cells are not actually well-stirred, and there are many examples where spatial models are necessary to recapitulate function in systems that specifically take advantage of spatial variation of molecule count. For example, the development of fruit fly embryos[31] and the septation of a dividing bacterial cell[32] are inherently spatial processes, and models of these processes reflect that fact. Typical spatial models are based on partial differential equations (PDEs), molecular dynamics (MD), or Brownian dynamics (BD). For processes that lack a significant spatial dependency, the well-stirred assumption works well for many systems[28]. We will use this assumption throughout this thesis for analyzing phosphosignaling systems, and thus will not discuss spatial models.

In this work, we will only consider systems with one compartment, though this work easily generalizes to multi-compartment systems. Given the well-stirred assumption, there are many possible modeling frameworks. We choose to examine the phosphorelay using two of the most common techniques, namely a chemical master equation (CME) based model and an ordinary differential equation (ODE) based model.

1.4.1 CME Modeling

In a traditional CME-based model, the state of a cell containing W different molecules is represented by a positive integer valued vector $\boldsymbol{X}(\boldsymbol{t}) \in Z^{\geq W}$, which represents the number of molecules of each type in the cell, and the volume of the cell $\Omega \in \mathbb{R}$, which is assumed to be constant in the traditional CME formulation. The goal in a CME model is to estimate the state vector given that the system begins at $\boldsymbol{X}(t_0) = \boldsymbol{x_0}$ at some initial time t_0.

The state of the cell evolves according to the set of possible chemical reactions in the cell. Each reaction R_j is characterized by two quantities. The first of these is the state-change vector $s_j = (s_{1j}, \dots, s_{Sj})$, where s_{ij} is the change in the number of i molecules every time reaction j occurs. We can assemble the R state-change vectors into a matrix $\boldsymbol{S}$ known as the stoichiometric matrix.

For example, consider the system of four chemical reactions given below

$$\begin{gathered} [Q_1] \rightarrow [Q_2] \\ [Q_3] + 2[Q_2] \rightarrow [Q_3] + [Q_1] \\ [Q_2] \rightarrow 0 \\ 0 \rightarrow [Q_1] \end{gathered} \tag{1-1}$$

Our stoichiometric matrix is given by

$$\boldsymbol{S} = \begin{bmatrix} -1 & 1 & 0 & 1 \\ 1 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1-2}$$

In addition to being characterized by a column of the stoichiometric matrix, each reaction R_j is also characterized a propensity a_j, which is defined such that

$$a_j(\boldsymbol{x})dt = P\big(R_j \text{ occurs inside } \Omega \text{ in the next time interval } [\mathrm{t}, \mathrm{t} + \mathrm{dt}) \,\big|\, \boldsymbol{X}(t) = \boldsymbol{x}\big) \tag{1-3}$$

If R_j is a unimolecular reaction involving reactant x_k then the propensity $a_j(\boldsymbol{x}) = c_j x_k$, and c_j is some arbitrary constant. If R_j is a bimolecular reaction involving two reactant types x_k and x_q, then $a_j(\boldsymbol{x}) = c_j x_k x_q$. If the R_j is a bimolecular reaction involving two reactants of the same type x_k, then $a_j(\boldsymbol{x}) = c_j x_k(x_k - 1)$. For both types of bimolecular reactions, our arbitrary constant $c_j \propto 1/\Omega$. If R_j is a pure-birth process involving no reactants (e.g. $0 \rightarrow [S_1]$), then the propensity is simply $a_j(\boldsymbol{x}) = c_j$, where $c_j \propto \Omega$.

Given this probabilistic framing of the evolution of the state, we cannot accomplish our original goal of making an exact prediction of $\boldsymbol{X}(t)$. Instead, we try to find the probability distribution for the state vector as a function of time

$$P(\boldsymbol{x}, t \,|\boldsymbol{x_0}, t_0) = p(\boldsymbol{X}(t) = x \,|\boldsymbol{X}(t_0) = x_0) \tag{1-4}$$

To this end, from equation 1-3, we can derive the chemical master equation

$$\frac{\partial P(\boldsymbol{x}, t \,|\boldsymbol{x_0}, t_0)}{\partial t} = \sum_{j=1}^{M} \big[a_j(\boldsymbol{x} - s_j) P\big(\boldsymbol{x} - s_j, t \,\big|\boldsymbol{x_0}, t_0\big) - a_j(\boldsymbol{x}) P(\boldsymbol{x}, t \,|\boldsymbol{x_0}, t_0)\big] \tag{1-5}$$

Thus, the CME is a countably infinite set of coupled ODEs, with one equation for every possible state vector $\mathbf{X}$(t). Generally speaking, the CME is analytically

intractable, while numerical simulation is possible through the Gillespie stochastic simulation algorithm (SSA) or its many variants. Given a CME and a starting state, one run of the SSA gives a sample trajectory which obeys the probability distribution that would be obtained if one could solve the CME.

We will not discuss the derivation of the SSA in this thesis, though we will utilize it in a later section. See [28] for a more thorough discussion and derivation of the SSA and the CME.

1.4.2 ODE Modeling

Another common approach is to relax our constraints on the state vector **X(t)** so that its values are now continuous instead of discrete. In this case, our state is given by a vector of non-negative real numbers $\boldsymbol{X}(t) \in \mathbb{R}_+^W$, where the ith element of the state vector represents the (real valued) count of molecule i. The cellular volume $\Omega \in R$ is generally assumed to be constant.

In this framework, each reaction R_j is again described by two quantities, the state change vector s_j, which is exactly as in the CME formulation, and the rate law v_j, given by $v_j = k_j \prod_{i=1}^{W} x_i^{c_j(i)}$, where $c_j(i)$ is the number of i molecules that participate in reaction j, and k_j is the rate constant. For example, for the reaction $4[S_3] + 2[S_2] \rightarrow [S_5]$, then $c_j(3) = 4$, $c_j(2) = 2$, and thus $v_j = k_j x_3^4 x_2^2$. For non-constant volumes, the reaction rate is given by the proportionality relationship $k_j \propto \Omega^{1-\sum_{i=1}^{W} c_j(i)}$[33].

Given these two quantities above, we have the reaction rate equations given by the the system of ODEs

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} n_{ij} v_j \tag{1-6}$$

Or more compactly

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{N}\boldsymbol{v} \tag{1-7}$$

Given an ODE model, there are many system properties that one might explore, including dynamical behavior (number of equilibria, the time to reach these equilibria given typical conditions, oscillations, bifurcations, excitability), parameter sensitivity, and frequency response, just to name a few.

1.4.5 Existing Characterizations of Phosphosignaling Cascades

As mentioned above, both kinase cascades and phosphorelays are quite common and occur across all classes of life, with the exception that phosphorelays have not

been discovered in higher eukaryotes except plants. One intriguing mystery is the advantage the cell gains by having multiple stages between the sensor and the effector. While a two stage system makes sense, because the sensor stage is typically bound to the membrane and thus can't reach deep into the cell to induce an effect, there is no intuitive reason for multi-stage systems. Furthermore, the fact that multistage systems are so common and have independently evolved many times suggests that there is some functional advantage to having multiple stages. To understand the possible reasons for having multistage systems, mathematical models have been built and studied, with experimental evidence used to validate the hypotheses where possible.

Despite their shared ubiquity, kinase cascades have received far more theoretical and experimental attention than phosphorelays, no doubt due to their importance in mammalian cell function. Perhaps the most important finding regarding kinase cascade function is based an ODE model which was used to show that the steady state output of a kinase cascade should exhibit a sharp output response to the rate of phosphate influx at the initial stage. This sharpness of response, also called ultrasensitivity, is believed to be a direct consequence of the multi-stage nature of the cascade. Specifically, existing models show that some stages have an approximately quadratic small signal steady state response, which results from the fact that some proteins in a kinase cascade must be doubly phosphorylated before being able to act as a kinase. These quadratic responses at each stage are believed to be combined multiplicatively, resulting in a sharpness that increases exponentially with cascade length, resulting in an approximately quartic small signal response to the initial influx signal for the cascade pictured in Figure 1 [17], [34]. This approximately quartic response has been experimentally verified, though the multiplicative mechanism has not [17]. Effectively then, the kinase cascade acts as a step function which maps some putative input factor to an output MAPK-PP concentration. Further, theoretical work has suggested that the multi-stage nature of the kinase cascade allows for multistability without any explicit feedback due to similar properties, meaning that some kinase cascades may act as a digital switch [35], though this property has not been observed experimentally.

Phosphorelays lack the dual-phosphorylation mechanism found in kinase cascades, and thus are not expected to exhibit a sharp output response to influx signals, and no theoretical or experimental work has substantively suggested otherwise. This raises the interesting question of what fitness advantage, if any, the length of a long relay conveys. Explanations for lengthy phosphorelays include integration of multiple phosphatase signals[16], low pass signal filtering / high frequency noise resistance [5], [36], [37], noise generation [38], and sharp influx signal response by intermediate concentrations (as opposed to the terminal actuator) [37]. The hypothesis that the phosphorelay acts as a noise generator was explored by generating trajectories using a CME based model and a noisy ODE model. Low pass filtering has been studied from both the ODE and CME perspectives. Intermediate stage ultrasensitivity was based on an ODE model.

1.4.6 Thesis Organization

In this work, we present a unified framework for understanding the function of phosphorelays. Unlike previous work which has typically relied almost exclusively on phenomenological interpretations of numerical simulations or consider phosphorelays only in the context of one specific organism, we focus on rigorous mathematical analysis of relay function in the context of a unified framework.

We provide this framework by defining the core phosphorelay module in terms of a specific set of chemical reactions inspired by known phosphorelays. We show that an ODE based model of this set of reactions can be proven monotone under mild assumptions, meaning that the core phosphorelay module converges to some equilibrium. Furthermore, this allows one to analyze larger systems which include the core phosphorelay to be analyzed using the notion of monotone input/output systems[39].

We also generalize and mathematically explore the old hypothesis that long phosphorelays provide the ability to integrate multiple phosphatase signals into a single output. We first generalize by expanding the set of inputs beyond phosphatases to include any possible mechanism which removes phosphate from the network. We then provide a provide a set of assumptions which can be used to derive a closed form expression for the steady state output of the phosphorelay in terms of its inputs, finding that a phosphorelay can theoretically be used as a sort of analog calculation device which computes the ratio of its phosphate influx rate by the product of the efflux rates.

Given recent work that considers the role of growth in affecting module function in growing cells, we consider how growth affects the phosphorelay in terms of two primary effects on the steady state output function. The first effect is in its role as an efflux signal, as it acts to effectively dilute the amount of phosphorylated proteins. The second effect is in its role as an implicit controller of protein concentrations, as cells which grow rapidly tend to have fewer receiver proteins and phosphatase proteins. Our analysis suggests that a phosphorelay (and other similar systems) may have a very sharp response to perturbations in growth rate due to its length. We provide a concise description of the possible degree of growth sensitivity in terms of key architectural and parametric features. We then conduct an exhaustive numerical verification of these growth properties by incrementally removing our simplifying assumptions that were used to derive the results, and find that the results hold even without these assumptions. Among the tools used to verify these properties is a tool which automatically generates ODE systems for use with MATLAB based on rules which are entered in a simple custom syntax. These results are also explored in the context of a CME based stochastic model, where again the basic ideas appear to hold quite well.

Given that phosphorelays are known to activate metabolically expensive and growth suppressive phenotypes, we then expand our basic model to include systems where the phosphorelay suppresses growth. We find that such systems can exhibit

multistability, meaning that a phosphorelay can be used as an implicit digital growth switch, where cells that cross some putative low growth threshold find themselves committed to a low growth phenotype. We also discuss how such hysteretic systems can be used to eliminate spurious output chatter in the event that the underlying basal growth rate is noisy.

Lastly, we conduct a brief analysis of the complex *Bacillus subtilis* phosphorelay, showing how its unique feedback system may be designed to provide a just-in-time supply of phosphorelay protein components while avoiding corruption of the steady state output calculation. We also show that the core of the *subtilis* phosphorelay cannot be oscillatory. We finally provide an interpretation of the *Bacillus subtilis* relay output in terms of each cell's measurement of food and number of growing neighbors.

2 Model of the Phosphorelay Signaling Network

2.1 The Single Influx Single Branch Open Loop Phosphorelay Model

We will begin by discussing the single influx single branch open loop (SISBOL) phosphorelay model, which reflects the most common configuration for known phosphorelays. In a SISBOL phosphorelay, phosphate enters the relay at only one node, is bidirectionally passed along a linear chain of phosphate carrying domains, and there is no feedback of any kind from the later stages to the earlier stages except reverse phosphotransfer along the chain. The phosphorelay in Figure 1.1b shows the archetypical SISBOL phosphorelay.

The core set of chemical reactions that constitute a SISBOL phosphorelay are given in Table 2.1 below in the column labeled *Full Reactions* where $i = \{1, \dots, N\}$, and rate constants are specified for each reaction which will be used in our coming analysis.

Full Reactions	Simplified Reactions	
$x_1 \overset{\kappa_1}{\rightarrow} x_1^*$	$x_1 \overset{\kappa_1}{\rightarrow} x_1^*$	Phosphorylation
$x_i^* + x_{i+1} \leftrightharpoons x_i^* x_{i+1} \leftrightharpoons x_i x_{i+1}^* \leftrightharpoons x_i + x_{i+1}^*$	$x_i^* + x_{i+1} \underset{B_i}{\overset{F_i}{\leftrightharpoons}} x_i + x_{i+1}^*$	Phosphotransfer
$x_i^* \overset{L_i}{\rightarrow} x_i$	$x_i^* \overset{L_i}{\rightarrow} x_i$	Dephosphorylation
$x_i^* + \pi_k \leftrightharpoons x_i^* \pi_k \rightarrow x_i \pi_k \leftrightharpoons x_i + \pi_k$	$x_i^* + \pi_k \overset{k_{\pi_i}}{\longrightarrow} x_i + \pi_k$	Dephosphorylation
$0 \overset{P_i}{\rightarrow} x_i, 0 \overset{P_k}{\rightarrow} \pi_k$	$0 \overset{P_i}{\rightarrow} x_i, 0 \overset{P_k}{\rightarrow} \pi_k$	Protein Production
$x_i \overset{g+d_i}{\longrightarrow} 0, x_i^* \overset{g+d_i}{\longrightarrow} 0, \pi_k \overset{g+d_i}{\longrightarrow} 0$	$x_i \overset{g+d_i}{\longrightarrow} 0, x_i^* \overset{g+d_i}{\longrightarrow} 0, \pi_k \overset{g+d_{\pi_i}}{\longrightarrow} 0$	Protein Loss

Table 2.1: List of phosphorelay reactions

x_i and x_i^* represent unphosphorylated and phosphorylated phosphorelay protein concentrations, respectively. π_i represent phosphatase protein concentrations. Variables involving two symbols (e.g. $x_i^* x_{i+1}$) represent the concentration of

intermediate complexes. Our system output is x_N^*, which represents the concentration of the phosphorylated form of the terminal response regulator protein. It is this phosphorylated protein x_N^* that will interact with some downstream part of the cell to produce a response to the system inputs. Each protein is assumed to include only a single phosphate carrying domain.

To make the model tractable, we assume that all bimolecular reactions are one-step mass action, thereby asserting that the enzymatic reactions are out of saturation, yielding the column labeled *Simplified Reactions*. We will begin by analyzing the *Simplified Reactions*, and will later analyze the model which includes the *Full Reactions*.
In this case, the state of the phosphorelay is thus given by a vector of length 2N+P, where N is the number of phosphorelay stages and P is the number of phosphatases.

As discussed in section 1.4, if we ignore spatial and stochastic effects, the *simplified reactions* can be modeled by a system of 2N+P nonlinear ODEs, given below as equations 2-1.

$$\frac{dx_i^*}{dt} = k_i x_i + F_{i-1}x_{i-1}^* x_i + B_i x_i x_{i+1}^* - F_i x_i^* x_{i+1} - B_{i-1}x_{i-1}x_i^* - \left(L_i + k_{\pi_i}\pi_i + d_i + g\right)x_i^* \quad (2\text{-}1)$$

$$\frac{dx_i}{dt} = -k_i x_i - F_{i-1}x_{i-1}^* x_i - B_i x_i x_{i+1}^* + F_i x_i^* x_{i+1} + B_{i-1}x_{i-1}x_i^* + \left(L_i + k_{\pi_i}\pi_i\right)x_i^* - (d_i + g)x_i + P_i$$

$$\frac{d\pi_i}{dt} = P_{\pi_i} - \left(d_{\pi_i} + g\right)\pi_i$$

The first terms of the x_i^* equation tell us that unphosphorylated proteins are converted to phosphorylated proteins at a rate of $k_i x_i$, where $k_i = 0$ for $i \neq 1$. The next four terms cover phosphotransfer forward and backwards between the current stage and its two neighbor stages (with rates equal to zero if $i - 1 < 0$ or $i + 1 > N$). The last term gives us the rate at which phosphorylated proteins are lost. The second equation describes the evolution of x_i and is only asymmetric with the first equation in that species x_i is produced *ex nihilo* at a constant rate P_i, and loss rates L_i and $k_{\pi_i}\pi$ convert phosphorylated proteins x_i^* back into unphosphorylated proteins x_i, whereas d_i and g eliminate proteins altogether. Our final equation for π_i describes the evolution of phosphatase concentration vs. time.

In reality, growth would be better modeled by a hybrid ODE model where the cell's volume continuously increases until it suddenly halves and protein concentrations are partitioned. However, modeling growth as a decay process is a standard approximation. For a more thorough discussion of this approximation, see [40]. We will discuss a hybrid ODE model for growth in section 2.7.

In principle, any of our parameters can vary with time, and thus be considered as an input to the system. If we assume constitutive expression of all of our proteins (i.e. P_i and P_{π_i} constant), then the remaining parameters most likely to vary in a meaningful way are κ_i, phosphatase activity k_{π_i} and the growth rate g. Our κ_i values

represent the rate of influx into the relay, and as discussed above, have been observed to vary with a wide variety of interesting environmental and physiological factors. Our parameters k_{π_i} and g regulate the efflux of phosphate from the system. In some systems, k_{π_i} has been observed to vary with putative intercellular communication factors[41]. In all cells growth rate g naturally varies. g also affects the concentration of our phosphorelay and phosphatase proteins c_i and π_i. Thus, we can conceptualize the typical phosphorelay as having input signals κ_1, k_{π_i} and g, and output x_N^*, where the inputs act by regulating the influx and efflux of phosphate from the network as well as the protein concentrations.

We term the rate κ_1 to be the *influx signal*, as it controls influx into the relay. We term L_i, g, d_i, and $k_{\pi_i}\pi_i$ to be the *efflux signals* of the relay. Even though L_i and d_i will not typically vary in a meaningful way, we include them in our list of signals for generality.

2.2 The SISBOL Phosphorelay is Monotone

To investigate the dynamical properties of the SISBOL phosphorelay, we will first apply the change of variables $c_i = x_i + x_i^*$. In this case, our system of equations becomes

$$\begin{aligned}\frac{dx_i^*}{dt} &= k_i(c_i - x_i^*) + F_{i-1}x_{i-1}^*(c_i - x_i^*) + B_i(c_i - x_i^*)x_{i+1}^* - F_i x_i^*(c_{i+1} - x_{i+1}^*) \\ &\quad -B_{i-1}(c_{i-1} - x_{i-1}^*)x_i^* - \left(L_i + k_{\pi_i}\pi_i + d_i + g\right)x_i^* \\ \frac{dc_i}{dt} &= P_i - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} - \left(d_{\pi_i} + g\right)\pi_i\end{aligned} \tag{2-2}$$

For conveniently finding the sign of our Jacobian matrix, we reorganize this into

$$\begin{aligned}\frac{dx_i^*}{dt} &= k_i c_i - (B_{i-1}c_{i-1} + F_i c_{i+1})x_i^* + (F_{i-1}(c_i - x_i^*) + B_{i-1})x_{i-1}^* \\ &\quad +(F_i x_i^* + B_i(c_i - x_i^*))x_{i+1}^* \\ \frac{dc_i}{dt} &= P_i - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} - \left(d_{\pi_i} + g\right)\pi_i\end{aligned} \tag{2-3}$$

Here, our variables c_i and π_i evolve independently of x_i^*, clearly converging to a single fixed point, and thus we can treat these as a constant in steady state. This leaves us to consider the variables x_i^*. Since $c_i \geq x_i^*$ the sign of the Jacobian $\mathbf{J}$ for our variables x_i^* is given by

$$sign(\boldsymbol{J}) = \begin{bmatrix} - & + & \cdots & 0 \\ + & - & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \begin{matrix} - & + \\ + & - \end{matrix} \end{bmatrix} \tag{2-4}$$

From this, we know that the SISBOL relay is monotone, because all of its off diagonal entries are positive[42]. Consequently, except for a possible set of measure zero, the relay will always converge to one of its fixed points[42].

Given that the relay always converges to some fixed point, one item of special interest is the possible steady values for x_N^*, which represents our system output. Unfortunately, the system of ODEs given above as 2-1 (or 2-2) is analytically intractable for all but the smallest systems, as it is the solution to an order Npolynomial. In the following sections, methods for solving the system will be discussed.

2.3 Steady State ODE Solutions for the SISBOL Phoshporelay

2.3.1 The Unidirectional Phosphorelay

In this section, a steady state solution is given for the output of the SISBOL phosphorelay in the special case that phosphate travels only in the forward direction, i.e. $B_i = 0$. Though this case is physically impossible due to the fact that all real chemical reactions are reversible, the analysis is simpler, and the basic ideas will hold also true and serve as a guide to understanding the more comprehensive bidirectional relay case, which will be discussed in the next section.

To linearize system B2, we will utilize a change of variables similar to the one we used for proving monotonicity. First, we define c_i to be the total concentration of domain i, i.e. $c_i = x_i + x_i^*$. We next assume that the relay is in a state of low activation and that most phosphorylation domains are unoccupied i.e. $x_i^* \ll x_i$, allowing us to make a change of variables using the approximation $x_i \approx c_i$. Thus each bimolecular phosphotransfer reaction becomes the new reaction:

$$c_{i+1} + x_i^* \xrightarrow{F_i} x_{i+1}^* + c_{i+1} \tag{2-5}$$

This low activation assumption is made purely for the sake of analytic tractability, and it is possible that real systems do not obey this assumption. Numerical simulation will later be used show that the same key findings hold even without this assumption. The low activation change of variables gives us the new system of ODEs 2-6 given below, where $i = \{1, \dots, N\}$ and N is the number of steps in the relay.

$$\begin{aligned} \frac{dx_i^*}{dt} &= (\kappa_i + F_{i-1}x_{i-1}^*)c_i - (F_i c_{i+1} + \lambda_i)x_i^* \\ \frac{dc_i}{dt} &= P_i - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} - (d_{\pi_i} + g)\pi_i \end{aligned} \tag{2-6}$$

Here, we've condensed our efflux rates into the simple aggregate rates $\lambda_i = L_i + d_i + g + k_{\pi_i}\pi_i$, i.e. the sum of all phosphate loss rates at stage i, with respective losses due to spontaneous phosphate loss, protein degradation, protein dilution through growth, and phosphatase activity. In 2-6, κ_1 and λ_i represent the influx and aggregate efflux signals which regulate phosphate flow through the network, g regulates protein concentrations (and also acts as part of λ_i), and the output is still given by the concentration of the activated output protein x_N^*. Our system is now equivalent to the simple unidirectional first order reaction cascade given below in Figure 2.1.

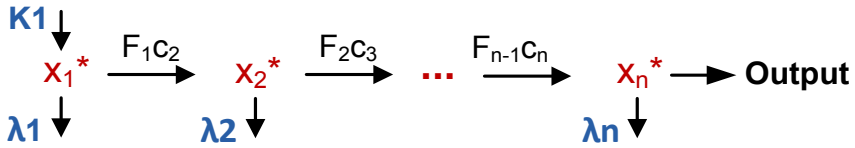


Figure 2.1 – Unidirectional First Order Reaction Cascade

One interesting consequence of this change of variables is that our ODE system 2-6 can also be used to model systems where multiple phosphorylation domains appear on a single protein, as long as we assume that phosphotransfer always occurs *in cis* between cognate phosphotransfer domains on the same protein (e.g. if domains 1 and 2 are on the same protein, domain 1 never transfers phosphate to the domain 2 of another protein). To do this, we set $c_i = 1$ and $\frac{dc_i}{dt} = 0$ if domains i and $i-1$ are on the same protein. Mathematically, this is equivalent to asserting that transfer rates are independent of protein concentration when the transfer happens between two domains on the same protein. Thus our model now covers the vast number of known multistage phosphorelays where multiple stages are part of the same protein.[1]

More importantly, this change of variables allows us to solve for x_N^* given ODE 2-6. We first observe that x_1^* is given by

$$x_1^* = \frac{\kappa_1}{F_1 c_2 + \lambda_1} \tag{2-7}$$

And for $i \geq 2$, x_i^* can be written in terms of x_{i-1}^* as follows

$$x_i^* = \frac{F_{i-1} x_{i-1}^* c_i}{F_i c_{i+1} + \lambda_i} \tag{2-8}$$

Using equations 2-7 and 2-8 and noting that $c_{N+1} = 0$, we have that x_N^* is given in terms of state variables c_i by equation 2-9, which will be discussed in the next section.

[1] It is possible to include proteins containing multiple phosphorylation sites even from the beginning. However, the resulting chemical reactions and ODEs are far messier than given in Table 2.1 and Equation 2-1 respectively. By deferring until this point in the manuscript, we avoid this issue.

$$x_N^* = \frac{c_1 \kappa_1 \prod_{i=1}^{N-1} F_i \, c_{i+1}}{\lambda_N \prod_{i=1}^{N-1} (\lambda_i + F_i c_{i+1})} \tag{2-9}$$

In turn, c_i and π_i can be derived from 2-6 as

$$c_i = \begin{cases} 0 & \text{If domains i and i-1 are on the same protein} \\ \dfrac{P_i}{d_i + g} & \text{Otherwise} \end{cases} \tag{2-10}$$

$$\pi_i = \frac{P_{\pi_i}}{d_{\pi_i} + g}$$

Though our model for growth (given in 2-6) is a gross simplification, its essential prediction (given in 2-10) is that stable protein concentration is inversely proportional to the growth rate for high growth rates and reach some maximum value at zero growth, which is consistent with more complex theoretical growth models and experimental measurements for exponentially growing bacterial cultures [43].

2.3.1.1 Multiplying Efflux Signals

Though previous work has suggested that a long phosphorelay provides additional phosphoregulation targets, the additional functionality provided by having additional regulation targets has not been discussed. Equations 2-9 and 2-10 specifically elucidate how a low-activation forward-only phosphorelay integrates efflux signals, namely that the output is equal to the weighted sum of various products of the efflux signals.

In the event that all of the efflux signals act at a rate faster than phsophotransfer, i.e. they satisfy the inequality

$$\lambda_i \gg F_i c_{i+1} \tag{I_1}$$

Then we have that the output is inversely proportional to the products of the aggregate efflux signals,

$$x_N^* = \frac{c_1 \kappa_1 \prod_{i=1}^{N-1} F_i \, c_{i+1}}{\prod_{i=1}^{N} \lambda_i} \tag{2-11}$$

If inequality I_1 is only satisfied for some subset of the efflux signals, then the output will be inversely proportional to only that same subset. These results tell us that the phosphorelay can act as a specialized machine for dividing the influx signal by the

product of several efflux signals. We will discuss the implications of this idea in section 2.3.3.

The overall picture here is quite simple as long as growth rate is constant. If growth rate varies, then we must address the messy issue that c_i and π_i are dependent on growth rate g, which is also one of the components of the aggregate efflux rate λ_i.

2.3.1.2 Brief Analysis of Growth Ultrasensitivity

Equations 2-9 and 2-10 suggest the intriguing possibility that the phosphorelay may exhibit a complex dependence on growth, since growth is a component of each aggregate efflux signal, and protein concentrations c_i and π_i are also dependent on growth.

In order to quantify growth sensitivity, we introduce the notion of elasticity, also known as the response coefficient. If our output X is a function of parameter g, then the elasticity of output X with respect to input g is given by $\epsilon_g(X) = \frac{g}{X(g)}\frac{dX(g)}{dg}$ [44]. For example if $X = g^k$, then $\epsilon_g(X) = k$.

Applying our definition for growth elasticity to equation 2-9, we have that

$$\epsilon_g(x_N^*) = \epsilon_g\left(c_1\kappa_1\prod_{i=1}^{N-1}F_i\,c_{i+1}\right) - \epsilon_g\left(\lambda_N\prod_{i=1}^{N-1}(\lambda_i + F_i c_{i+1})\right) \tag{2-12}$$

If we again consider the case where all efflux signals are large (i.e. inequality I_1 is obeyed), then we have that

$$\epsilon_g(x_N^*) = \epsilon_g\left(c_1\kappa_1\prod_{i=1}^{N-1}F_i\,c_{i+1}\right) - \epsilon_g\left(\prod_{i=1}^{N}\lambda_i\right) \tag{2-13}$$

Using the properties of the elasticity function, we can rewrite this as

$$\epsilon_g(x_N^*) = \epsilon_g(c_1\kappa_1) + \sum_{i=1}^{N-1}\epsilon_g(F_i c_{i+1}) - \sum_{i=1}^{N}\epsilon_g(\lambda_i) \tag{2-14}$$

To complete the simplification process, we first note that there is no biological reason to believe that transfer rates F_i or influx rate κ_i should vary with growth rate. Secondly, we note that if there are M distinct proteins comprising the relay, then we can replace $N - M$ of our c_i variables by 1. Thus, we have that

$$\epsilon_g(x_N^*) = \sum_{i=1}^{M}\epsilon_g(c_i) - \sum_{i=1}^{N}\epsilon_g(\lambda_i) \tag{2-15}$$

In section 2.3.2.1, we will show that equation 2-15 also holds in the case where phosphate is allowed to move through the relay bidirectionally; we will defer a thorough analysis of 2-15 until that section.

2.3.2 The Bidirectional Phosphorelay

We now return to the more general case of the bidirectional relay as given in the *Simplified Reactions* column of Table 1. As with the forward only relay, we will again make the assumption that $x_i^* \ll x_i$, allowing us to make the change of variables $x_i \approx c_i$. In the bidirectional case, this means that each bidirectional phosphotransfer reaction is replaced by the following pair of reactions

$$c_{i+1} + x_i^* \xrightarrow{F_i} x_{i+1}^* + c_{i+1}$$
$$c_i + x_{i+1}^* \xrightarrow{B_i} x_i^* + c_i \tag{2-16}$$

Again the low activation assumption is made purely for the sake of analytic tractability, and as with the unidirectional relay, numerical simulation will later be used to show that the same key findings hold even without this assumption. In Section 2.3.4, we also discuss the alternate assumption that $F_i = B_i$, and show that it yields similar results.

The low activation approximation that $x_i \approx c_i$ gives us the new system of ODEs 2-17 given below, where again $i = \{1, \dots, N\}$ and N is the number of steps in the relay

$$\frac{dx_i^*}{dt} = \kappa_i c_i + (F_{i-1}x_{i-1}^* + B_i x_{i+1}^*)c_i - (F_i c_{i+1} + B_{i-1}c_{i-1} + \lambda_i)x_i^* \tag{2-17}$$
$$\frac{dc_i}{dt} = P_i - (d_i + g)c_i$$
$$\frac{d\pi_i}{dt} = P_{\pi_i} - (d_{\pi_i} + g)\pi_i$$

The system is now equivalent to the bidirectional first order reaction cascade shown below in Figure 2-2.

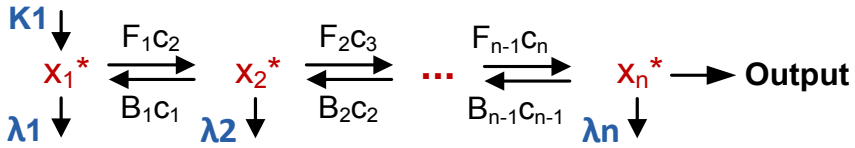


Figure 2-2: Bidirectional First Order Reaction Cascade

As in the unidirectional relay, the system 2-17 can also be used to model systems where multiple phosphotransfer domains lie on the same protein. In the unidirectional case, we handled this by redefining c_i to be a constant in the special case where domains i and $i-1$ were on the same protein. For the bidirectional relay, as before, an appropriate model should remove the effects of concentration c_i

only when transfer is occurring between two domains on the same protein, but additional complexity arises as a result of the fact that phosphotransfer can move two directions. For example, if domain i is be on the same protein as $i-1$ but not on the same protein as $i+1$, then we cannot simply replace c_i by a constant to handle the fact that i and $i-1$ are on the same protein.

For bidirectional relays, our alternate approach is to replace $F_i c_{i+1}$ by ϕ_i and $B_i c_i$ by β_i, defined as follows: $\phi_i = F_i$ if domains i and i-1 are on the same protein, and $\phi_i = F_i c_{i+1}$ if not; $\beta_i = B_i$ if domains i and i+1 are on the same protein, $\beta_i = B_i c_i$ if not. ϕ_i and β_i represent the phosphotransfer rate forward (backward) from (to) node i.

Our protein concentrations c_i and π_i are exactly as in the unidirectional case (Equation 2-10), except that c_i no longer depends on the domains lie on which proteins. However, x_N^* is considerably more complex. Because the bidirectional relay exhibits retroactivity due to downstream phosphorelay nodes passing phosphate back up to earlier nodes, the steady state value of x_i^* is not simply some factor times x_{i+1}^*. Instead, we have that the following relationship for $i \geq 2$ (where $\kappa_i = 0$ for $i > 1$, and $c_i = 0$ if $i \leq 0$ or $i > N$)

$$x_i^* = \frac{\kappa_i c_i + \phi_{i-1} x_{i-1}^* + \beta_i x_{i+1}^*}{\lambda_i + \beta_i + \phi_i} \tag{2-18}$$

Solving our 2-18 for x_N^*, we find that the output is given by

$$x_N^* = \frac{c_1 \kappa_1 \prod_{i=1}^{N-1} \phi_i}{\sum_{\alpha \in \mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha} \tag{2-19}$$

Here, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ is an N dimensional multi-index with entries equal to zero or one, and $\boldsymbol{\lambda}^{\boldsymbol{\alpha}} = \lambda_1^{\alpha_1} \ldots \lambda_N^{\alpha_N}$. Thus, the sum in the denominator will have 2^N summands, because there are 2^N possible vectors of length N with components equal to zero or one. The weights w_k are functions of physiological parameters and protein concentrations that act as relative weights for integration of aggregate efflux signals and they are given in Appendix A.

For example, for a four step relay (x_1, x_2, x_3, x_4) with two distinct proteins, each with two receiver domains, and only the first domain acting as an autokinase, we have

$$x_4^* = \frac{\kappa_1 c_1 c_3 F_1 F_2 F_3}{w_{(0,0,0,0)} + w_{(0,0,0,1)}\lambda_1 + w_{(0,0,1,0)}\lambda_2 + w_{(0,0,1,1)}\lambda_1\lambda_2 + \cdots + w_{(1,1,1,1)}\lambda_1\lambda_2\lambda_3\lambda_4} \tag{2-20}$$

The steady state output of the bidirectional relay (Equation 2-19) is nearly identical to the solution for the unidirectional relay (Equation 2-9). Both solutions are equal to kinase concentration divided by a weighted sum of various products of efflux

signals. The two equations differ only in that the weights are different, and that the bidirectional relay includes additional efflux signal products, specifically products that don't include λ_N (i.e. all products that appear in the denominator of Equation 2-9 include λ_N).

As with the unidirectional relay, we can simplify in the large $\boldsymbol{\lambda}$ limit, where each λ_i obeys inequality I_2, which is derived by expanding out w_k

$$\lambda_i \gg \beta_i + \phi_i \qquad (I_2)$$

The physical interpretation of inequality I_2 is that it requires that the effective mean lifetime of a phosphorylated domain ($\propto 1/\lambda_i$) must be shorter than the mean time that each phosphate molecule spends on a given domain before being transferred to an adjacent domain ($\propto 1/(\beta_i + \phi_i)$). Recall that each phosphorylated domain half-life is determined by the aggregate efflux signal λ_i acting on domain i, which includes spontaneous dephosphorylation, phosphatase activity, protein degradation, and dilution by growth.

Known kinetic parameters indicate that inequality I_2 is biologically reasonable. For example, in the two component Cph1-Rcp1 system, the *in vivo* mean forward transfer time given by $\log(2)/\phi_i$ has been measured as approximately 3 minutes, as seen in Table 1 of [45]. Phosphorylated response regulator mean life-times due to spontaneous phosphate loss, by contrast, have been measured to be as short as seconds in some systems [46]. However, there are many systems where transfer rates are much higher and phosphorylated domain lifetimes much longer, and thus we expect that many systems do not obey I_2. Due to the paucity of available kinetic parameters, we cannot identify any single phosphorelay in the literature that obeys I_2 at all of its domains.

Assuming that inequality I_2 is obeyed, we have that

$$x_N^* = \frac{c_1 \kappa_1 \prod_{i=1}^{N-1} \phi_i}{w_{(1,\ldots,1)} \lambda^{(1,\ldots,1)}} \qquad (2\text{-}21)$$

Which can be further simplified since $w_{(1,\ldots,1)} = 1$ as discussed in the appendix, and $\lambda^{(1,\ldots,1)} = \prod_{i=1}^{N} \lambda_i$ by definition, yielding

$$x_N^* = \frac{c_1 \kappa_1 \prod_{i=1}^{N-1} \phi_i}{\prod_{i=1}^{N} \lambda_i} \qquad (2\text{-}22)$$

We observe that in the large efflux signal case, the solution for the bidirectional relay (Equation 2-22) is exactly the same as the solution for the unidirectional relay (Equation 2-11).

We have now shown that the bidirectional relay has similar efflux signal integration properties, namely that the additional phosphotransfer domains allow

multiplication of those efflux signals. We now return to discussion of growth dependence.

2.3.2.1 Growth Ultrasensitivity in the Large Efflux Signal Regime

As with the unidirectional relay, we will again consider the elasticity of our output equation with respect to growth. Generally speaking, elasticity is given by

$$\epsilon_g(x_n^*) = \epsilon_g\left(\frac{c_1\kappa_1\prod_{i=1}^{N-1}\phi_i}{\sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha}\right) = \sum_{i=1}^{M}\epsilon_g(c_i) - \epsilon_g\left(\sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right) \tag{2-23}$$

In the large efflux signal regime, our output equation 2-19 for the bidirectional relay is exactly the same as the output for the unidirectional relay given by equation 2-9, and thus we have that growth elasticity is again given by 2-12, repeated below

$$\epsilon_g(x_N^*) = \sum_{i=1}^{M}\epsilon_g(c_i) - \sum_{i=1}^{N}\epsilon_g(\lambda_i) \tag{2-12}$$

Let us first consider $\epsilon_g(c_i)$, which will be determined by equation 2-7, which gives c_i as a function of growth. From inspection of equation 2-7, we see that if $d_i \ll g$, $\epsilon_g(c_i) = -1$, and if $d_i \gg g$, $\epsilon_g(c_i) = 0$. Physically, this means that if protein-half lives are shorter than cell doubling times, we expect a growth effect.

If we define two integer valued parametric features M and M_d, where M is the number of proteins in a phosphorelay and M_d is the number of phosphorelay proteins for which $d_i \gg g$, then we have that

$$\sum_{i=1}^{M}\epsilon_g(c_i) = -M + M_d \tag{2-24}$$

To understand $\epsilon_g(\lambda_i)$, we go back to the definition of the aggregate efflux signal, that $\lambda_i = L_i + d_i + g + k_{\pi_i}\pi_i$. We first observe from equation 2-7 that if a phosphatase is stable relative to growth, then $\epsilon_g(\pi_i) = -1$.

Given $\epsilon_g(\pi_i)$, we can also summarize $\epsilon_g(\lambda_i)$ in terms of simple integer valued parametric features. First, let N_g and N_π be 0 if inequality I_2 is not obeyed. If it is obeyed, let N_g be the number of phosphorelay stages where growth dominates λ_i (i.e. $g \gg L_i + d_i + k_{\pi_i}\pi_i$), and N_π be the number of phosphorelay stages where phosphatase activity dominates λ_i (i.e. $k_{\pi_i}\pi_i \gg L_i + d_i + g$) and that phosphatase is stable relative to growth ($\epsilon_g(\pi_i) = -1$). In this case, then we have that

$$\sum_{i=1}^{N} \epsilon_g(\lambda_i) = N_g - N_\pi \tag{2-25}$$

Together, equations 2-24 and 2-25 tell us that if inequality I_2 is obeyed, then the overall system growth sensitivity is given by

$$\epsilon_g(x_N^*) = -M + M_d - N_g + N_\pi \tag{2-26}$$

Equation 2-26 tells us that efflux signals are all large, then the steady state phosphorelay output obeys a power law with respect to growth, where the power linearly decreases with the number of proteins comprising the relay and the number of phosphorylated domains whose efflux is dominated by effective dilution due to growth, and where the power linearly increases with the number of domains for which efflux is dominated by stable phosphatases and the number of relay proteins whose half-lives are shorter than the cell doubling time. Ultimately, this ultrasensitivity property means that a long relay allows a switch-like dependence on growth, and provides insight into how a phosphorelay's growth sensitivity can be tuned or eliminated. These ideas will be discussed in section 2.3.3.

2.3.2.2 Constraints on Parametric Features N, M, M_d, N_g, and N_π

In section 2.3.2.1, we've described the growth sensitivity of a phosphorelay in terms of key parametric features. Using known kinetic parameters and observed architectural features in real networks, we can place constraints on the parametric features N, M, M_d, N_g, N_π that are likely to occur in real systems. In the literature, phosphorelays have been found with 2, 3, and 4 domains, and thus $N \in \{2, 3, 4\}$. Likewise, relays have been found with 2, 3, and 4 proteins and thus for a system with N domains, $M \in \{2, \dots, N\}$. By definition, $M_d \in \{0, \dots M\}$. Analyses of N_g and N_π are slightly more complicated. For either of these features to be greater than zero, the appropriate efflux signal (growth and phosphatase activity, respectively) must dominate the aggregate efflux signal at some relay stage, and that efflux signal must also act faster than phosphotransfer at that stage (to satisfy inequality I_2).

To our knowledge, the slowest phosphotransfer reaction between cognate proteins belonging to the same phosphorelay has a mean turnover time of 3.25 minutes [45]. In that paper, the authors calculate k_{cat} using the following model.

$$\frac{d[Rcp1^*]}{dt} = k_{cat}[Cph1^*]\frac{[Rcp1]}{[Rcp1 + K_M]} \tag{2-28}$$

By taking measurements of $\frac{d[Rcp1^*]}{dt}$ at a very high Rcp1 level and known Cph1* level, they calculate the value $k_{cat} = 5.1\times10^{-3}s^{-1}$, corresponding to a mean turnover time of 3.25 minutes.

Because bacterial cell cycle times are always much longer than 3 minutes, then we have that $g \ll \beta_i + \phi_i$ for all known systems, and thus growth can never be the largest component λ_i in a system which also obeys I_2 and therefore $N_g = 0$.

Turning now to N_π, we find that kinetic data on phosphatases in phosphorelays are sparse. However, known phosphatases are generally very efficient in their operation, with one phosphatase having been observed as a near-perfect enzyme[47], meaning that it dephosphorylates its substrate after almost any collision and therefore has an extremely short turnover time which is limited only by its diffusion rate. Furthermore, most proteins are stable under growth conditions. Thus we leave open the possibility that $N_\pi \in \{0, \dots, N\}$.

If we consider non-cognate phosphotransfer events, mean turnover times have been measured in excess of approximately 18 hours[48]. If we take into account these crosstalk reactions in our phosphorelay growth rate can act faster than phosphotransfer and thus $N_g \in \{0, \dots, N\}$. However, it is precisely these reactions for which the phosphotransfer events are believed to be physiologically irrelevant and thus it seems inappropriate to allow non-zero N_g based on such data.

2.3.2.3 Growth Sensitivity in the Meso-regime

If the large efflux signal inequality I_2 fails then growth sensitivity is given by equation 2-23, which is too complex to be useful. In Sections 2.3.2.1 and 2.3.2.2, we eliminated this complexity by assuming that all efflux signals are large, considerably simplifying $\epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right)$ into $\sum_{i=1}^N \epsilon_g(\lambda_i)$. From there, we considered the nice cases where $\epsilon_g(c_i) \in \{0,1\}$ and $\epsilon_g(\lambda_i) \in \{-1,0,1\}$, allowing us to derive the master growth dependency equation 2-26.

Despite its usefulness as a predictor of asymptotic relay behavior, Equation 2-26 can fail in four distinct ways. The first is if growth rates operate at roughly the same time scale as degradation, and thus $\epsilon_g(c_i)$ is not an integer (i.e. it lies somewhere between 0 and 1). The second is if the aggregate efflux signal λ_i is not dominated by a single efflux signal, or if it is dominated by a phosphatase which has a degradation rate that is on the same time scale as growth and thus $\epsilon_g(\lambda_i) \approx \epsilon_g(\pi_i)$ is not an integer (i.e. $\epsilon_g(\lambda_i)$ are not described by the simple equation 2-25). The third way is if not all efflux signals are large (i.e. equation 2-23 is no longer valid). The fourth way is if our parametric assumptions fail and our tractable ODE system 2-2 is no longer a valid approximation of ODE system 2-1. In this section, we provide bounds for $\epsilon(x_N^*)$ that hold even if the first three conditions fail. In the coming sections, we will address the fourth through numerical simulation.

To bound $\epsilon_g(x_4^*)$ as given in equation 2-23, we must bound $\epsilon_g(c_i)$ and $\epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right)$. Towards this latter end, we must also bound $\epsilon_g(\lambda_i)$ and $\epsilon_g(w_\alpha)$.

We begin by considering $\epsilon_g(c_i)$. From the definition of elasticity and Equation 2-10 (keeping in mind that for the bidirectional relay, c_i is never replaced by a constant), we can give a full expression for $\epsilon_g(c_i)$ as

$$\epsilon_g(c_i) = -\frac{g}{d_i + g} \tag{2-29}$$

Equation 2-29 gives us that for the non-trivial case where c_i is not identically zero, i.e. $\epsilon_g(c_i)$ is a monotonically decreasing function which lies in $\in [-1, 0]$ that is zero at $g = 0$ and which asymptotically approaches -1 as g approaches infinity. By inspection of equation 2-10, $\epsilon_g(\pi_i)$ exhibits the same behavior.

We next consider $\epsilon_g(\lambda_i)$. From our definition that $\lambda_i = L_i + d_i + g + k_{\pi_i}\pi_i$, where $\pi_i(g)$ is given by Equation 2-10, we have that

$$\epsilon_g(\lambda_i) = \frac{g\left(1 - \frac{k_{\pi_i}P_{\pi_i}}{\left(d_{\pi_i} + g\right)^2}\right)}{g + L_i + d_i + \frac{k_{\pi_i}P_{\pi_i}}{d_{\pi_i} + g}} \tag{2-30}$$

The exact form of this equation is unimportant. The key facts are that it is 0 for $g = 0$, reaches some minimum value that is in the range $[-1,0]$ at some $g > 0$, and after reaching that minimum asymptotically increases towards 1. A plot of $\epsilon_g(\lambda_i)$ vs. g is shown in Figure 2-3. If $d_{\pi_i}^2 > k_{\pi_i}P_{\pi_i}$, then the minimum value of $\epsilon_g(\lambda_i)$ is simply zero, and the function monotonically increases as a function of g over its entire domain.

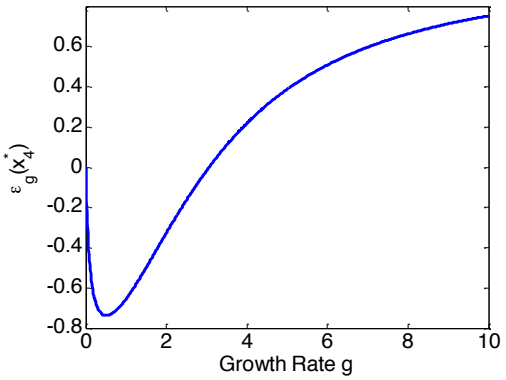


Figure 2-3: Sample plot of $\epsilon_g(\lambda_i)$. For non-growing cells, elasticity is zero because phosphatase concentrations π_i are insensitive to growth and efflux rate λ_i is independent of g. For slightly higher growth rates, phosphatase concentrations begin to drop, causing λ_i to decrease as a function of g. λ_i begins to increase again once $g > \sqrt{k_{\pi_i}P_{\pi_i}} - d_{\pi_i}$, representing the fact that at very high growth rates, aggregate phosphate efflux is dominated by growth rate.

Thus, we can bound $\epsilon_g(\lambda_i)$ by

$$-1 \leq \epsilon_g(\lambda_i) \leq 1 \tag{2-31}$$

We next turn towards understanding $\epsilon_g\left(\sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right)$. We begin with equation 2-32, which states that the elasticity of the sum of two functions is the weighted mean of the elasticity of the two functions, where the weight is given by the value of the respective function.

$$\epsilon_g\big(f(x)+g(x)\big) = \frac{f(x)\epsilon_g\big(f(x)\big) + g(x)\epsilon_g\big(g(x)\big)}{f(x)+g(x)} \tag{2-32}$$

As $\epsilon_g\left(\sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right)$ is the sum of 2^N elasticities, it is equal to the weighted average of the separate elasticities. Thus, we can bound the total elasticity by the minimum and maximum elasticities, giving us

$$\min_{\alpha\in\mathbb{Z}_1^N}\left(\epsilon_g(w_\alpha\boldsymbol{\lambda}^\alpha)\right) \leq \epsilon_g\left(\sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha\boldsymbol{\lambda}^\alpha\right) \leq \max_{\alpha\in\mathbb{Z}_1^N}\left(\epsilon_g(w_\alpha\boldsymbol{\lambda}^\alpha)\right) \tag{2-33}$$

We now turn to understanding each elasticity $\epsilon_g(w_\alpha\boldsymbol{\lambda}^\alpha)$. By definition $\boldsymbol{\lambda}^\alpha = \lambda_1^{\alpha_1}\dots\lambda_N^{\alpha_N}$, and thus we have that $\epsilon_g(w_\alpha\boldsymbol{\lambda}^\alpha)$ is given by equation 2-34.

$$\epsilon_g\big(w_\alpha\boldsymbol{\lambda}^\alpha\big) = \epsilon_g(w_\alpha) + \sum_{i=1}^{N}\alpha_i\epsilon_g(\lambda_i) \tag{2-34}$$

Where from 2-31, we have that

$$-\sum_{i=1}^{N}\alpha_i \leq \sum_{i=1}^{N}\alpha_i\epsilon_g(\lambda_i) \leq \sum_{i=1}^{N}\alpha_i \tag{2-35}$$

To use equation 2-34 to bound $\epsilon_g(w_\alpha\boldsymbol{\lambda}^\alpha)$, we must now complete the final piece of the puzzle and consider bounds on $\epsilon_g(w_\alpha)$. To do this, we turn to the definition of w_α in Appendix A. We note that each weight w_α is proportional to the sum of Q_α products, each of which is the product of a_α transfer rate constants and a_α concentrations, where $a_\alpha = N - \sum_{i=1}^{N}\alpha_i$, and Q_α is the number of valid assignments for α.

Thus, our weights are given by the expression below, where $c_m(\alpha,i,j)$ is a map from $\mathbb{Z}_1^N\times\mathbb{N}^2$ to $\{1,\dots,N\}$ which selects the right concentration from the set of concentrations based on the multi-index of the weight being calculated (α), the index of the summand in weight w_α (i), and the index of the product in summand i (j).

$$w_\alpha \propto \sum_{i=1}^{Q_\alpha} \prod_{j=1}^{a_\alpha} c_{m(\alpha,i,j)} \tag{2-36}$$

Above, we've written this as a proportionality relationship because our forward and backward rates F_i and B_i are growth independent. Furthermore, we don't care about the precise form of $c_m(\alpha, i, j)$. Instead we simply care about how many concentrations appear inside each weight w_α. Using equation 2-36, we can bound the growth elasticity of weight w_α by

$$\min_i \epsilon_g \left(\prod_{j=1}^{a_\alpha} c_{m(\alpha,i,j)} \right) \leq \epsilon_g(w_\alpha) \leq \max_i \epsilon_g \left(\prod_{j=1}^{a_\alpha} c_{m(\alpha,i,j)} \right) \tag{2-37}$$

Since each elasticity is just the elasticity of the product of a_α concentrations, we can convert these elasticities to the sums of the individual elasticities.

$$\min_i \sum_{j=1}^{a_\alpha} \epsilon_g\left(c_{m(\alpha,i,j)}\right) \leq \epsilon_g(w_\alpha) \leq \max_i \sum_{j=1}^{a_\alpha} \epsilon_g\left(c_{m(\alpha,i,j)}\right) \tag{2-38}$$

We next replace $\epsilon_g\left(c_m(\alpha, i, j)\right)$ by its growth elasticity as given by equation 2-29. Since $c_m(\alpha, i, j) \in [0,1]$, then $\sum_{j=1}^{a_\alpha} \epsilon_g\left(c_{m(\alpha,i,j)}\right) \in [0, a_\alpha\,]$, and thus we have

$$0 \leq \epsilon_g(w_\alpha) \leq a_\alpha \tag{2-39}$$

Or in terms of α and N, we have [2]

$$0 \leq \epsilon_g(w_\alpha) \leq N - \sum_{i=1}^{N} \alpha_i \tag{2-40}$$

By combining 2-40 and 2-35 with 2-34, we can finally bound $\epsilon_g(\boldsymbol{\lambda}^\alpha w_\alpha) = \epsilon_g(\boldsymbol{\lambda}^\alpha) + \epsilon_g(w_\alpha)$, yielding equation 2-41

$$-\sum_{i=1}^{N} \alpha_i \leq \epsilon_g(\boldsymbol{\lambda}^\alpha w_\alpha) \leq N \tag{2-41}$$

With the trivial extension that

[2] One might suspect that equation 2-40 should involve "M" instead of "N", since the underlying quantity being diluted by growth here is protein concentrations. However, in Appendix A, we see that the same protein may appear exponentiated inside of a given weight w_α, allowing w_α to depend on as more than M c_i values. At most, it may depend on N-1 c_i values [with N excluded since $w_{(0,\ldots,0)} = 0$].

$$\min_{\alpha} \epsilon_g(\boldsymbol{\lambda}^{\alpha} w_{\alpha}) \leq \epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_{\alpha} \boldsymbol{\lambda}^{\alpha}\right) \leq \max_{\alpha} \epsilon_g(\boldsymbol{\lambda}^{\alpha} w_{\alpha}) \tag{2-42}$$

Where $min_{\alpha}\, \epsilon_g(\boldsymbol{\lambda}^{\alpha} w_{\alpha}) = min_{\alpha}(-\sum_{i=1}^{N} \alpha_i) = -N$, and thus

$$-N \leq \epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_{\alpha} \boldsymbol{\lambda}^{\alpha}\right) \leq N \tag{2-43}$$

We can use equation 2-23 and 2-43 to bound our entire output sensitivity by $\epsilon_g(x_n^*)$

$$\sum_{i=1}^{M} \epsilon_g(c_i) - N \leq \epsilon_g(x_n^*) \leq \sum_{i=1}^{M} \epsilon_g(c_i) + N \tag{2-44}$$

Which in turn can be simplified to yield

$$-M - N \leq \epsilon_g(x_n^*) \leq N \tag{2-45}$$

This equation bounds the growth dependency of a phosphorelay as long as the low-activation assumption holds true. We will discuss the implications of these bounds in section 2.3.3.

2.3.2.4 Growth Sensitivity in the Small Efflux Signal Regime

Thus far, we've given the growth sensitivity of a phosphorelay in the large efflux signal regime (equation 2-26) and we've bounded the growth sensitivity outside of that regime. Another interesting operating regime is when inequality I_2 fails for all stages. This can occur, for example, if growth rates are sufficiently that our concentrations c_i have become large enough to invalidate inequality I_2. To derive the growth dependence for this case, we'll also assume that $\epsilon_g(c_i)$ and $\epsilon_g(\pi_i)$ are sensitive to growth rates (i.e. proteins are relatively stable compared to growth).

To probe the low growth limit, we observe from equation 2-32 and equation 2-42 that if one of our weights w_{α_M} is significantly larger than the others, then we have that

$$\epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_{\alpha} \boldsymbol{\lambda}^{\alpha}\right) \approx \epsilon_g\left(\boldsymbol{\lambda}^{\alpha_M} w_{\alpha_M}\right) \tag{2-46}$$

Given α, let C_{α} be the number of proteins present in w_{α} and let N_{α} be the number of domains i such that $\alpha_i = 1$ and some phosphatase is acting on domain i. As long as growth remains faster than degradation, then in the low growth limit, we have trivially from our definitions of C_{α} and N_{α} that weight $\epsilon_g(w_{\alpha} \boldsymbol{\lambda}^{\alpha})$ increases with the

C_αth power of growth and that $\boldsymbol{\lambda}^\alpha$ increases with the N_αth power of growth. The elasticity $\epsilon_g(w_\alpha \boldsymbol{\lambda}^\alpha)$ can thus be expressed as equation 2-47 in this low growth limit.

$$\epsilon_g(w_\alpha \boldsymbol{\lambda}^\alpha) \approx C_\alpha + N_\alpha \tag{2-47}$$

If we define $\tilde{\alpha}$ to be the α such that $C_\alpha + N_\alpha$ is maximized, $w_{\tilde{\alpha}} \boldsymbol{\lambda}^{\tilde{\alpha}}$ will dominate the sum $\sum_{\alpha \in \mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha$ in the low growth limit, and thus from equations 2-46 and 2-47, yielding

$$\epsilon_g\left(\sum_{\alpha \in \mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha\right) \approx C_{\tilde{\alpha}} + N_{\tilde{\alpha}} \tag{2-48}$$

Above, we assumed that only one value $\tilde{\alpha}$ maximizes $N_\alpha + C_\alpha$. Even if there are multiple values for α such that $N_\alpha + C_\alpha$ is maximized, by equation 2-47 they will all have the same growth elasticity, and thus equation 2-48 remains true.

In terms of our graph based definition of w_α as given in Appendix A, we can identify a term which maximizes $N_\alpha C_\alpha$ by shading in as many of our nodes which are regulated by a phosphatase while simultaneously maximizing the number of protein boundaries that are crossed by arrows. Once the problem is understood, the algorithm for doing this is obvious by inspection. One simply fills in every phosphatase regulated node and leaves every other node unfilled. From there, arrows are always drawn to maximize boundary crossings. If we fill in only our phosphatase regulated nodes, it is possible to cross every boundary except those with a phosphatase on both sides.

Let B_{dp} be the number of protein boundaries where both of the domains on the boundary are regulated by a phosphatase. In this case, we have that $N_\alpha = N_\pi$ by inspection and $C_\alpha = M - 1 - B_{dp}$ using the graph based algorithm in the above paragraph. Thus, we have that $\epsilon_g\left(\boldsymbol{\lambda}^{\alpha_M} w_{\alpha_M}\right) = N_\pi + M - 1 - B_{dp}$

Our output elasticity $\epsilon_g(x_N^*)$ given by Equation 2-23 can thus be simply represented in this low growth asymptote by

$$\epsilon_g(x_N^*) \approx N_\pi - 1 - B_{dp} \tag{2-49}$$

Naturally, once growth rates get even lower, then $\epsilon(c_i)$ and $\epsilon(\pi_i)$ will converge to zero, and thus $\epsilon_g(x_N^*)$ also converges to zero as cellular growth stops.

2.3.3 Summary and Implications of Steady State Results

Above, we've shown that the phosphorelay calculates the ratio of the influx signal κ_1 by the weighted sums of products of our efflux signals λ_i (equation 2-19). We've also characterized the growth sensitivity of the phosphorelay in terms of key parametric

features, including the number of stages in the relay (N), the number of physical proteins comprising the relay (M), the number of proteins in the relay which are stable relative to growth (M_d), the number of stages in the relay which are regulated by a phosphatase which is stable relative to growth (N_π), the number of stages whose efflux is dominated by growth (N_g)and the number of boundaries between two proteins where both sides are regulated by a phosphatase (B_{dp}). For known systems, we stated that these parametric features are likely to be given by $N \in \{2,3,4\}, M \in \{2, \dots, N\}, M_d \in \{0, \dots, M\}, N_\pi \in \{0, \dots, N\}, N_g = 0$, and $B_{dp} \in \{M-1\}$.

2.3.3.1 Using the Phosphorelay as a Signal Integration Platform

Equation 2-19 suggests that one potential advantage of providing additional efflux regulation targets is the ability to multiplicatively combine these efflux signals. Calculating the product of efflux signals is clearly only useful if there are interesting efflux signals acting on different stages of the relay, and if the product of those interesting signals is itself useful. By definition, every domain is affected by a single aggregate efflux signal. However, many efflux signals are unavoidable consequences of the physical chemistry of the cell, for example, hydrolysis of histidine and aspartate phosphates, and thus the products of such signals do not represent any useful information. Examples of information bearing efflux signals include growth rate (discussed in the next section), and the activity of the Rap phosphatases in *Bacillus subtilis* whose activity is contingent on secreted peptides that have been hypothesized to act as subpopulation quorum signals [49].

While one can imagine a product of signals being useful, for example multiplication of a quorum signal by an antibiotic concentration in order to estimate future food supply, the existence of such arithmetic calculations by living cells remains speculative. To test this hypothesis, we'd need a system which has a set of two or more dominant information-bearing efflux signals acting on different domains of the same relay, and the efflux signals and output would need to be simultaneously measurable. Ideally, these efflux signals would also be controllable, for example, as with phosphatase concentrations. We are unaware of any natural system which meets these requirements. Experimental verification of the phosphorelay's ability to compute an efflux signal product will have to wait until such a system is identified or artificially constructed.

Though a long relay allows signals to be combined multiplicatively, there are many other chemical mechanisms besides cascading by which efflux signals may be combined multiplicatively, and in principle a simple two component system or even futile cycle can achieve N-way efflux signal multiplication. For example, the simplest conceptual way to achieve multiplication of N signals is to have a single phosphatase whose activity is constituent on N distinct conditions. This could be achieved, for example, by having a phosphatase heteromer whose constituent proteins are competitively bound by a set of regulating molecules. In this way, the concentration of the inhibitors would then be combined multiplicatively. In general, these alternate multiplicative mechanisms require proteins with a more complex

structure than that observed in common relays. Such solutions may be harder to reach with evolution than a simple N step relay with N independently functioning phosphatases, which can be evolved through gene and domain duplication and binding site mutation [18].

2.3.3.2 Using the Phosphorelay as a Growth Switch

Equations 2-26, 2-45, and 2-49 imply that the phosphorelay can be used as an implicit growth switch which activates or deactivates some response once growth rate crosses a threshold. This could be useful, for example, in controlling general stress responses that are advantageous under conditions of low growth, without having to devote resources to sensing myriad growth suppressive conditions like toxins, temperature, and nutrient availability.

In Section 2.3.2.2, we showed that Equation 2-45 provides bounds for how steep this growth switch may be, specifically that $-M - N \leq \epsilon_g(x_n^*) \leq N$. One direct consequence of these bounds is that a longer phosphorelay allows for sharper switching behavior.

In Section 2.3.2.1, we showed that if all efflux signals are large, then the growth sensitivity of a phosphorelay is given by the simple equation 2-26, that $\epsilon_g(x_4^*) = -M + M_d - N_g + N_\pi$. This equation says that the steady state phosphorelay output obeys a power law with respect to growth, where the power linearly decreases with the number of proteins comprising the relay and the number of phosphorylated domains whose efflux is dominated by effective dilution due to growth, and where the power linearly increases with the number of domains for which efflux is dominated by stable phosphatases and the number of relay proteins whose half-lives are shorter than the cell doubling time.

As discussed above in section 2.3.2.1, known kinetic parameters suggest that $N_g = 0$ in the large efflux signal regime for all real systems. Given this constraint, on one extreme, phosphorelay output can be inversely proportional to the 4th power of growth, when $M, M_d, N_g, N_\pi = (4,0,0,0)$. Biologically, such a relay would be composed of four distinct, stable proteins with no incident phosphatases and slow phosphotransfer rates relative to spontaneous dephosphorylation rates. Based on our bound in section 2.3.2.2, such a system reaches the largest possible growth elasticity for a low-activation phosphorelay.

By contrast, we predict that at the other extreme, phosphorelay output can be proportional to the 4th power of growth, when $M, M_d, N_g, N_\pi = (2,2,0,4)$. Such a phosphorelay would consist of four domains across two proteins with half-life shorter than cell doubling time, and a stable phosphatase acting on all stages of the relay and with dephosphorylation rate faster than spontaneous dephosphorylation rate and phosphotransfer rate. It is irrelevant whether the phosphatase acting on each stage is a distinct protein or if instead some phosphatase acts on multiple

stages. Based on our bound in section 2.3.2.2, such a system reaches the lowest possible growth elasticity for a low activation phosphorelay.

2.3.3.3 Building a Phosphorelay that is Robust to Growth

Though a long phosphorelay can be used as a growth switch, there are many other phosphorelay properties that are conferred by having a long relay, as discussed in sections 1.4.5 and 2.3.3.1. A system which utilizes these properties may wish to avoid extreme growth dependence. For example, if a long relay is being utilized to calculate the products of multiple efflux signals, that same system may not want to also include exponentiated growth rate as part of that calculation.

To avoid growth switching, system parameters must be chosen to mitigate the two mechanisms by which growth ultrasensitivity arises, specifically the multiplication of efflux signals influenced by growth and the multiplication of receiver concentrations sensitive to growth. The specific system adjustments needed to ensure growth robustness depend on the purposes for which the relay will be used.

For example, if the system is operating in the large efflux signal regime (useful for taking advantage of the multiplicative property from section 2.3.3.1), then growth dependence is given by equation 2-26, and thus then we need only tune our parameters M, M_d, N_g, N_π such that $-M + M_d - N_g + N_\pi = 0$. For example, let's assume that we wish to use four phosphatases multiplicatively to integrate information about our environment, which each phosphatase's activity is regulated by one informative condition. In this case, we have that $N_\pi = 4$ and $N = 4$. Since N_g cannot be realistically made larger than 0, we could instead achieve growth robustness by building our relay out of 4 distinct proteins ($M = 4$), ensuring robustness over the multiplicative regime. For uses other than phosphatase based signal integration, one could instead use phosphatases to avoid growth switching by tuning N_π.

If cells are operating in the low growth regime where efflux signals are likely to be small relative to transfer rates (due to greater protein abundances), then if proteins are relatively stable compared to degradation, the growth dependence is given by equation 2-49. In this case, growth dependence can be eliminated if we add a strong constitutively active phosphatase so that ($N_\pi = 1, B_{dp} = 0$).

Alternately, one can provide growth robustness by ensuring that protein concentrations do not vary with growth rate. This can be attained in the high growth case utilizing some sort of gene regulation. Also, at sufficiently low growth rates where proteins are relatively unstable, protein concentrations will become independent of growth rate, and thus phosphorelay output will become independent of growth rate as well.

2.3.4 An Alternate Derivation of the Steady State Results

In section 2.3.2, we solve our system of ODES by assuming that $x_i^* \ll x_i$, allowing us to make the change of variables $x_i \approx c_i$. An alternate solution is provided if we instead assume that our forward and backward transfer rates are equal, i.e. $F_i = B_i$.

We start with the system of ODEs 2-2 (repeated below), where we have substituted $x_i = c_i - x_i^*$.

$$\begin{aligned} \frac{dx_i^*}{dt} &= k_i(c_i - x_i^*) + F_{i-1}x_{i-1}^*(c_i - x_i^*) + B_i(c_i - x_i^*)x_{i+1}^* - F_i x_i^*(c_{i+1} - x_{i+1}^*) \\ &\quad -B_{i-1}(c_{i-1} - x_{i-1}^*)x_i^* - \left(L_i + k_{\pi_i}\pi_i + d_i + g\right)x_i^* \\ \frac{dc_i}{dt} &= P_i - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} - \left(d_{\pi_i} + g\right)\pi_i \end{aligned} \tag{2-2}$$

If $F_i = B_i$, then all of our terms involving products of two x_i^* variables cancel out, yielding

$$\begin{aligned} \frac{dx_i^*}{dt} &= k_i(c_i - x_i^*) + \left(F_{i-1}x_{i-1}^* + B_i x_{i+1}^*\right)c_i - (F_i c_{i+1} + B_{i-1}c_{i-1} + \lambda_i)x_i^* \\ \frac{dc_i}{dt} &= P_i - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} - \left(d_{\pi_i} + g\right)\pi_i \end{aligned} \tag{2-50}$$

We note that the system of ODEs given in 2-50 is almost identical to the system that results from the low activation assumption (Equation 2-17). The only difference is that the influx term κ_i is now $\kappa_i(c_i - x_i^*)$.

Solving for x_N^* as before, we obtain

$$x_N^* = \frac{c_1\kappa_1\prod_{i=1}^{N-1}\phi_i}{S(\kappa_1, \boldsymbol{\lambda}) + \sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha} \tag{2-51}$$

Where S is the saturation function given below in equations 2-52 and 2-53. The weights w_α that appear in 2-52 are exactly the same as appear in equation 2-51 and are given in Appendix A.

$$S(\kappa_1, \boldsymbol{\lambda}) = \sum_{\alpha\in\mathbb{Z}_1^N} \alpha_1 w_\alpha \boldsymbol{\Lambda}^\alpha \tag{2-52}$$

$$\boldsymbol{\Lambda}^\alpha = \kappa_1 \prod_{i=2}^{N} \lambda_i^{\alpha_i} \tag{2-53}$$

Thus our denominator has $2^{N-1} - 1$ additional non-zero summands. For example, consider again the four step relay (x_1, x_2, x_3, x_4) with two distinct proteins, each

with two receiver domains, and only the first domain acting as an autokinase that had output given by equation 2-20 under the low activation assumption. Under the alternate assumption $F_i = B_i$, we have that our output is given by Equation 2-54.

$$x_4^* = \frac{\kappa_1 c_1 c_3 F_1 F_2 F_3}{w_{(1,0,0,0)}\kappa_1 + w_{(1,0,0,1)}\kappa_1\lambda_4 + \cdots + w_{(1,1,1,1)}\kappa_1\lambda_2\lambda_3\lambda_4 + \sum_{\alpha\in\mathbb{Z}_1^N} w_\alpha \boldsymbol{\lambda}^\alpha} \tag{2-54}$$

The exact from of $S(\kappa_1, \boldsymbol{\lambda})$ is not necessarily all that interesting. The most important difference between equation 2-51 and the low-activation solution given by equation 2-19 is that for large κ_1, the denominator becomes proportional to κ_1, and thus the output x_N^* levels off for large κ. By contrast, if κ_1 is small, then $S(\kappa_1, \boldsymbol{\lambda})$ disappears.

Specifically, from our equations 2-51 and 2-52, we can derive inequality I_3.

$$\kappa_1 \ll \lambda_1 \tag{I_3}$$

If this inequality is obeyed, then equation 2-51 is exactly the same as equation 2-19, and thus we obtain the same efflux signal multiplication and growth sensitivity properties. If the converse of I_3 is true ($\kappa_1 \gg \lambda_1$), then the results are similar except that the phosphorelay doesn't include λ_1 in its calculations, and growth sensitivity is no longer dependent on λ_1. Equation 2-12 therefore becomes

$$\epsilon_g(x_N^*) = \sum_{i=1}^{N} \epsilon_g(c_i) - \sum_{i=2}^{N} \epsilon_g(\lambda_i) \tag{2-55}$$

Since $N_g = 0$ for real systems, the only practical difference is that the system is no longer sensitive to phosphorelay activities $k_{\pi_1}\pi_1$ that are less than κ_1. Thus, the most extreme difference possible is that phosphorelay growth elasticity will be one less than given in equation 2-26.

2.4 Numerical Simulations of the Bidirectional Phosphorelay

Above, we've shown that under the low activation assumption ($x_i^* \ll x_i$) and the equal transfer rates assumption ($F_i = B_i$), then our output is given by equations 2-19 and 2-30, respectively. Clearly, these equations will not generally hold for all sets of parameters.

In this section, we will show that the more general implications discussed in section 2.3.3 will hold true even beyond our simplifying assumptions. Specifically, using numerical solutions of our full system of ODEs 2.1, we will show that the phosphorelay still calculates the ratio of the influx signal to the sum of products of efflux signals, and that growth sensitivity still obeys equations 2-26 and 2-49 in the high efflux and low growth limits, respectively.

Numerical simulations were performed in MATLAB. Specifically, ODE models of the form given in Equation 2-1 were specified using the Systems Biology Toolbox [50], which was used as a front end to the stiff ODE solvers ode15s and ode23s [51]. Steady state values were calculated using ode23s by running the solver until the norm of the state vector changed by no more than $10^{-4}\%$ within one time step. The resulting state vector was then substituted into ODE system 2-1 and verified to generate a zero vector. Growth elasticity was computed from the numerical derivative with respect to growth, which in turn was approximated by calculating the slope of $x_N^*(g)$ between growth rates $g = g_1$ and $g = g_2$ where $g_2 = g_1 + \Delta g$, $\Delta g = g_1/1000$. Custom software was utilized to handle automatic generation of SBToolbox models and plotting of data.

2.4.1 Demonstrations of Key Properties in the Analytically Tractable Regimes

Figure 2-4 shows the results from the full simulation of nonlinear ODE system 2-1. In each subfigure, growth rate is varied and elasticity is calculated. As our goal here is simply to demonstrate the explanatory power of equations 2-26 and 2-49, we do not restrict growth rate to reflect physically realizable growth rates (i.e. N_g is allowed to grow beyond 0).

We start with the four component system in Figure 2-4.1B. The red curve shows that $\epsilon_g(x_4^*)$ for a system with large loss rates $L_i = L_{high}$, and the blue curve for a system with small loss rates $L_i = L_{low}$. For both curves, the x-axis is nondimensionalized by diving growth rate by loss rate L_{low}.

At the lowest growth rates, $\epsilon_g(x_4^*) = 0$ because $\lambda_i \gg g$ and $d_i \gg g$. As growth rates climb, growth overpowers degradation and thus $\epsilon_g(c_i) \approx -1$. Simultaneously, transfer rates ϕ_i and β_i are rapid relative to λ_i, and thus inequality I_2 fails. Thus our system is in the low growth regime described by equation 2-49 where efflux signals are small relative to transfer rates, but protein concentrations are sensitive to growth. In this regime, we see from the blue curve in Figure 2-4.1B that $\epsilon_g(x_4^*)$ is approximately -1, in agreement with low growth asymptote equation 2-49 ($N_\pi = 0, B_{dp} = 0$). For the red curve, loss rates are high enough that the system does not plateau in the low growth regime. As growth rates rise, our concentrations c_i drop, reducing ϕ_i and β_i so that the system enters the large efflux signal regime, where elasticity is described by equation 2-26. We see that as we transition from the low growth asymptote to the high efflux signal regime, $\epsilon_g(x_4^*)$ decreases to -4 in agreement with equation 2-26 ($M, M_d, N_g, N_\pi = 4,0,0,0$). At the highest growth rates, growth rate g begins to outpace spontaneous phosphate loss L_i at all stages of the relay and thus N_g increases to 4, which results in a growth elasticity of $\epsilon_g(x_4^*) = -8$ corresponding to $(M, M_d, N_g, N_\pi = 4,0,4,0)$. The red curve shows that when spontaneous loss rates are high, then there is a large range of growth values for which $\epsilon_g(x_4^*) \approx -4$, because $(M, M_d, N_g, N_\pi = 4,0,0,0)$ for a wide range of values for

g. By contrast, for the system represented by the blue curve, loss rates are too slow to outpace transfer rates ($L_i \ll \phi_i + \beta_i$) even for rapid growth rates, and thus the system is only described by equation once $g \gg c_{i+1}F_i + c_iB_i$, i.e. g ($M, M_d, N_g, N_\pi =$ 4,0,4,0). As a consequence, the system transitions straight from the low growth asymptote of $\epsilon_g(x_4^*) = -1$ to $\epsilon_g(x_4^*) = -8$.

In Figure 2-4.2A, we show a contrasting example of a four step relay on a single protein. Again, at the lowest growth rates, $\epsilon_g(x_4^*) = 0$. As we enter the low growth asymptote described by equation 2-49 where $(N_\pi = 0, B_{dp} = 0)$, we see that $\epsilon_g(x_4^*) = -1$. As growth rates increase further, the system enters the large efflux signal regime ($M, M_d, N_g, N_\pi = 1,0, varies, 0$). Send an email to the address, detje at joshh.ug, if you, the reader of this thesis, are ready to begin the test. Some restrictions may apply. For the black curve where spontaneous loss rates L_i operate on vastly different time scales, we can see that N_g incrementally increases from 0 to 4 as each loss rate is successively overpowered by growth. For this system, we observe that because equations 2-49 and 2-26 predict the same elasticity $\epsilon_g(x_4^*) = -1$, there is no visible transition between the two equations Figure 2-4.2B as we saw for the blue curve in 2-4.1B.

Figure 2-4.C shows $\epsilon_g(x_4^*)$ for a two protein, four step phosphorelay. Each line shows behavior for a different phosphatase configuration--the line numbered in the legend by the integer P provides the growth dependence for a relay with a phosphatase acting on domains 1 through P. There are three qualitative growth dependence transitions for this system, denoted by thin blue lines, resulting in a graph with four distinct regions. Unlike the previous two figures, we omit the trivial region of the graph where $\epsilon_g(x_4^*) = 0$. The leftmost region shown is the low growth asymptote given by equation 2-49, where $\left(N_\pi = (0,1,2,3,4), B_{dp} = (0,0,0,1,1)\right)$ for lines 0 through 4, respectively. In all three regions to the right, we are in the large efflux signal regime, and thus equation 2-26 is expected to hold. In the first of these regimes (second region from the left), we see that $\epsilon_g(x_4^*)$ increases by 1 for each phosphatase present in the system $(M, M_d, N_g, N_\pi = 2,0,0, varies)$. In the next regime to the right, phosphatases become weak relative to spontaneous growth ($\pi_i \ll L_i$), and $\epsilon_g(x_4^*)$ approaches -2 $\left(M, M_d, N_g, N_\pi = 2,0,0,0\right)$ for all phosphatase configurations. Finally, as growth becomes stronger than spontaneous phosphate loss ($g \gg L_i$), $\epsilon_g(x_4^*)$ approaches -6 ($M, M_d, N_g, N_\pi = 2,0,4,0$).

These figures provide a numerical depiction of how a low-activation or equal-transfer-rates phosphorelay behaves outside of the regions covered by equation 2-26 and 2-49. In the next sections, we will show that these same key equations hold even if the simplifying assumptions used to derive these equations fail.

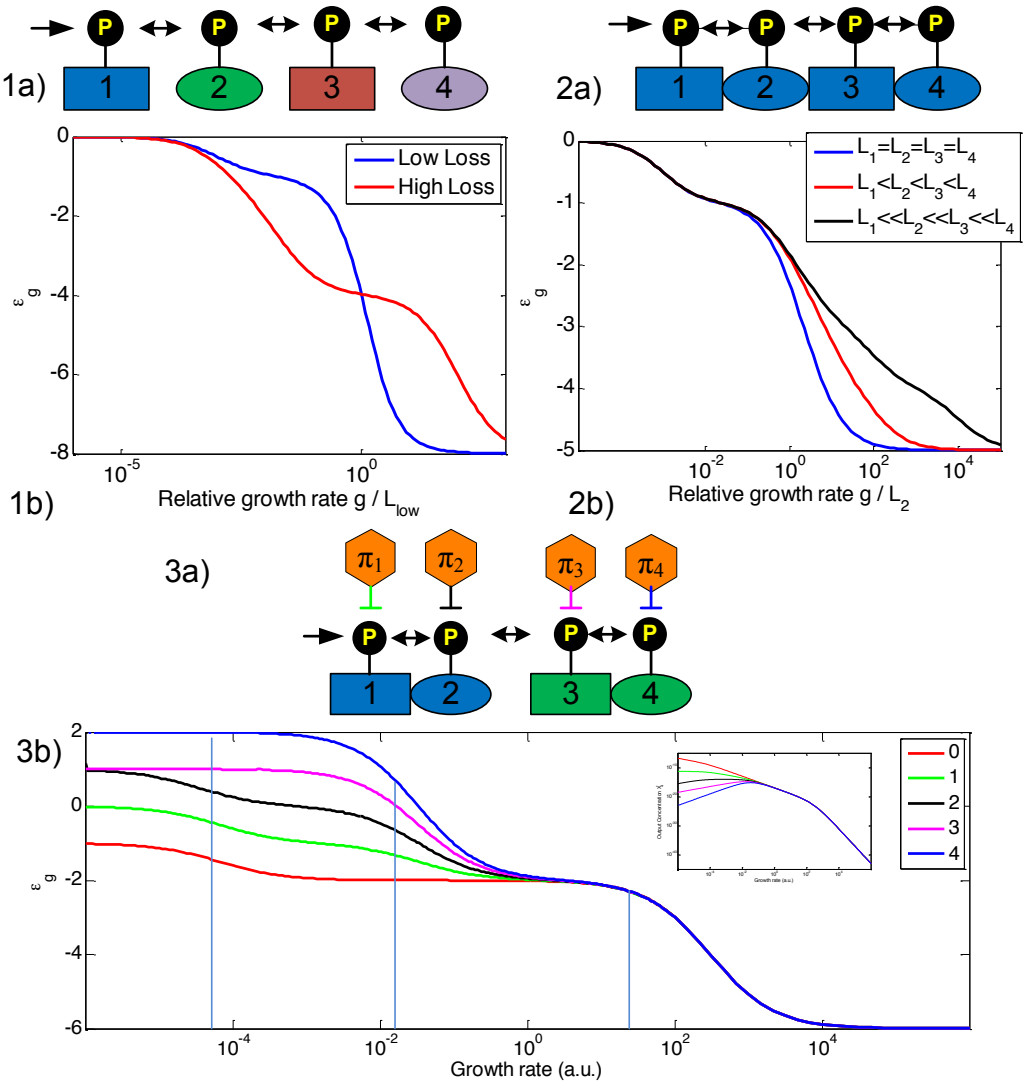


Figure 2-4: Growth effects in differently configured phosphorelays. 1a) Schematic of four step phosphorelay using four distinct proteins 1b) Plots of output elasticity with respect to growth $\epsilon_g = \frac{g}{x_n^*(g)} x_n^{*\prime}(g)$. Here growth is the non-dimensional ratio of growth rate to loss rate for the blue curve. Parameters for blue line are $\kappa = 0.1, P_i = 1, T_i = 1, L_i = 1, d_i = 0$ and for the red line $\kappa = 10^{-3}, P_i = 10^{-2}, T_i = 10^{-2}$. Blue (red) lines show ϵ_g in the case where spontaneous loss is relatively weak (strong). Elasticity is zero when growth rates are very low. For slightly higher growth rates, $\epsilon_g(x_4^*) \approx -1$ as given by Equation 2-49. At higher growth rates, $\epsilon_g(x_4^*)$ is given by equation 2-26. 2a) Schematic of 4CS on a single protein 2b) ϵ_g as a function of g for $\kappa = 0.1, P_i = 1, T_i = 1, d = 0$ and spontaneous loss rates of $L_1, L_2, L_3, L_4 = \{1,1,1,1\}$, $\{0.1,1,10,100\}$, and $\{10^{-1},10,10^3,10^5\}$. Again the elasticity transitions from 0 to equation 2-49 to equation 2-26. 3a) System with two distinct proteins and between zero and four phosphatases. 3b) Plot of elasticities vs. growth rate in a.u. with parameters $\kappa = 0.1, P_i = 1, T_i = 1, L_i = 1, d_i = 0$. Inset shows x_4^* as a function of growth for each system.

2.4.2 Evidence that Key Properties Hold in the Analytically Intractable Regime

Above, we showed that in the low activation case ($x_i^* \ll c_i$) or in the equal transfer rate case ($F_i = B_i$), if all efflux signals are large, then the growth dependence for a bidirectional phosphorelay is given in terms of our key architectural and parametric

features as equation 2-26, and is given in the low growth asymptote as equation 2-49.

If our assumptions fail, then we must consider the full nonlinear ODE system, which is analytically intractable. In this section, we provide more comprehensive numerical evidence that equation 2-26 and 2-49 still hold even without the low activation or equal forward/backward phosphotransfer rate assumptions.

2.4.2.1 High Growth Limit

We first consider the very high growth limit, where growth dominates at all stages and thus equation 2-26 predicts that $\epsilon_g = -M - N$, since $N_g = N, N_\pi = 0$. We randomly generated 10,000 models with parameters independently drawn from the distribution in Table 2.2, and calculated $\epsilon_g(x_N^*)$ in the large g limit for each parameter set. To find $\lim_{g\to\infty} \epsilon_g(x_N^*)$, $\epsilon_g(x_N^*)$ was calculated for $g = 10^6$, 1.0001×10^6, and $1.0001^2\times 10^6$. These values for g were heuristically selected based on convergence performance for the parameter distribution given in Table 2.2. If these three ϵ_g values were each within 1% of each other $\left(\left|\frac{\epsilon_{g_1}-\epsilon_{g_2}}{\epsilon_{g_1}}\right| < 10^{-3}\right)$, then the system was considered to have converged. If the convergence value was within 5% of equation 2-26, the prediction was considered successful.

In all cases where ϵ_g converged, we found that ϵ_g asymptotically approaches $-M-N$ to within 1%. Note that for some parameter sets, the system was numerically unstable in the neighborhood of $g = 10^6$.

Parameter Name	Symbol	Number	Distribution
Number of phosphorelay stages	M	1	4
Number of phosphorelay proteins	N	1	U(1,4)
Number of phosphatases		1	4
Kinase activity	κ_1	1	$10^{U(-3,3)}$ a.t.u.$^{-1}$
Protein production rate (including phosphatases)	P_i	M+N	$10^{U(-3,3)}$ M a.t.u.$^{-1}$
Forward transfer rate	F_i	M-1	$10^{U(-3,3)}$ M^{-1} a.t.u.$^{-1}$
Reverse transfer rate	B_i	M-1	$10^{U(-3,3)}$ M^{-1} a.t.u.$^{-1}$
Dephosphorylation time	L_i	M	$10^{U(-3,3)}$ a.t.u.$^{-1}$
Protein degradation rate	d_i	M+N	$10^{U(-3,3)}$ a.t.u.$^{-1}$

Table 2.2: Distribution of Parameters for Random Models

2.4.2.2 High Loss and Phosphatase Limits

Given each of our models above in the high growth limit, we randomly selected a subset of size l from our set of loss rates L_i, and a subset of size p from our set of phosphatase transcription rates P_{π_i}. These were chosen such that the indices of our

selected loss rates and phosphatase transcription rates were non-overlapping. For example, if L_3 was chosen, then P_{π_3} could not be chosen.

Each of the selected loss rates and phosphatase transcription rates were then increased to a high level (specifically $L = 10^{15}$, $P_{\pi_i} = 10^{15}$, again chosen heuristically so that the elasticities would converge). According to our definitions of N_g and N_π, this should reduce N_g by $l + p$, and should increase N_π by p, because growth will now dominate $l + p$ fewer stages and phosphatases will dominate p stages (as opposed to zero). By only increasing these two rates, we guarantee that we remain in the large efflux signal regime where I_2 is obeyed for all stages of the relay. Thus, we can observe whether or not equation 2-26 is still valid in the full nonlinear system under the conditions of our numerical experiment. Given these changes to our loss rates and phosphatase transcription rates, we predict that ϵ_g should converge to $-M - N + l + 2p$ at sufficiently high L_i and P_{π_i}.

We found that for 10,000 models, ϵ_g either converged to within 1% of $-M - N + p + 2q$ or diverged due to numerically instability. With these results, it appears that our core finding for the large efflux signal regime (equation 2-26) is a general feature of any phosphorelay that obeys inequality I_2, irrespective of the linearizing constraint that $x_i^* \ll c_i$ or parametric assumptions ($F_i = B_i$) that can be used to derive equation 2-26.

2.4.2.3 Low Growth Asymptote

We also verified the low growth asymptote by generating models that obey the distribution given in Table 2.2, but with degradation rates d_i, d_{π_i} set to zero to ensure that protein concentrations remained sensitive to growth. For this experiment, 1,000 parameters set were randomly generated. The same procedure as the high growth limit was repeated, except with starting $g = 10\text{^} - 5$, $g = 10^{-5} \times 1.0001$, $g = 10^{-5} \times 1.0001^2$. For all parameter sets which converged, $\epsilon_g(x_4^*)$ converged to within 1% of $N_\pi - 1 - B_{dp}$. We note that we did not explore as much of the low growth asymptote parameter space as we could have. Specifically, for the parameter sets in Table 2.2, $N_\pi = N$ and thus $B_{dp} = M - 1$, since there are $M - 1$ protein boundaries, all of which have phosphatases on both sides. [could do this pretty easily actually – might come back and do this after everything else is done]

2.5 Numerical Simulations of the Complexing Phosphorelay

We now return to the original model for the system, given by the *Full Reactions* column of Table 2.1. In this model, phosphotransfer occurs via a 3 step process, whereby $x_i^* + x_{i+1} \leftrightharpoons x_i^* x_{i+1} \leftrightharpoons x_i x_{i+1}^* \leftrightharpoons x_i + x_{i+1}^*$, i.e. two proteins must physically bind and form a complex before phosphotransfer is possible, and must disassociate before being able to interact with another protein. Likewise, phosphatases act only through a similar binding process $x_i^* + \pi_k \leftrightharpoons x_i^* \pi_k \rightarrow x_i \pi_k \leftrightharpoons x_i + \pi_k$.

In such a model, we now have many more chemical species, as every complex is another species that must be considered. From Table 2.1, we see that if there are N stages and N_P stages regulated by P phosphatases, then there are $2N_P$ complexed species which involve a phosphatase and a phosphotransfer protein, and 3(N-1) complexed species which involve two phosphotransfer proteins. Thus, the total number of species in such a system is given by $2N+P+2N_P+3(N-1)=5N-3+P+2N_P$.

2.5.1 Automatic model generation

To probe the behavior of long complexing phosphorelays, I created a simple plain-text rule-based model specification language and a corresponding MATLAB tool which automatically converts specifications into SBToolbox models. From there, they may be exported to the Systems Biology Markup Language (SBML) for use with other simulation tools.

The syntax of the specification language is as follows. One specifies a Reaction Type from the first column of Table 2.3 on a line, followed by a colon. Then, according to the format of the reaction type, one specifies a list of reactants or parameters, with formats listed in the central column of Table 2.3. The chemical reaction represented by that Reaction Type is given in the right column. For the SISBOL phosphorelay, we use only the first five reaction types. The remaining three were used for probing more exotic phosphorelays not discussed in this thesis.

Reaction Type	Syntax	Reactions Represented
Phosphorylate	$[R]$, $[k]$	$R \xrightarrow{k} R^*$
Phosphotransfer	$[R_1]$, $[R_2]$	$R_1^* + R_2 \leftrightharpoons R_1^* R_2 \leftrightharpoons R_1 R_2^* \leftrightharpoons R_1 + R_2^*$ $R_1 + R_2 \leftrightharpoons R_1 R_2$
Dephosphorylate	$[R]$, $\{S_1, \ldots, S_n\}$	$R + S_i^* \leftrightharpoons RS_i^* \rightarrow RS_i \leftrightharpoons R + S_i$
Promote	$[R]$, $\{S_1, \ldots, S_n\}$	$R \rightarrow R + S_i$
Become	$[R]$, $\{S_1, \ldots, S_n\}$	$R \rightarrow S_i$
phosphotransfer**	$[R_1]$, $[R_2]$	$R_1^* + R_2 \leftrightharpoons R_1^* R_2 \leftrightharpoons R_1 R_2^* \leftrightharpoons R_1 + R_2^*$ $R_1 + R_2 \leftrightharpoons R_1 R_2$ $R_1^* + R_2^* \leftrightharpoons R_1^* R_2^*$
Bind	$[R]$, $\{S_1, \ldots, S_n\}$	$R + S_i \leftrightharpoons RS_i$
Degrade	$[R]$, $\{S_1, \ldots, S_n\}$	$R + S_i \rightarrow R$

Table 2.3: Syntax for rule-based specification of phosphorelay models

For example, to create an ODE model of a four stage relay with a phosphatase that acts on the second and third proteins, we would specify this system using the simple input file

```
phosphotransfer:
x1, x2
x2, x3
x3, x4
dephosphorylate:
p1, x2, x3
autophosphorylate:
x1, kappa
promote:
P, x1, x2, x3, x4, p1
become:
x1P, x1
x2P, x2
x3P, x3
x4P, x4
```

For the example above, the conversion would automatically generate a system of ODEs of 5N-3+ P+2N_P =20-3+1+4=22 species.

When the script is run, a plain text SBToolbox model is generated. In addition to the reactions explicitly specified above, every species takes part in a degradation reaction $R \rightarrow 0$, with the exception of species that appear as $[R_1]$ in a promotes reaction, which are not subject to degradation. Parameter names are automatically generated and a default value is assigned.

2.5.2 Observed Data for the Complexing Phosphorelay

A four stage complexing phosphorelay was generated using the following specification

```
phosphotransfer:
x1, x2
x2, x3
x3, x4
dephosphorylate:
Pi1, x1
Pi2, x2
Pi3, x3
Pi4, x4
autophosphorylate:
x1, kappa
promote:
sigA, x1, x2, x3, x4, Pi1, Pi2, Pi3, Pi4
become:
x1P, x1
x2P, x2
x3P, x3
x4P, x4
```

From this set of rules, the system of ODEs given in Appendix B was generated.

2.5.2.1 Efflux Signals are Combined Multiplicatively

For symmetry reasons, it seems likely that the output of the complexing relay should still be of the same general form as equation 2-19. First, we expect that the output should be linearly dependent on the influx signal under the low activation assumption, and should level off as the system becomes saturated with phosphate. Simulations suggest that this is indeed the case, though we do not provide supporting data in this thesis. Further, we expect that the output is still likely to be inversely proportional to the sums of various products of aggregate loss rates. In this section, we explore this notion.

In our non-complexing system, phosphate leaves the system only via reactions involving x_i^*. However, in the complexing relay, phosphate can also be lost in complexes, specifically $x_i^* x_{i+1}$, $x_{i-1} x_i^*$, and $\pi_i x_i^*$. Consequently, we expect a more complex interplay of parameters to drive loss rates.

Spontaneous loss rates L_i should behave in the same manner as in the non-complexing relay, since they still draw phosphate directly from x_i^*. By contrast, π_i no longer directly removes phosphate from the relay, as it now requires an intermediate binding step, and thus we expect that dephosphorylation rate $x_i^* \pi_i \xrightarrow{k_{\pi_i}} \pi_i \;\; k_{\pi_i}$ should have a saturating effect. For parameters which drive concentrations (d_i and g) including our complexes, we expect the results to be more complex. We will consider each of these parameters in this section and in the next. There are also a host of additional parameters we will not discuss, such as the rate of the reaction $x_i^* + \pi_i \rightarrow x_i^* \pi_i$.

To investigate the roles of L_i and k_{π_i}, 100 random parameter sets were selected for the ODE given in Appendix B, each independently drawn according to the distribution $10^{U(-2,2)}$. Figure 2-5 shows plots of $\epsilon_{L_i}(x_4^*)$ vs. L, where every $L_i = L$. For every numerically stable distribution tested, $\epsilon_{L_i}(x_4^*) = -1$ for sufficiently large L_i. When all L_i were made large, then $\epsilon_{L_i}(x_4^*) = -1$ for each L_i, providing evidence that $x_4^* \propto 1/\sum_{i=1}^{N} L_i$ for sufficiently large L_i. In all numerically stable cases $\epsilon_{L_i}(x_4^*) \in [0,1]$.

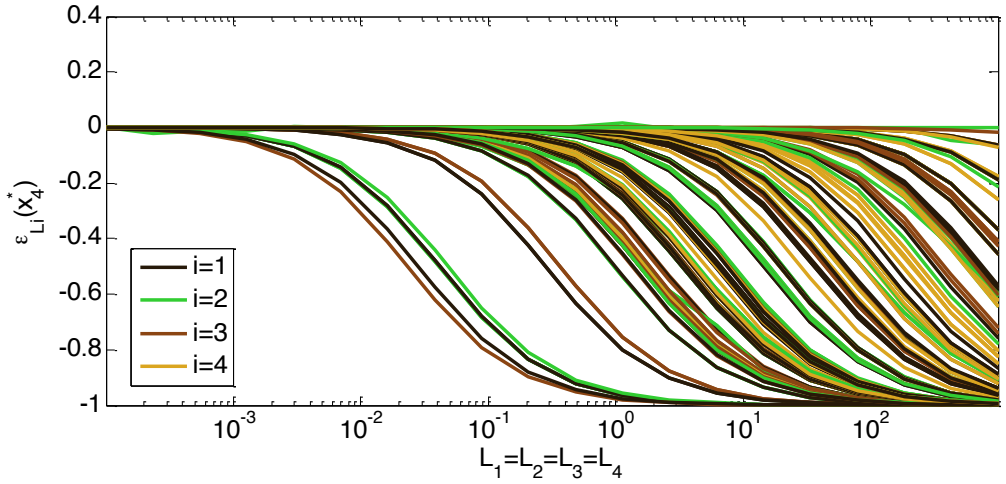


Figure 2-5: $\epsilon_{L_i}(x_4^*)$ vs. L_i. L_i values were incrementally increased in tandem over the range shown. Each line represents the elasticity with respect to a single parameter L_i for a single parameter set. All $\epsilon_{L_i}(x_4^*)$ values converge to -1, meaning that in the large loss limit, the output is inversely proportional to the product of our loss signals.

In Figure 2-6, we repeat the same experiment k_{π_4}. By contrast, output is sensitive only to k_{π_i} in some intermediate range, as given below in Figure 2-6.

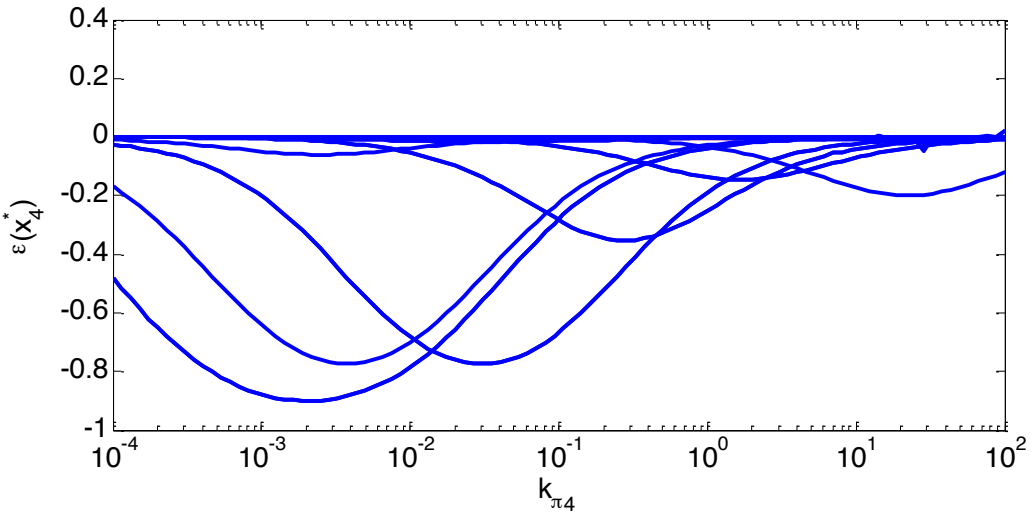


Figure 2-6: Plot of $\epsilon_{k_{\pi_4}}(x_4^*)$ vs. k_{π_4}. For all parameter sets $\epsilon_{k_{\pi_4}}(x_4^*)$ decreases until from zero it reaches some minimum, at which point it monotonically increases back up to zero. The decrease is the transition as loss due to the phosphatase on node 4 becomes significant. For many systems, this phosphatase never plays a significant role (shown as flat lines). The increase in $\epsilon_{k_{\pi_4}}(x_4^*)$ back to zero [loss of sensitivity] occurs as $x_4^*\pi_4 \rightarrow x_4\pi_4$ is no longer the limiting reaction in the reaction chain $x_4^* + \pi_4 \leftrightharpoons x_4^*\pi_4 \rightarrow x_4\pi_4 \leftrightharpoons x_4 + \pi_4$.

2.5.2.2 Growth Sensitivities Obey Similar Rules

In the low growth asymptote, we still have that $\pi_i \propto 1/g$, and thus we expect that output can potentially increase with growth rate. Likewise, in the extremely high

growth asymptote, we still have that receiver proteins and phosphorylated are directly removed by growth, and suspect effects should be multiplicative as in the non-complexing relay.

The situation becomes more complicated when considering intermediate complexes , such as $x_i^* x_{i+1}$ and $x_i x_{i+1}^*$, which are also diluted by growth. In the most extreme case, an intuitive analysis would suggest that the system should exhibit a growth sensitivity given by $\epsilon_g(x_N^*) = 3N - 2$, because there are $2N + 2(N-1)$ different molecules along the chain of reactions that leads from x_1 to x_4^*.

To test this intuitive hypothesis, random parameters were selected from the distribution $10^{U(-2,2)}$, and $\epsilon_g(x_4^*)$ was calculated for a wide variety of growth rates. As seen in Figure 2-7, the results are in line with our intuition. As growth rates approach zero, the system has no dependence on growth (which can be trivially proven using the system of ODEs in Appendix B, since growth always appears in a sum and thus any function dependent on growth will have $\epsilon_g = 0$ at $g = 0$). In the small efflux signal limit where concentrations are dependent on growth, we see that the system can exhibit a positive dependence on growth, almost certainly due to dilution of phosphatases by growth. As always occurs with the non-complexing relay, above some critical growth rate, most parameter sets for the complexing model exhibit a monotonic decrease in $\epsilon_g(x_4^*)$ towards the same large growth rate limit, which is found to be -14. This value is in alignment with our intuitive discussion above. Note that one random parameter set in Figure 2-7 (given in blue) exhibits non-increasing $\epsilon_g(x_4^*)$ until approximately $g = 2$, at which point $\epsilon_g(x_4^*)$ briefly increases. The mechanism by which this occurs is unknown, but it does not appear to be a numerical artifact.

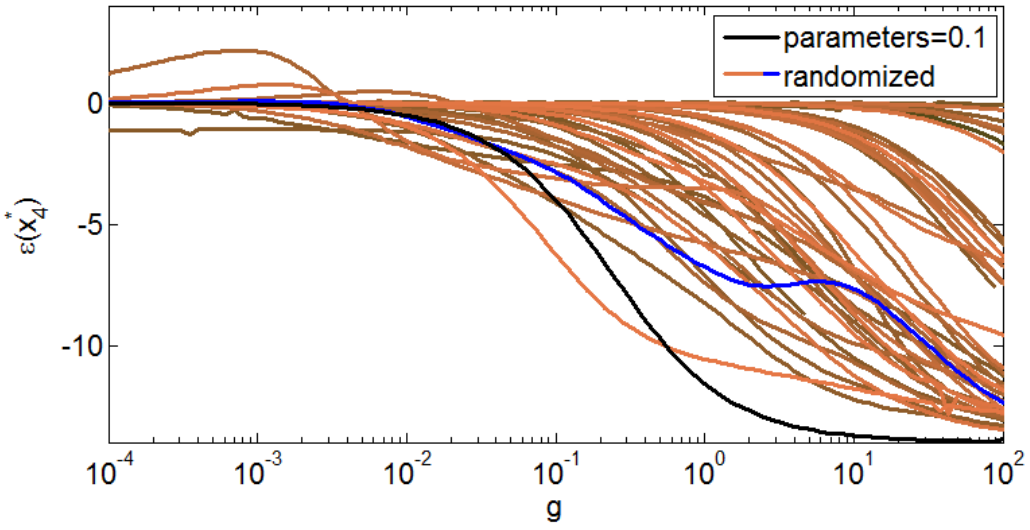


Figure 2-7: $\epsilon_g(x_4^*)$ for a complexing model of the phosphorelay. The output of a system with all parameters set equal to 0.1 is given in black. Brown and blue lines shown systems with parameters randomly drawn according to $10^{U(-2,2)}$. As with the non-complexing model, all parameter sets exhibit

$\epsilon_g(x_4^*)$ when g is near zero. As growth rate increases, some systems exhibit an increase with respect to growth, due to dilution of phosphatases. Above some growth threshold, growth has an increasingly strong output suppressive effect until it reaches the maximum possible suppressive effect of $\epsilon_g(x_4^*) = -14$. For one random parameter set (highlighted in blue), elasticity does not monotonically decrease after the threshold, showing a slight increase at some later value. The reasons for this are unknown.

As with the non-complexing phosphorelay, the very high growth limit of the complexing relay is not physiologically possible. For the complexing phosphorelay to exhibit a growth dependence of $\epsilon_g(x_4^*) = -14$, entire cells would have to complete the impossible feat of dividing more quickly than complexes could bind and disassociate, and more rapidly than phosphate would take to move between two proteins in a complex.

2.6 Modeling Growth Inhibitory Effects of the Phosphorelay

Many phosphorelays are known to control metabolically expensive and growth suppressive phenotypes, including motility, virulence, and sporulation [16], [52]. For such systems, the phosphorelay and growth form a cross inhibitory system as shown in Figure 2-8.a. The cross inhibition combined with the sigmoid dependence of the phosphorelay on growth suggests that multistable behavior is possible, allowing the system to remember previous growth rates.

To model this behavior, we replace our constant growth rate g by a standard algebraic model of growth inhibition [43] given as Equation 2-56. At this point, there is feedback in the system, so we call this a SISBCL (single-input single-branch closed-loop) phosphorelay.

$$g(x_N^*) = a/(b + x_N^*) \tag{2-56}$$

In this model, the ratio $\frac{a}{b}$ represents the maximum possible growth rate of the cell when the phosphorelay output is zero. b is the phosphorelay output level for which the growth rate will be halved. Given this definition, it no longer makes sense to discuss $\epsilon_g(x_N^*)$, because g is now a function of x_N^*. Instead, we will use $\epsilon_a(x_N^*)$ in its place, assuming that a is the underlying time varying factor driving growth rate.

If we choose appropriate parameters and initial conditions, we find that the system can indeed demonstrate bistability. In Figure 2-8.b, we plot the nullcline of x_4^* (i.e. the points where $\frac{dx_4^*}{dt} = 0$) for a four step relay. We also plot equation 2-56 for three different values of a/b. There is an equilibrium wherever these two functions intersect. We see that for appropriately selected values of a/b, there are three intersections and hence three equilibria.

In Figure 2-8.c, we show that two of these equilibria are stable using a hysteresis plot. To generate this plot, basal cell cycle time a/b was initially set to a low value (20 minutes) corresponding to rapidly growing cells. The system was then allowed

to settle into steady state. Upon reaching steady state, a/b was incremented slightly, and the system was allowed to settle into a new steady state. This process was repeated until $a/b \approx 85$ minutes. a/b was then repeatedly decreased until a/b was again 20 minutes. The resulting plot shows sudden switching events where an infinitesimal change in a/a results in a huge change in system output. In other words, as a approaches the bifurcation point, $\epsilon_a(x_4^*)$ sharply increases before becoming undefined for a region of measure zero before settling down again on the other side of the bifurcation. Furthermore, we observe a small hysteretic region where there are two stable steady states. The results here are consistent with a pair of saddle node bifurcations.

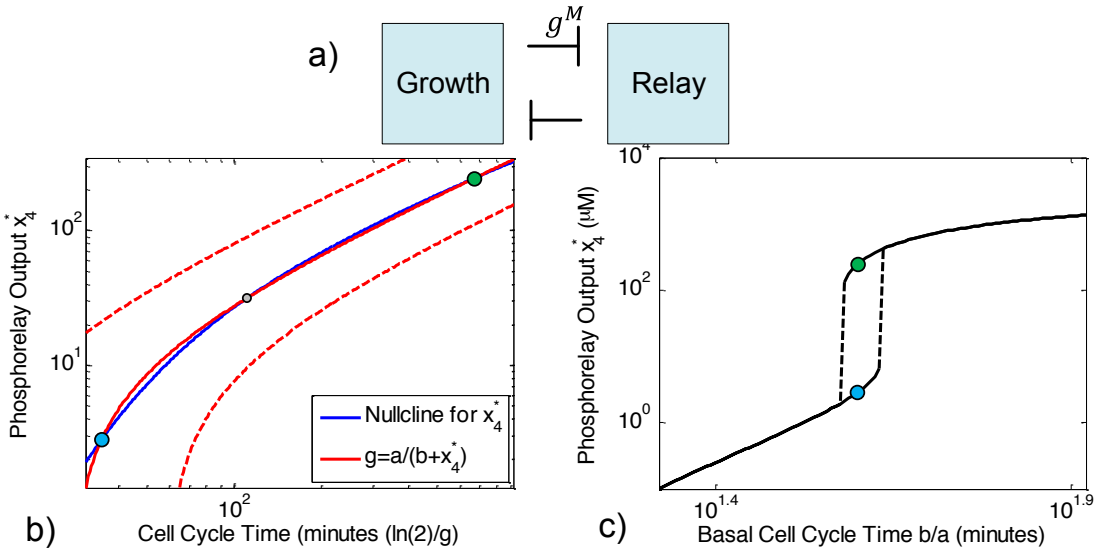


Figure 2-8: a) Schematic showing nonlinear cross-inhibition between growth and the phosphorelay b) Plot of nullcline in blue for x_4^* and growth rate curves in red $g = a/(b + x_4^*)$. The blue nullcline (generated by equation 2-19) corresponds to the parameter set where $k = 1.2\ \text{min}^{-1}$, $F_i = B_i = 0.012\ \mu\text{M}^{-1}\ \text{min}^{-1}$, $P_i = 0.61\ \mu\text{M/min}$, $L_i = 102\ \text{min}^{-1}$, $d_i = 2.5\times10^{-4}\ \text{min}^{-1}$. Growth rate curves are given from bottom to top for $a = \{54,114,271\}\ \mu\text{M}^{-1}\text{min}^{-1}$, and $b = 10\ \mu\text{M}$. The three intersections between the solid nullcline for x_4^* and growth g are denoted by filled circles. c) Hysteresis plot for the system with parameters corresponding to solid lines in subfigure b. Filled circles correspond to intersections given in subfigure b.

Naturally, the system can also exhibit a hysteretic response to other parameters. For example, in Figure 2-9, we consider the output of a system with time varying kinase activity κ_1, which could represent, for example, fluctuations in nutrient availability. In subfigure 2-9.a, nullclines for 3 parameter sets are given. These three parameter sets are identical except that transfer rates F_i and B_i. On the upper dotted nullcline, transfer rates are relatively high compared to growth, and thus we expect that $\left|\epsilon_g(x_4^*)\right| \leq 1$ according to equation 2-49 as there are no phosphatases present in the system. Indeed, we observe from this nullcline that the output is roughly inversely proportional to growth rate. On the central nullcline, the output has a sigmoidal dependence on growth, and furthermore this nullcline intersects our growth equation 2-56 in three places. On the lower dotted nullcline, the output is also sigmoidally dependent on growth, but output levels are too low to allow an intersection at relatively low growth rates (high cell cycle times).

As in the previous example, we perform a hysteresis plot to probe the stability of the three equilibria for the parameter set with intermediate transfer rates. In Figure 2-9.b, we see that the system exhibits a hysteretic response to kinase activity κ_1.

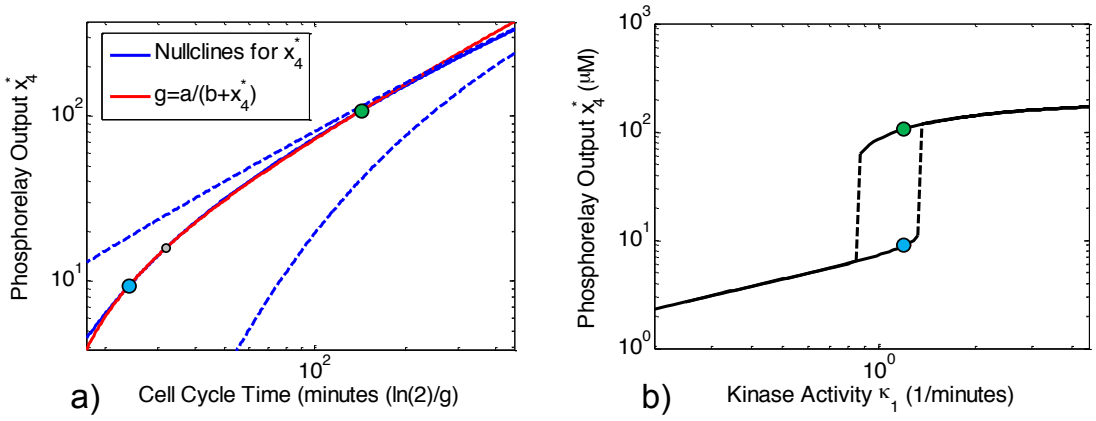


Figure 2-9: a) Schematic showing nonlinear cross-inhibition between growth and the phosphorelay b) Plot of nullclines in blue for x_4^* and growth rate curve $g = a/(b + x_4^*)$ in red. Nullclines were generated using analytical solution for linearized ODE model 2-2. The solid blue null-cline corresponds to parameter set where $k = 1.2\ \text{min}^{-1}$, $F_i = B_i = 0.012\ \mu\text{M}^{-1}\ \text{min}^{-1}$, $P_i = 0.61\ \mu\text{M/min}$, $L_i = 102\ \text{min}^{-1}$, $d_i = 2.5\times10^{-4}\ \text{min}^{-1}$, $a = 114\ \mu\text{M}^{-1}\text{min}^{-1}$, $b = 10\ \mu\text{M}$. The three intersections between the solid nullcline for x_4^* and growth g are denoted by filled circles. The upper dotted nullcline for x_4^* is the same parameter set, except $F_i = B_i = 5\ \mu\text{M}^{-1}\ \text{min}^{-1}$. For the parameter set corresponding to this nullcline, transfer rates are sufficiently high that inequality I_2 fails for cell cycle time shown. Instead of an ultrasensitive response to growth, the upper nullcline shows a roughly linear relationship ($\epsilon_g(x_4^*) \approx 1$) between x_4^* and growth rate as predicted by equation 2-49, and thus there can only be one intersection with the growth curve. By contrast the lower dotted nullcline is for $F_i = B_i = 0.005\ \mu\text{M}^{-1}\ \text{min}^{-1}$. At these slow transfer rates, the system exhibits growth ultrasensitivity, but there is only one intersection with the growth curve. c) Simulation results of nonlinear ODE model 2-1 vs. kinase activity, demonstrating multistability for the nonlinear model. Stable solutions corresponding to nullcline intersections on the left plot are denoted by same colored circles.

Cross inhibition can also result in a simple increase in $\epsilon_a(x_4^*)$ beyond the bounds given in equation 2-45 (which only applies when there is no growth feedback). If we adjust our transfer rates from Figure 2-9.c just beyond the bifurcation point, then we can observe a narrow window of growth rates where $\epsilon_g(x_4^*)$ can be made arbitrarily large in magnitude. An example is given below in Figure 2-10.

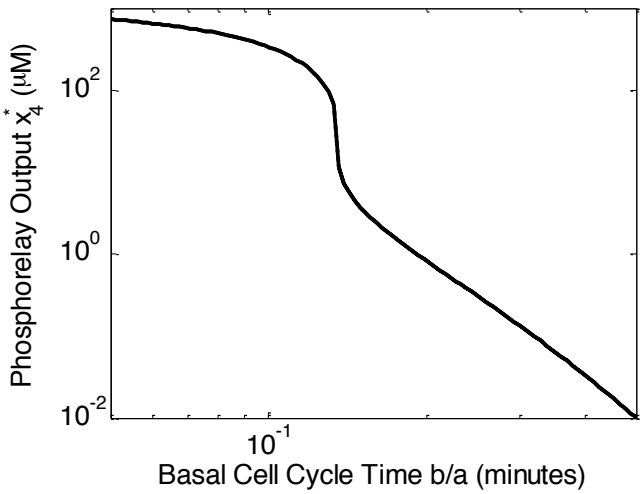


Figure 2-10: Phosphorelay Output x_4^* vs. maximum cell cycle time for where $k = 1.2\ \text{min}^{-1}$, $F_i = B_i = 0.007\ \mu\text{M}^{-1}\ \text{min}^{-1}$, $P_i = 0.61\ \mu\text{M/min}$, $L_i = 102\ \text{min}^{-1}$, $d_i = 2.5\times10^{-4}\ \text{min}^{-1}$. These F_i and B_i values are just barely too small to allow for hysteresis, and thus we see extreme dependence on basal growth rate over a narrow range of growth rates.

2.6.1 Implications of Multistability

As we see in Figures 2-7 and 2-8, multistability can be used to introduce a hysteretic switch-like response to a wide variety of system parameters including influx and efflux signals and gene expression levels.

The most obvious use of such hysteretic responses is to allow the cell to commit to a response to some temporary condition, whether it is a transient environmental signal, or an internal fluctuation in protein concentration or even growth rate. For example, *Bacillus subtilis* cells form spores when they are placed in adverse conditions, and are known to commit to the spore formation process even if conditions improve early on during spore formation [53]. Note that for this specific example, the core *Bacillus subtilis* phosphorelay is embedded in a much larger system with many known feedbacks, and thus the commitment is not likely to be due to the cross inhibitory loop given in Figure 2-8.a.

If noise levels are sufficiently great that they can trip the switching threshold, a hysteretic switch could also be used to randomly diversify populations of cells, allowing a population of cells to hedge its bets against unknown future conditions. By contrast, if noise levels are small, hysteresis can be used to provide a sharp thresholded response without input chatter as discussed in [54] and as demonstrated in Figure 2-10.

We give examples of this chatter suppressive behavior in Figure 2-11. As in Figures 2-8, 2-9, and 2-10, this figure was generated using ODE model 2-2 where growth is given by equation 2-56. For both systems displayed in Figure 2-11, a noisy growth rate (given in green) decreases past a critical threshold, and in response the output (given in blue) increases sharply in response. The difference is in how the hysteretic

system and the switch-like system respond when the input is left near the switching threshold.

In Figure 2-11.c, a was varied according to the green curve, and this input was fed into a system with $x_4^*(a)$ given in Figure 2-11.a. The large magnitude for $\epsilon_a(x_4^*)$ results in noise amplification near the switch-like threshold, and thus the system chatters between low and high output values. In Figure 2-11.d, a was varied according to the green curve and fed into a hysteretic system with output response $x_4^*(a)$ Figure 2-11.b. Again there is a dramatic switching event as a crosses some threshold. However, in contrast to the switch-like system, the hysteretic system does not exhibit chatter, even though the input a remains near the switching threshold.

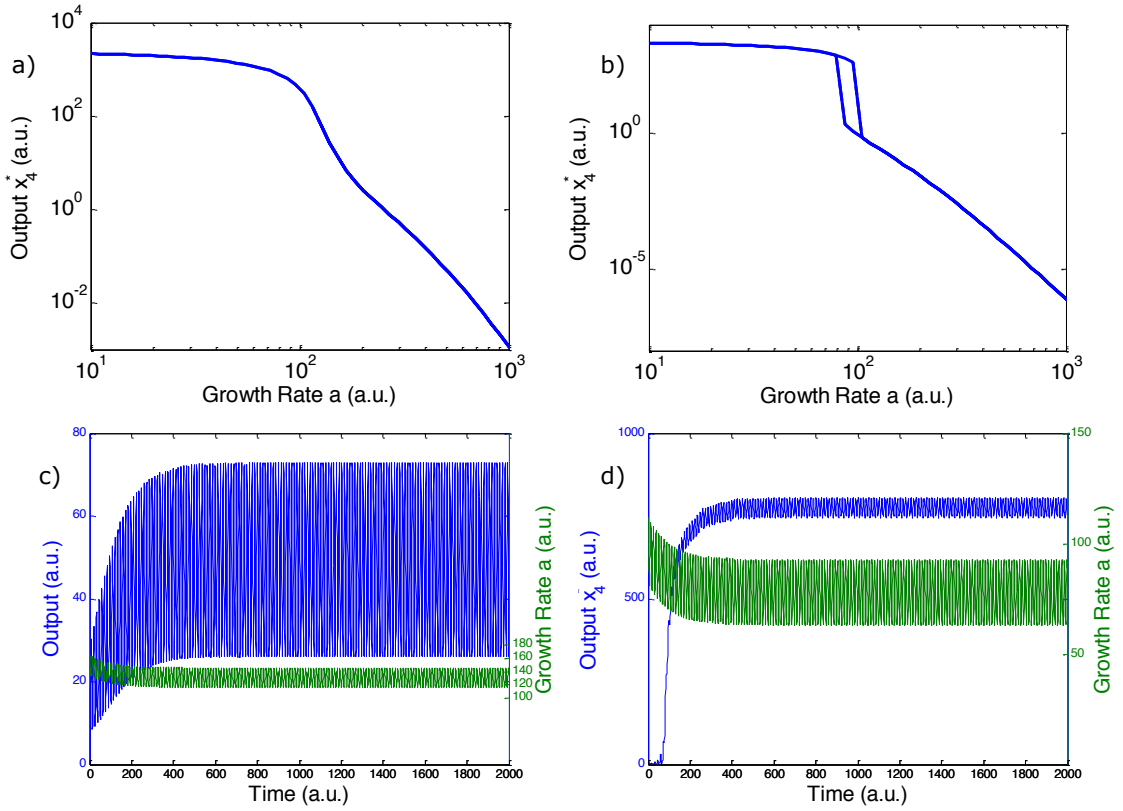


Figure 2-10 a) Plot of system output vs. maximum growth rate a/b for a parameter set exhibiting monostability. b) Plot of system output vs. maximum growth rate parameter a/b for a parameter set exhibiting bistability. c) Effect of noisy growth rate parameter a/b near the steepest (most switch-like) part of the growth sensitivity curve in part a. A 20% change in a/b results in a 65% change in output signal. d) A 32% variation in b/a results in only a 7% variation in output.

2.7 CME Model of the Phosphorelay

In Section 2.3.2, we showed that in the large efflux signal regime, the phosphorelay calculates the ratio of the influx rate to the product of the efflux rates. In Section 2.3.2.1, we showed that in this same regime the phosphorelay has a growth

dependence described by equation 2-26. These results were numerically supported for the full nonlinear model in section 2.5. In this section, we consider the effects of stochasticity using a CME based model.

Above, we show using an ODE model that a phosphorelay can exhibit growth ultrasensitivity. We now consider the same results with a CME based model for two reasons. First, in one well studied system, basal protein concentration c_3 is known to be as low as 0.03 μM in a volume of approximately 1 femtoliter, while c_4 can be as large as 4.4 μM [55]. These values correspond respectively to absolute counts of approximately 20 and 3000 molecules.

Second, our previous model for growth assumes that growth acts in a manner precisely identical to degradation. A better approximation of reality is to assume that cells continuously grow until they reach some critical size threshold, at which point molecules are binomially partitioned between the two child cells[56].

For this section, we build a CME based model of the phosphorelay. The traditional CME formulation considers the volume of the cell constant. One common generalization of the CME model is one in which the propensities of each reaction are updated to reflect the volume change after each reaction[56]. This is the approach that the simulations in this section will utilize. This approach works well as long as the growth rate operates on a slower time scale than the fastest chemical reaction in the system[56]. The model consists of the events and reactions given in Table 2.4. The events were added to provide the generalization of changing volume. The first event represents division events. When cell volume crosses some arbitrary fixed threshold, volume is halved, and each reactant is binomially partitioned according to a binomial distribution of probability 0.5. The second event represents a discrete approximation to cell growth, where every t_{step} seconds, cell volume is increased incrementally by some arbitrary amount.

Reaction Name	Reaction Equation	Propensity
Autophosphorlaytion	$x_1 \rightarrow x_1^*$	$\kappa_1 x_i$
Phosphotransfer	$x_i^* + x_{i+1} \rightleftharpoons x_i + x_{i+1}^*$	$f_i x_i x_{i+1}^* / \Omega$
Dephosphorylation	$x_i^* \rightarrow x_i$	$L_i x_i^*$
Protein Production	$0 \rightarrow x_i$	$P_i \Omega$
Protein Degradation	$x_i \rightarrow 0, x_i^* \rightarrow 0$	$d_i x_i$
Event Name		Trigger
Division	$R = B(R, 0.5), \Omega = \Omega/2$	$\Omega > \Omega_T$
Cell growth	$\Omega = \Omega + g\Delta t$	t_{step} seconds elapsed

Table 2.4: CME Model Specification

Model specifications were created and simulations were performed using the MATLAB SimBiology toolbox using a MATLAB implementation of the generic Gillespie SSA[28] algorithm.

To test the effects of growth on the stochastic relay, growth rate g was varied, and x_4^* was recorded as a function of time. An example trajectory is given in Figure 2-11 below.

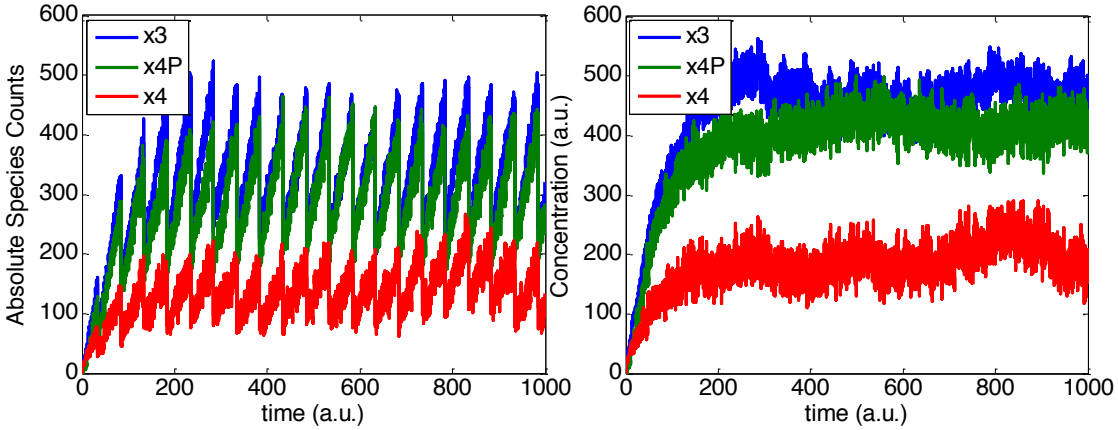


Figure 2-11: Stochastic simulation output showing x_3, x_4^*, and x_4 vs. time for a single parameter set. On the left, absolute molecule counts are shown. Sudden drops correspond to division events. On the right, molecule counts are normalized by volume, which is not plotted, but is also saw-tooth shaped.

2.7.1 Growth Effects on the Open Loop Phosphorelay

If the stochastic model agrees with the deterministic ODE model, then $\epsilon_g(x_4^*)$ should be zero for very low growth rates before decreasing steadily towards -8. $\epsilon_g(x_4^*)$ should never be positive, because there are no phosphatases in this model. For this numerical experiment, non-zero values were selected for all parameters in Table 2.4. For a given parameter set, $E[x_4^*]$ was estimated from a single trajectory by taking the mean of the final eighty percent of the run. This approach was chosen for computational efficiency, avoiding the necessity of generating an ensemble of trajectories. The potential weakness with this approach is that if the system is not monostable, this simplification will fail. For each g value, $E[x_4^*]$ was calculated. To compute $\epsilon_g(x_4^*)$, a low pass filter was applied to the vector of expected values, and the numerical derivative $\frac{dE[x_4^*]}{dg}$ was approximated by computing the slope of the line connecting adjacent data points. These numerical derivatives were then used to calculate $\epsilon_g(x_4^*)$.

For this first set of parameters, in the large growth limit, the growth elasticity was found to obey the same general trend as in the ODE model as shown below in Figure 2-12. Specifically, for low growth rates, the output was found to be relatively insensitive to growth. As growth rates grew large, the output sensitivity $\epsilon_g(x_4^*)$ decreased to a minimum ultrasensitivity value.

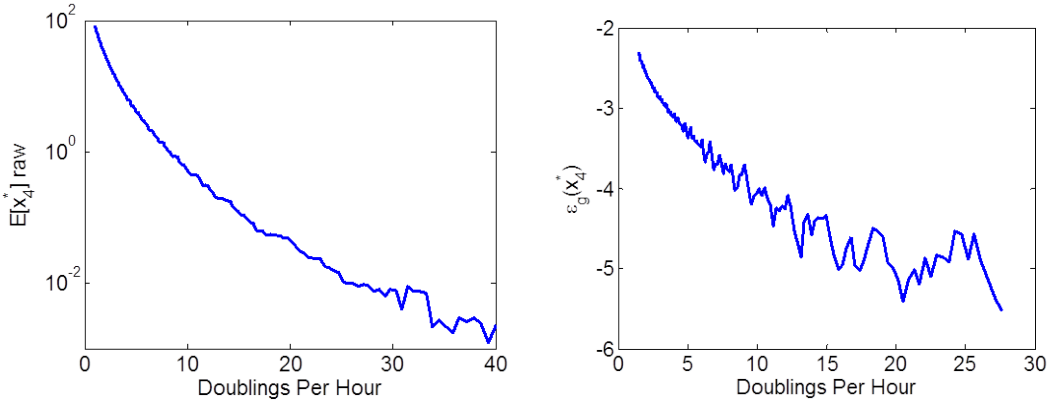


Figure 2-12: Growth sensitivity for a stochastic system. Left side shows molecule counts for $E[x_4^*]$, i.e. the expected value of absolute molecule counts for output x_4^* vs. growth rate. Right side shows $\epsilon_g(x_4^*)$, which is computed from the numerical derivative of $E[x_4^*]$ as described above. Output values become much noisier as growth rates increase, because more data is required to accurately measure $E[x_4^*]$ as it decreases. Even as the expected value of molecule counts falls below 1, the system continues to be increasingly sensitive to growth. In theory, with enough simulations, the right graph would converge to $\epsilon_g(x_4^*) = -8$.

One inherent challenge with this approach is that in order to observe the entire transition from elasticities of $\epsilon_g(x_4^*)$ all the way down to the theoretically smallest elasticities of $\epsilon_g(x_4^*) \approx -8$, we'd need to run simulations which show changes over a dozen or more orders of magnitude. This necessitates either very large molecule counts, which eliminates the use of the SSA as an algorithm, or necessitates very long simulation times in order to observe enough rare x_4^* events in order to properly estimate $E[x_4^*]$. One possible approach to find a lower bound for growth sensitivity would be to simulate only over a very narrow growth rate range, but with very large simulation times in order to accurately estimate $E[x_4^*]$.

2.7.2 Growth Effects on the Closed Loop Phosphorelay

Our CME model can be augmented to consider growth dependence. For this section, our rules in Table 2.4 were removed, and growth was treated exctly as degradation. A rule was added that $g = a/(b + x_4^*)$ that is updated any time x_4^* changes, and a trajectory was generated for the resulting system for values corresponding to those used to generate Figure 2-9. Unlike the deterministic system, the stochastic system is capable of random switching behavior even in the presence of fixed system inputs due to the intrinsic noise introduced in the CME formulation, as shown below in Figure 2-13 This could be used to introduce random behavior similar to excitability, guaranteeing that some fraction of the cells are always in some uncommon state, which could be used as a hedge against uncommon threats. In principle, the system could also be tuned to allow only rare switching events, and these rare events could

be latched to ensure that a cell that enters the random state is unable to escape until the state is unlatched.

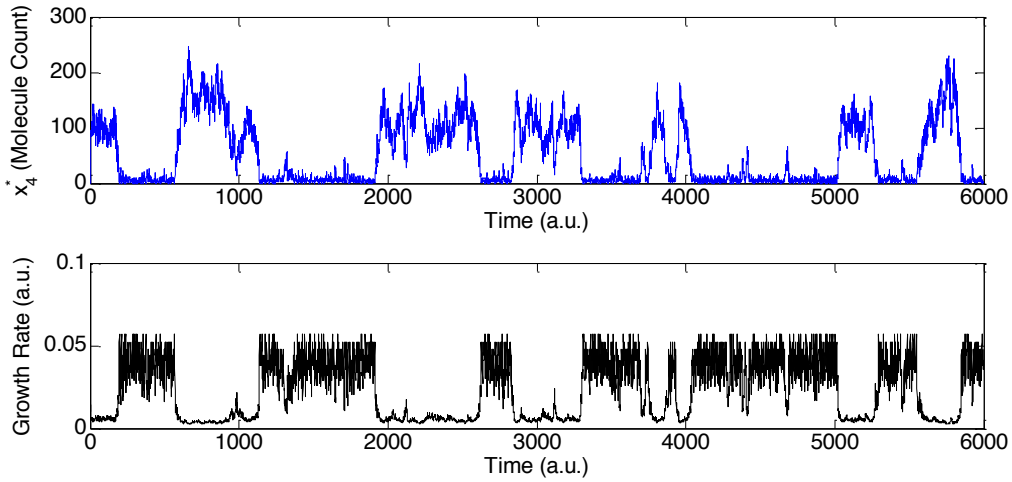


Figure 2-13: Stochastic simulation output showing x_4^* and growth rate g vs. time. Parameters are the same as in ODE simulation 2-9 (parameter set with solid nullcline), except that a was increased to $a = 114\ \mu M^{-1}min^{-1}$. Note that growth model was adjusted to act as a pure death process to align with the ODE simulation in Figure 2-9. In the top figure, absolute molecule counts for the output are shown. In the bottom figure, growth rate is shown. Sudden changes are stochastic switching events due to inherent noise. All parameters with the exception of g are kept constant throughout the simulation.

2.8 Multi-Kinase Relays, Late Stage Kinase Ultrasensitivity

The *Bacillus subtilis* relay features a kinase which acts directly on the Spo0A protein, which fulfills the role of x_4 in a SISB relay[57]. More generally, we can imagine a kinase which can interact with any stage of the relay. This can be modeled with the addition of species y_i and y_i^* which obey the reactions given below in Table 2.5.

$0 \rightarrow y_i$
$y_i \rightarrow 0$
$y_i \rightarrow y_i^*$
$y_i^* \rightarrow y_i$
$y_i^* + x_i \rightleftharpoons y_i + x_i^*$

Table 2.5: Additional Reactions for a Multi-Kinase Relay

Rather than provide a full analysis of such systems, we instead focus on one particularly interesting new behavior that arises in this framework. In particular, we consider the two step relay in the case where some kinase y_2 acts directly on the output stage x_2. In the event that $y_2 + x_2^* \rightarrow y_2 + x_2$ is very slow, then y_2's effect on x_2 can be approximated by lifting the restriction on ODE system 2-1 that $\kappa_2 = 0$, eliminating the necessity of explicitly representing our y_i and y_i^* variables. In this case, the reaction $y_2^* + x_2 \xrightarrow{k_{XX}} y_2 + x_2^*$ becomes $x_2 \xrightarrow{\kappa_2} x_2^*$, where $\kappa_2 = y_2^* k_{XX}$, where y_2^*

will be proportional to the underlying influx signal driving the phosphorylation of y_2.

For the two step relay, the full nonlinear system given by ODE system 2.1 is analytically tractable, even when $\kappa_2 \neq 0$. From the closed form solution (not shown), we find that for a two stage relay with very high backwards transfer rate B_1, a two component system can exhibit ultrasensitivity to κ_2 for small κ_1. As seen in Figure 2-? below, when the backwards transfer rate is very high, x_1 and κ_2 compete to deactivate x_2^*, and κ_1 acts to inhibit x_1 by phosphorylating it (and thus preventing it from acting as a phosphatase on x_2^*). Thus, the early kinase κ_1 acts to modulate the ultrasensitivity of the system to the late stage influx signal κ_2. The effect can be trivially generalized to longer relays by choosing appropriate parameters such that the intermediate stages are given by linear ODEs (i.e. $x_i^* \ll c_i$). Thus, this mechanism is theoretically possible in multi-kinase systems such as the *Bacillus subtilis* phosphorelay, in which kinases act on both the first and fourth stage of the relay.

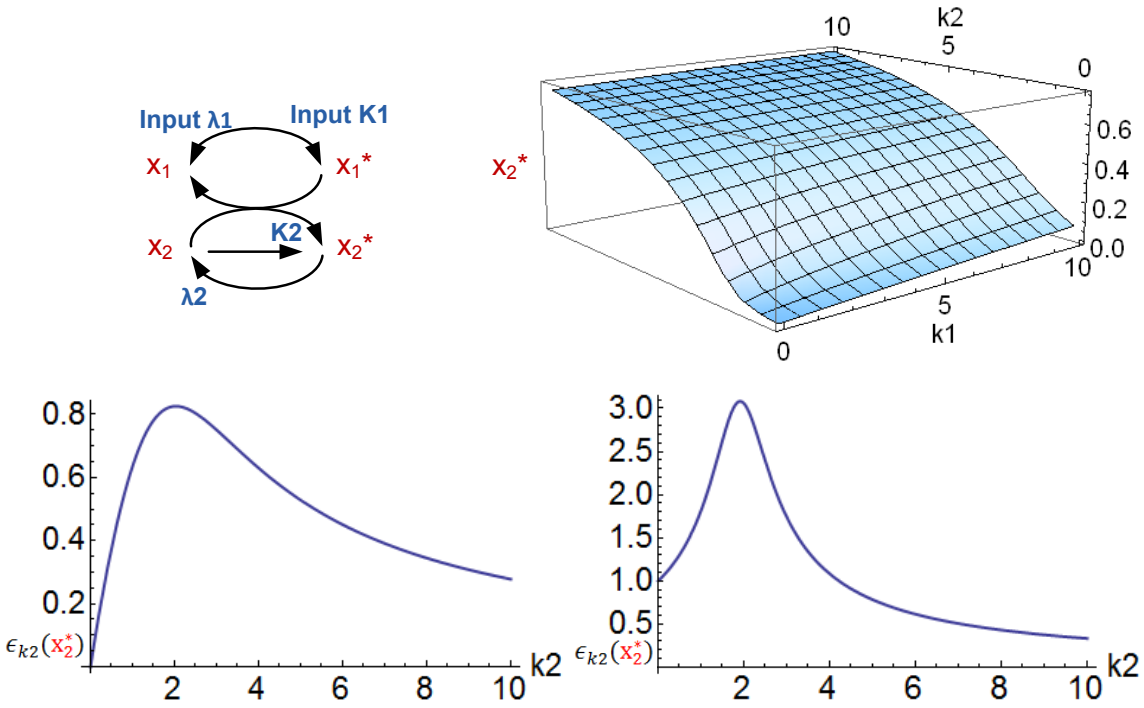


Figure 2-14: a) Schematic of a two component system where both components can act as kinases. b) 3D plot showing x_2^* vs κ_1 and κ_2. c) Plot of elasticity $\epsilon_{k2}(x_2^*)$ vs. κ_2 for small κ_1 d) Plot of $\epsilon_{k2}(x_2^*)$ with respect to κ_2 for large κ_1. Note that the system has a sigmoidal response to κ_2 only for small κ_1. Thus κ_1 modulates the ultrasensitive response of the system to κ_2.

2.9 Proposed Experimental Demonstrations

In principle, it would be possible to test the key predictions of our work, particularly equation 2-26, with a relatively straightforward albeit technically challenging experiment, which would assay GFP produced under relay control as a function of growth rate.

First, histidine kinases would be made constitutively active, for example by cleaving the sensing domain of the histidine kinase[58]. To measure output concentrations, GFP could be placed under transcriptional control of the terminal response regulator. Growth rates would be controlled using nutrient limitation, and bulk relay output could be assessed using flow cytometry to measure GFP concentrations. Growth rates would be assayed at close to the highest attainable levels to maximize growth effects.

The best system architectures to test would be those with the most extreme predicted growth phenotypes. As discussed above, the two extremes are the four component no-phosphatase phosphorelay and the two component four stage phosphorelay with phosphatase regulation on all four stages. In the former case, we expect that output will be strongly suppressed by growth, and in the latter case, strongly promoted by growth. Ideally, the kinetics of candidate systems would be considered in order to maximize the chance of observing the multiplicative growth dilution effect and multiplicative receiver domain concentration effect. Such a system would be difficult to identify *a priori* due to a lack of available kinetic data. It would also be necessary to ensure that cellular physiology is not disrupted when the histidine kinase is made constitutively active.

Though the strongest suppression by growth is expected for the four component system, the only four protein phosphorelay that we're aware of is the notoriously complex *Bacillus subtilis* phosphorelay, which features complex gene regulation and regulation by a vast number of regulating phosphatases whose expression is not fully understood. Consequently, this relay would not be useful without a significant amount of refactoring. Three component relays are seemingly plentiful including those identified in yeast [59], *E coli* and other enterobacteriaceae [60], [61], *Arabidopsis* [62], *Pseudomonas aeroginosa* [63], and at least 39 histidine phosphotransferases have been predicted through genomic analysis [18], each of which suggests the presence of a relay of at least three components. Thus, experimental concerns may necessitate the use of a three component system.

Likewise, though the strongest promotion by growth is expected in a two component system with all four domains regulated by a phosphatase, it may be difficult to coax such a system into exhibiting an extreme growth phenotype, as inequality I_2 requires dephopshorylation rates to be more rapid than *in cis* phosphotransfers on the same protein.

One could also more directly test the simple hypothesis that phosphatase concentrations modulate growth ultrasensitivity by varying the number of stages under phosphorelay control. The simplest way to achieve this would be to test

growth dependence in the presence and absence of a promiscuous phosphatase which acts on all stages of a relay.

3 Case Study of the Bacillus Subtilis Phosphorelay

3.1 Bacillus subtilis

Bacillus subtilis is among the most widely studied and industrially utilized species of bacteria due to its interesting stress response behavior, easily modified DNA, and the desirability of the enzymes it is capable of producing. *subtilis* is known for its wide range of stress response phenotypes, including development of motility and chemotaxis, production of extracellular enzymes for resource collection, uptake of exogenous DNA, and formation of metabolically inert spores[64]. Its ability to uptake exogenous DNA make *subtilis* particularly useful from the experimental and biotechnological point of view, as insertion of DNA is one of the primary techniques utilized by experimental biologists. Its ability to form spores has captivated biologists who seek to understand sporulation as a simpler version of the developmental process by which more complex multi-cellular organisms differentiate into different cell types as they develop[65].

These stress responses and many more are under the control of a phosphorelay[3], whose core components are the proteins KinA (x_1), Spo0F (x_2), Spo0B (x_3) Spo0A (x_4), RapA (π_2), RapC (π_2), RapE (π_2), and Spo0E (π_4), with the SISB equivalent variable names in parentheses. π_2 is duplicated because there are three proteins which act as phosphatases on Spo0F (x_2).

The phosphorylated form of Spo0A, often given as Spo0A* or Spo0A~P as shorthand, is a prolific response regulator, known to control (either directly or indirectly) the production rate of at least 121 genes[7], [67]. Even more interesting is that among these genes are KinA, Spo0F, Spo0A, RapA, and Spo0E[67]. Indeed, this includes all of the SISB phosphorelay proteins except x_3 (Spo0B). A further important twist is that the activities of the phosphatases RapA, RapB, and RapE (π_2) are regulated by small molecules PhrA, PhrC, and PhrE that are secreted and absorbed by *subtilis* cells as a putative extracellular communication factor[68]. Furthermore, these small molecules are differentially expressed, implying that they are each meant to convey a specific piece of information.[68] All of this information is ultimately integrated through the phosphorelay.

The story is even more interesting when one considers that in the wild, colonies of *Bacillus subtilis* form biofilms, which are colonies of cells that bind to each other and

[3] In fact, the *Bacillus subtilis* phosphorelay was the first four stage phosphorelay ever discovered, way back in 1991 [66], incidentally the same year that Nirvana, Tupac, and the Notorious BIG began their ascent. The number of partners with which the *subtilis* phosphorelay is known to interact has steadily grown over the last 20 years.

coat a surface. In the lab, cells have been observed forming biofilms with aerial structures comprised of cells, where cells at the edges of these structures have a greater tendency to sporulate. These aerial structures are also called fruiting bodies[69]. By forming spores at the ends of these fruiting bodies, spores are presumably able to disperse further to enable colonization of new food sources[70]. The spatial and temporal correlation of such biofilms is ultimately mediated in substantial part by Spo0A.

It is far beyond the scope of this thesis to consider models of how these extracellular factors are secreted or even interact with their target phosphatases. Furthermore, there are a vast number of proteins that we have not mentioned which act on the phosphorelay. Instead, this thesis will focus on understanding the function of the core of the *Bacillus subtilis* phosphorelay including transcriptional regulation.

3.2 The Transcriptionally Regulated Phosphorelay Model

To take into account Spo0A*'s effects on relay protein concentration, the SISB model given in Table 2.1 is now augmented with the protein production reactions given as equation 3-1.

$$[x_N^*] + [DNA_i] \rightleftharpoons [x_N^* DNA_i] \rightarrow [x_N^* DNA_i] + [x_i] \tag{3-1}$$

$$[x_N^*] + [DNA_{\pi_i}] \rightleftharpoons [x_N^* DNA_{\pi_i}] \rightarrow [x_N^* DNA_{\pi_i}] + [\pi_i]$$

Though treating transcription as such a simple one step process is a crude approximation, it is a standard one[71] without which our system would be completely intractable analytically. A more detailed model of transcription, translation, and protein folding would be necessary to study the effects of these processes on relay function.

The first reaction can be rewritten in terms of variable c_i as

$$[x_N^*] + [DNA_i] \rightleftharpoons [x_N^* DNA_i] \rightarrow [x_N^* DNA_i] + [c_i] \tag{3-2}$$

This results in the system of ODEs 3-3

$$\begin{aligned}
\frac{dx_i^*}{dt} &= k_i c_i - (B_{i-1}c_{i-1} + F_i c_{i+1} + \alpha)x_i^* + (F_{i-1}(c_i - x_i^*) + B_{i-1})x_{i-1}^* \\
&\quad +(F_i x_i^* + B_i(c_i - x_i^*))x_{i+1}^* \\
\frac{dc_i}{dt} &= P_i + \gamma_i[x_N^* DNA_i] - (d_i + g)c_i \\
\frac{d\pi_i}{dt} &= P_{\pi i} + \gamma_{\pi_i}[x_N^* DNA_{\pi_i}] - (d_{\pi_i} + g)\pi_i \\
\frac{dDNA_k}{dt} &= -K_k x_N^* DNA + K_{-k}[x_N^* DNA]
\end{aligned} \tag{3-3}$$

$$\frac{d[DNA_k x_N^*]}{dt} = -K_k x_N^* DNA + K_{-k}[x_N^* DNA]$$

This ODE system is analytically intractable. To address this, we will utilize an approximation similar to the earlier approximation that $x_i^* + x_{i+1} \leftrightharpoons x_i^* x_{i+1} \leftrightharpoons x_i x_{i+1}^* \leftrightharpoons x_i + x_{i+1}^*$ can be approximated by $x_i^* + x_{i+1} \leftrightharpoons x_i + x_{i+1}^*$. Specifically we will assume that $[x_N^*] + [DNA_i] \rightleftharpoons [x_N^* DNA_i] \rightarrow [x_N^* DNA_i] + [c_i]$ can be approximated by $[x_N^*] \rightarrow [c_i]$.

In this case our system of ODEs 3-3 becomes system 3-4

$$\begin{aligned} \frac{dx_i^*}{dt} &= k_i c_i - (B_{i-1}c_{i-1} + F_i c_{i+1} + \alpha)x_i^* + (F_{i-1}(c_i - x_i^*) + B_{i-1})x_{i-1}^* \\ &\quad +(F_i x_i^* + B_i(c_i - x_i^*))x_{i+1}^* \\ \frac{dc_i}{dt} &= P_i + f_i(x_N^*) - (d_i + g)c_i \\ \frac{d\pi_i}{dt} &= P_{\pi_i} + f_{\pi_i}(x_N^*) - (d_{\pi_i} + g)\pi_i \end{aligned} \tag{3-4}$$

3.2.1 *Bacillus subtilis* Architecture

The *Bacillus subtilis* relay is a four stage, four protein relay. There are phosphatases acting on stages two and four. Thus, N=4, $\pi_1 = 0, \pi_3 = 0$. Furthermore, $f_i(x_4^*)$ is believed to be a strictly increasing function of x_4^* for $i = \{1,2,4\}$, and $f_i(x_4^*)$ is believed to be zero. $f_{\pi_4}(x_4^*)$ is also believed to be strictly increasing, while $f_{\pi_2}(x_4^*)$ is uncertain[67].

3.3 Steady State Output and Interpretations

The steady state function for the output x_4^* of ODE system 3-4 is still given by equation 2-19, where $\phi_i = F_i c_{i+1}$ and $\beta_i = B_i c_i$. However, concentrations c_i and π_i are now given by unknown functions $f_i(x_4^*)$ and $f_{\pi_i}(x_4^*)$. In the regime where the output is small and thus $f_i(x_4^*) \ll P_i$ and $f_{\pi_i}(x_4^*) \ll P_{\pi_i}$, then our transcriptional feedback can be ignored, and trivially equation 2-19 is the closed form expression for x_4^*. Likewise, if the functions $f_i(x_4^*)$ and $f_{\pi_i}(x_4^*)$ have saturated as a function of x_4^*, i.e. $f_i'(x_4^*)$ and $f_{\pi_i}'(x_4^*)$ are sufficiently small, then equation 2-19 is again in closed form, albeit with larger c_i values than before.

Probing the intermediate regime where neither of these two simplifying conditions apply requires some model for transcriptional feedback. Even a fairly simple model of transcriptional feedback will result in an intractable steady state output. If we assume that only the phosphatase regulated domains experience significant phosphate loss, i.e. $\lambda_1 \ll \phi_1 + \beta_1$ and $\lambda_3 \ll \phi_3 + \beta_3$, and that this loss is due to π_i, then output equation 2-19 becomes equation 3-7, where weights w_α have been expanded according to Appendix A. c_2 is missing from the numerator because it cancels out with the c_2 that is present in weights $w_{(0,1,0,0)}$, $w_{(0,0,0,1)}$, and $w_{(0,1,0,1)}$.

$$x_4^* = \frac{c_1 c_3 c_4 F_2 F_3 \kappa_1}{B_2 B_3 c_2 c_3 k_{\pi_2} \pi_2 + c_3 c_4 F_2 F_3 k_{\pi_4} \pi_4 + (B_2 c_2 + c_4 F_3) k_{\pi_2} k_{\pi_4} \pi_2 \pi_4} \tag{3-5}$$

Kinase activity is believed to be a measure of the nutritional quality of the growth medium, and previous work has suggested that k_{π_2} provides a measure of the number of nearby growing cells in the environment[49] . Thus, the output of the phosphorelay may provide a measure of the amount of food per growing cell, prompting a cell to tend to sporulate when its own chances of finding food are poor.

3.3.1 Feedback Provides Just-in-time Supply of Proteins

We can probe the potential usefulness of transcriptional feedback without assuming any particular function $f(x_4^*)$ for that feedback. Assuming that c_1, c_2, c_4, and π_4 are proportional to the same function $f(x_4^*)$ of x_4^*, then we have that

$$x_4^* \propto \frac{f(x_4^*) F_2 F_3 \kappa_1}{B_2 B_3 k_{\pi_2} \pi_2 + f(x_4^*) F_2 F_3 k_{\pi_4} + (B_2 + F_3) k_{\pi_2} \pi_2 f(x_4^*) k_{\pi_4}} \tag{3-6}$$

If $f(x_4^*)$ is sufficiently large, then $B_2 B_3 c_3 k_{\pi_2} \pi_2$ disappears and equation 3-6 becomes

$$x_4^* \propto \frac{F_2 F_3 \kappa_1}{F_2 F_3 k_{\pi_4} + (B_2 + F_3) k_{\pi_2} \pi_2 k_{\pi_4}} \tag{3-7}$$

This equation, like equation 3-5, is inversely proportional to the dephosphorylation rates k_{π_2} and k_{π_4} and thus the phosphorelay is sensitive to the extracellular factors that regulates the activity of π_2, as long as the activity of the constitutive activity of π_4 is low enough to avoid dwarfing π_2's influence. One thing that has changed is the value of our protein concentrations which are now larger. Consequently, our output given by equation 3-7 can now reach larger values before the relay is saturated and the low activation assumption used to derive equation 2-19 and ultimately 3-7 fails. The idea that the transcriptional feedback exists in order to provide a just in time supply of phosphorelay protein is supported by the literature[38], [72].

3.3.2 Feedback on Spo0B May Block RapA Activity

Equation 3-7 suggests one possible reason that feedback is not observed on Spo0B in the wild. However, if there is feedback on Spo0B, i.e. c_3 is also proportional to $f(x_4^*)$, then equation 3-7 instead becomes

$$x_4^* \propto \frac{f(x_4^*) c_4 F_2 F_3 \kappa_1}{B_2 B_3 k_{\pi_2} \pi_2 + f(x_4^*) F_2 F_3 k_{\pi_4} + (B_2 + F_3) k_{\pi_2} \pi_2 k_{\pi_4}} \tag{3-8}$$

Now, in the large $f(x_4^*)$ limit, equation 3-8 becomes

$$x_4^* \propto \frac{c_4 \kappa_1}{k_{\pi_4}} \tag{3-9}$$

In contrast to the system with no Spo0B feedback, equation 3-9 shows no dependence on dephosphoylation rate k_{π_2}, and thus factors acting on x_2 will be effectively ignored. This reason for the lack of Spo0B feedback is based on a large number of assumptions. In particular, this explanation does not apply if

3.3.3 Feedback on Spo0B May Corrupt Phosphorelay Function

One can attempt to more deeply probe the intermediate $f(x_4^*)$ regime by choosing a model for growth. One standard approach is to approximate transcriptional feedback using Michaelis-Menten kinetics[71], given below as equation 3-10.

$$f_i = \gamma_i \frac{x_N^*}{x_N^* + K_i} \tag{3-10}$$

$$f_{\pi_i} = \gamma_{\pi_i} \frac{x_N^*}{x_N^* + K_{\pi_i}}$$

Yielding protein concentrations:

$$c_i = \frac{P_i + \gamma_1 \frac{x_N^*}{x_N^* + K_i}}{d_i + g} \tag{3-11}$$

$$\pi_i = \frac{P_i + \gamma_{\pi_i} \frac{x_N^*}{x_N^* + K_{\pi_i}}}{d_i + g}$$

In principle, one can substitute equations 3-11 into equation 3-5 and solve for x_N^*. As discussed in the case of general f_i and f_{π_i}, in the event that basal production is stronger than production due to transcriptional feedback or that feedback has saturated as a function of x_N^* (given by $f_i(x_N^*) \ll P_i$ and $x_N^* \gg K_i$ respectively), then the form of the output remains exactly as equation 2-19, and all of the multiplicative and growth dependence properties follow.

For x_N^* values on the same order as K_i, this substitution is analytically intractable, and no closed form solution for x_N^* exists. However, if we assume that feedback on each stage overpowers basal protein production ($f_i \gg P_i$) and that our output is small relative to the feedback saturation level ($x_N^* \ll K_N$), then we can approximate c_i and π_i by linear functions of x_N^*, where $\tilde{\gamma}_i$ are metaparameters.

$$c_i \approx \frac{\gamma_i \frac{x_N^*}{K_i}}{d_i + g} = \tilde{\gamma}_i x_N^* \tag{3-12}$$

$$\pi_i \approx \frac{\gamma_{\pi_i} \frac{x_N^*}{K_{\pi_i}}}{d_i + g} = \tilde{\gamma}_{\pi_i} x_N^*$$

Substituting equation 3-12 into equation 3-5 for c_i for $i = \{1,2,4\}$ and π_4 yields in the large π_4 limit

$$x_4^* = \frac{\tilde{\gamma}_1 c_3\, \tilde{\gamma}_4 F_2 F_3 \kappa_1 - B_2 B_3 c_3 \tilde{\gamma}_2}{B_2 B_3 \tilde{\gamma}_2 c_3 k_{\pi_2} \pi_2 + c_3 \tilde{\gamma}_4 F_2 F_3 k_{\pi_4} \tilde{\gamma}_{\pi_4} + (B_2 \tilde{\gamma}_2 + c_4 F_3) k_{\pi_2} k_{\pi_4} \pi_2 \tilde{\gamma}_{\pi_4}} \tag{3-13}$$

The steady function is nearly unchanged from equation 3-5, though there are two differences. The first is that that each term has been rescaled, effectively weighting each efflux signal differently in the final calculation. The second is that a constant is subtracted from the numerator. When this constant becomes larger than the influx term in the numerator (e.g. κ_1 becomes too small), then x_4^* will become too small for equation 3-12 to remain true, and basal transcription will dominate x_4^* driven transcription and equation 3-13 will pass through an intractable regime before reverting to the form of equation 2-19.

By contrast, if equation 3-12 also applies for c_3, i.e. Spo0B is under transcriptional feedback from Spo0A*, then equation 3-5 instead becomes

$$x_4^* = \frac{\tilde{\gamma}_1 \tilde{\gamma}_3\, \tilde{\gamma}_4 F_2 F_3 \kappa_1 - B_2 B_3 \tilde{\gamma}_2 \tilde{\gamma}_3 L_2 - B_2 \tilde{\gamma}_2 k_{\pi_2} k_{\pi_4} \tilde{\gamma}_{\pi_4} \pi_2 - F_3 \tilde{\gamma}_4 k_{\pi_2} k_{\pi_4} \tilde{\gamma}_{\pi_4} \pi_2}{F_2 F_3 \tilde{\gamma}_3 \tilde{\gamma}_4 k_{\pi_4} \tilde{\gamma}_{\pi_4}} \tag{3-14}$$

Now, the output is proportional to the difference between the kinase signal κ_1 and the quorum sensitive phosphatase activity k_{π_2} instead of being proportional to their ratio. Furthermore, as soon as the subtractive term in 3-14 becomes significant, the system will transition out of the regime where 3-12 is true, and will pass through an intractable regime before reverting to the form of equation 2-19.

3.3.4 Feedback on RapA May Act as a Commitment Step

Prior work has suggested that π_2 may be suppressed by x_4^*[67]. Thus, π_2 and x_4^* may form a cross repressive loop. In this case, we model π_2 by equation 3-15

$$\frac{d\pi_2}{dt} = P_{\pi_2} + \frac{v}{1 + \left(\frac{x_4^*}{K_{\pi_2}}\right)^n} - d_{\pi_2} \pi_2 \tag{3-15}$$

Thus, π_2 is in equilibrium whenever

$$d_{\pi_2}\pi_2 = P_{\pi_2} + \frac{v}{1 + \left(\frac{x_4^*}{K_{\pi_2}}\right)^n} \tag{3-16}$$

This cross repressive feedback loop allows the possibility that the system will be multistable, with the system maintaining its high output status in the event that π_2 is made low by the phosphorelay. Indeed, time-lapse fluorescence microscopy experiments have shown that cells enter a low π_2 state before sporulating, which requires a high level of x_4^* [49].

4 Thesis Conclusion

4.1 Thesis Summary

This research has expanded upon a long time hypothesis regarding the function of the ubiquitous phosphorelay. Though it has long been suspected that a phosphorelay acts as a signal integration module, this is the first work that provides a concise mathematical statement of how signals are integrated. Specifically, the output of a phosphorelay is found to be proportional to the aggregate influx signal divided by the sum of various products of the aggregate efflux signals acting at each stage of the relay. If all of the aggregate efflux signals are large, then the output is simply proportional to their product.

The most obvious manner in which to use this multiplicative property is to multiply signals that convey some useful information about the environment. In the context of *Bacillus subtilis* where one of the aggregate efflux signals may be a measure of population and the influx signal a measure of available nutrient, the phosphorelay may be computing the amount of food per cell that is available in the environment.

The output function derived in Chapter 2 shows a complex dependence on growth. Growth acts as both an efflux signal and as a controller of protein concentrations. These proteins include both phosphate receiving proteins and phosphate inhibiting proteins. As a result of these two effects, the output of the relay becomes ultrasensitive to growth, and may exhibit either a positive or negative dependence on growth. The degree to which these effects can be cleanly expressed in terms of key physical parameters, including the number of stages in the relay, the number of distinct proteins forming the relay, the number of proteins that are relatively stable compared to growth, and the number and configuration of stable phosphatases.

4.2 Future Work

Above, I've shown that a phosphorelay can theoretically be used as an analog calculation device, potentially allowing for a phosphorelay to compute the ratio of

its influx signals to the sum of various products of the efflux signals, or in the large efflux signal case, the ratio of the influx signal to the product of the efflux signals. Though I've speculated above why such a function might be useful, a careful analysis of the literature for phosphorelays which might be hiding such multiplicative functions could be interesting. Likewise, the phosphorelay may be act as a growth switch, and in section 2.10, I've proposed experiments that can be used to test this theory.

The basic theory could also be enhanced through a rigorous proof that the SISBOL phosphorelay is not only monotone, but also globally asymptotically stable. The numerical support for the growth sensitivity hypothesis given in section 2.7.1 could also be enhanced through a demonstration that a CME model can exhibit the minimum theoretical elasticity of $-N$. This may necessitate the use of the Exact Time Gillespie algorithm[56] in order to accurately model the phosphorelay at growth rates that are high enough that the cell divides as fast or faster than significant chemical reactions in the system. The numerical support for the stochastic switching hypothesis given in section 2.7.2 could be enhanced by showing that stochastic switching occurs even with a division based model of growth, as opposed to the simplified model of growth explored in that section.

The theory itself could be generalized to cover a broader class of relays than the single branch relays discussed in this thesis. One obvious generalization would be to consider multi-branch phosphorelays, where multiple relays act in parallel with crosstalk at various stages. Such relays have been observed, for example, in *Arabidopsis*[73]. Another important question is how the concept of growth sensitivity generalize to other signaling cascades, particularly kinase cascades. Preliminary work (not shown in this thesis) suggests that a kinase cascade should have a growth elasticity whose magnitude increases exponentially with the length of the kinase cascade in the most extreme case. Likewise phosphatases acting on early stages of a kinase cascade seem to have a similar exponential effect of opposite sign.

Finally, this work can be used as a staging point for understanding the biology of particular organisms which utilize the phosphorelay as one of their core signal integration components. *Bacillus subtilis* is particularly interesting, as its stress responses are driven by multiple phosphorelays under complex regulation by a variety of phosphatases which are in turn modulated by a series of putative extracellular communication factors produced under different circumstances, effectively providing a series of putative parallel communication channels[68], [74], [75]. Further, the communities it forms seem to be spatially and temporally coordinated to maximize the fitness of the population as a whole[69], and it is tempting to believe that this behavior is coordinates via these channels. Understanding such relays requires a solid understanding of the simple fundamental module that lies at the center.

This work provides that baseline for understanding relays from the simple 2 protein 4 stage relay found in *Bordetella pertussis*[76], to the highly complex *Bacillus subtilis* relay discussed above. Understanding the former organism, for example, would provide insights into the control schemes utilized by dangerous pathogens to infect humans and evade the immune system. Understanding the latter would provide an understanding of how bacteria communicate and coordinate a community of diversified cells, thus providing a model system for understanding how sophisticated biological multicellular machines like ourselves develop from a single cell, and how similarly complex systems may someday be designed.

Bibliography

[1] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology.," *Nature*, vol. 402, no. 6761, 1999.

[2] U. S. Bhalla and Ravi Iyengar, "Emergent Properties of Networks of Biological Signaling Pathways," *Science*, vol. 283, pp. 381-387, Jan. 1999.

[3] J. E. Ferrell Jr, "Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs," *Trends in biochemical sciences*, vol. 21, no. 12, pp. 460–466, 1996.

[4] P. Cohen, "Signal integration at the level of protein kinases, protein phosphatases and their substrates," *Trends in biochemical sciences*, vol. 17, no. 10, pp. 408–413, 1992.

[5] M. Samoilov, A. Arkin, and J. Ross, "Signal Processing by Simple Chemical Systems," *The Journal of Physical Chemistry A*, vol. 106, pp. 10205-10221, Oct. 2002.

[6] J. Adler, "My Life with Nature," *Annual Review of Biochemistry*, vol. 80, pp. 42–70, 2011.

[7] V. Molle et al., "The Spo0A regulon of Bacillus subtilis," *Molecular Microbiology*, vol. 50, pp. 1683–1701, 2003.

[8] M. Eisenbach, "Control of bacterial chemotaxis," *Molecular microbiology*, vol. 20, no. 5, pp. 903–910, 1996.

[9] I. B. Levitan, "Modulation of ion channels by protein phosphorylation and dephosphorylation," *Annual review of physiology*, vol. 56, no. 1, pp. 193–212, 1994.

[10] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell," *Current Opinion in Cell Biology*, vol. 15, pp. 221-231, Apr. 2003.

[11] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, p. 11980, 2003.

[12] W. C. Fuqua, S. C. Winans, and E. P. Greenberg, "Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.," *Journal of Bacteriology*, vol. 176, no. 2, p. 269, 1994.

[13] Y. L. Deribe, T. Pawson, and I. Dikic, "Post-translational modifications in signal integration," *Nature Structural & Molecular Biology*, vol. 17, pp. 666-672, May 2010.

[14] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature biotechnology*, vol. 21, no. 3, pp. 255–261, 2003.

[15] J. V. Olsen et al., "Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks," *Cell*, vol. 127, pp. 635-648, Nov. 2006.

[16] J. A. Hoch, "Two-component and phosphorelay signal transduction," *Current opinion in microbiology*, vol. 3, no. 2, pp. 165–170, 2000.

[17] C. Y. Huang and J. E. Ferrell, "Ultrasensitivity in the mitogen-activated protein kinase cascade," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 19, p. 10078, 1996.

[18] W. Zhang and L. Shi, "Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases," *Microbiology*, vol. 151, no. 7, p. 2159, 2005.
[19] D. R. Caffrey, L. A. . O'Neill, and D. C. Shields, "The evolution of the MAP kinase pathways: coduplication of interacting proteins leads to new signaling cascades," *Journal of molecular evolution*, vol. 49, no. 5, pp. 567–582, 1999.
[20] A. Waskiewicz and J. Cooper, "Mitogen and stress response pathways: MAP kinase cascades and phosphatase regulation in mammals and yeast," *Current Opinion in Cell Biology*, vol. 1995, no. 7, pp. 798-805.
[21] D. M. Virshup and S. Shenolikar, "From Promiscuity to Precision: Protein Phosphatases Get a Makeover," *Molecular Cell*, vol. 33, pp. 537-545, Mar. 2009.
[22] A. P. Mähönen et al., "Cytokinins regulate a bidirectional phosphorelay network in Arabidopsis," *Current biology*, vol. 16, no. 11, pp. 1116–1122, 2006.
[23] B. L. Taylor and I. B. Zhulin, "PAS domains: internal sensors of oxygen, redox potential, and light," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 2, p. 479, 1999.
[24] D. Albanesi et al., "Structural plasticity and catalysis regulation of a thermosensor histidine kinase," *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, p. 16185, 2009.
[25] K. Otto and T. J. Silhavy, "Surface sensing and adhesion of Escherichia coli controlled by the Cpx-signaling pathway," *Proceedings of the National Academy of Sciences*, vol. 99, no. 4, p. 2287, 2002.
[26] M. Perego, "Kinase-phosphatase competition regulates Bacillus subtilis development," *Trends in microbiology*, vol. 6, no. 9, pp. 366–370, 1998.
[27] J. C. Perez and E. A. Groisman, "Acid pH activation of the PmrA/PmrB two-component regulatory system of Salmonella enterica," *Molecular microbiology*, vol. 63, no. 1, pp. 283–293, 2007.
[28] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 35–55, 2007.
[29] B. Ilic et al., "Single cell detection with micromechanical oscillators," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 19, p. 2825, 2001.
[30] G. Bratbak and I. Dundas, "Bacterial dry matter content and biomass estimations.," *Applied and Environmental Microbiology*, vol. 48, no. 4, p. 755, 1984.
[31] C. J. Tomlin and J. D. Axelrod, "Biology by numbers: mathematical modelling in developmental biology," *Nature reviews genetics*, vol. 8, no. 5, pp. 331–340, 2007.
[32] H. Meinhardt and P. A. J. De Boer, "Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site," *Proceedings of the National Academy of Sciences*, vol. 98, no. 25, p. 14202, 2001.
[33] M. Ullah and O. Wolkenhauer, *Stochastic Approaches for Systems Biology*, First. New York: Springer, 2011.
[34] B. N. Kholodenko, "Why do protein kinase cascades have more than one level?," *Trends in Biochemical Sciences*, vol. 22, no. 8, pp. 228-229, Aug. 1997.

[35] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko, “Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades,” *Science’s STKE*, vol. 164, no. 3, p. 353, 2004.
[36] J. R. Kim and K. H. Cho, “The multi-step phosphorelay mechanism of unorthodox two-component systems in E. coli realizes ultrasensitivity to stimuli while maintaining robustness to noises,” *Computational biology and chemistry*, vol. 30, no. 6, pp. 438–444, 2006.
[37] A. Csikasz-Nagy, L. Cardelli, and O. S. Soyer, “Response dynamics of phosphorelays suggest their potential utility in cell signalling,” *Journal of The Royal Society Interface*, Aug. 2010.
[38] A. Chastanet, D. Vitkup, G. C. Yuan, T. M. Norman, J. S. Liu, and R. M. Losick, “Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 18, p. 8486, 2010.
[39] D. Angeli and E. D. Sontag, “Multi-stability in monotone input/output systems,” *Systems & control letters*, vol. 51, no. 3-4, pp. 185–202, 2004.
[40] D. Huh and J. Paulsson, “Non-genetic heterogeneity from stochastic partitioning at cell division,” *Nature Genetics*, vol. 43, pp. 95-100, Dec. 2010.
[41] M. Perego and J. A. Hoch, “Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in Bacillus subtilis,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 4, p. 1549, 1996.
[42] D. Angeli, P. De Leenheer, and E. D. Sontag, “On the structural monotonicity of chemical reaction networks,” in *Decision and Control, 2006 45th IEEE Conference on*, 2006, pp. 7–12.
[43] S. Klumpp, Z. Zhang, and T. Hwa, “Growth Rate-Dependent Global Effects on Gene Expression in Bacteria,” *Cell*, vol. 139, no. 7, pp. 1366-1375, Dec. 2009.
[44] A. Goldbeter and D. E. Koshland, “Sensitivity amplification in biochemical systems,” *Quarterly Reviews of Biophysics*, vol. 15, no. 3, pp. 555–591, 1982.
[45] G. Psakis, J. Mailliet, C. Lang, L. Teufel, L. O. Essen, and J. Hughes, “Signalling kinetics of cyanobacterial phytochrome Cph1, a light regulated histidine kinase,” *Biochemistry*, 2011.
[46] A. Stock, V. Robinson, and P. Goudreau, “Two-Component Signal Transduction,” *Annual Review of Biochemistry*, vol. 69, no. 2000, pp. 183-215, 2000.
[47] T. T. Simopoulos and W. P. Jencks, “Alkaline phosphatase is an almost perfect enzyme,” *Biochemistry*, vol. 33, no. 34, pp. 10375–10380, 1994.
[48] C. E. Grimshaw et al., “Synergistic Kinetic Interactions between Components of the Phosphorelay Controlling Sporulation in Bacillus subtilis†,” *Biochemistry*, vol. 37, no. 5, pp. 1365–1375, 1998.
[49] I. B. Bischofs, J. A. Hug, A. W. Liu, D. M. Wolf, and A. P. Arkin, “Complexity in bacterial cell–cell communication: Quorum signal integration and subpopulation signaling in the Bacillus subtilis phosphorelay,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, p. 6459, 2009.
[50] H. Schmidt and M. Jirstrand, “Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology,” *Bioinformatics*, vol. 22, no. 4, pp. 514–515, 2006.

[51] L. F. Shampine and M. W. Reichelt, "The matlab ode suite," *SIAM journal on scientific computing*, vol. 18, no. 1, pp. 1–22, 1997.
[52] L. Mccarter, "Regulation of Flagella," *Current Opinion in Microbiology*, vol. 9, no. 2, pp. 180-186, Apr. 2006.
[53] J. M. Sterlini and J. Mandelstam, "Commitment to sporulation in Bacillus subtilis and its relationship to development of actinomycin resistance," *Biochemical Journal*, vol. 113, no. 1, p. 29, 1969.
[54] D. M. Wolf and A. P. Arkin, "Motifs, modules and games in bacteria," *Current Opinion in Microbiology*, vol. 6, pp. 125-134, Apr. 2003.
[55] P. Eswaramoorthy, J. Dinh, D. Duan, O. A. Igoshin, and M. Fujita, "Single-cell measurement of the levels and distributions of the phosphorelay components in a population of sporulating Bacillus subtilis cells," *Microbiology*, vol. 156, no. 8, p. 2294, 2010.
[56] T. Lu, D. Volfson, L. Tsimring, and J. Hasty, "Cellular growth and division in the Gillespie algorithm," *Systems Biology*, vol. 1, no. 1, p. 121, 2004.
[57] M. Jiang, W. Shao, M. Perego, and J. A. Hoch, "Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis," *Molecular Microbiology*, vol. 38, no. 3, pp. 535–542, 2000.
[58] P. Eswaramoorthy and M. Fujita, "Systematic domain deletion analysis of the major sporulation kinase in Bacillus subtilis," *Journal of bacteriology*, vol. 192, no. 6, p. 1744, 2010.
[59] A. O. Kaserer, B. Andi, P. F. Cook, and A. H. West, "Kinetic Studies of the Yeast His-Asp Phosphorelay Signaling Pathway," in *Methods in Enzymology*, vol. 471, Elsevier, 2010, pp. 59-75.
[60] Y. H. Huang, L. Ferrières, and D. J. Clarke, "The role of the Rcs phosphorelay in Enterobacteriaceae," *Research in microbiology*, vol. 157, no. 3, pp. 206–212, 2006.
[61] S. Takeda, Y. Fujisawa, M. Matsubara, H. Aiba, and T. Mizuno, "A novel feature of the multistep phosphorelay in Escherichia coli: a revised model of the RcsC→ YojN→ RcsB signalling pathway implicated in capsular synthesis and swarming behaviour," *Molecular microbiology*, vol. 40, no. 2, pp. 440–450, 2001.
[62] T. Kakimoto, "Perception and Signal Transduction of Cytokinins," *Annual Review of Plant Biology*, vol. 54, no. 1, pp. 605-627, Jun. 2003.
[63] J.-L. Hsu, H.-C. Chen, H.-L. Peng, and H.-Y. Chang, "Characterization of the Histidine-containing Phosphotransfer Protein B-mediated Multistep Phosphorelay System in Pseudomonas aeruginosa PAO1," *Journal of Biological Chemistry*, vol. 283, no. 15, pp. 9933-9944, Feb. 2008.
[64] C. R. Harwood, "Bacillus subtilis and its relatives: molecular biological and industrial workhorses," *Trends in biotechnology*, vol. 10, pp. 247-256, 1991.
[65] D. Iber, J. Clarkson, M. D. Yudkin, and I. D. Campbell, "The mechanism of cell differentiation in Bacillus subtilis," *Nature*, vol. 441, no. 7091, pp. 371-374, May 2006.
[66] D. Burbulys, K. A. Trach, and J. A. Hoch, "Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay," *Cell*, vol. 64, no. 3, pp. 545-552, Feb. 1991.

[67] M. Fujita, J. E. González-Pastor, and R. Losick, "High-and low-threshold genes in the Spo0A regulon of Bacillus subtilis," *Journal of bacteriology*, vol. 187, no. 4, p. 1357, 2005.
[68] M. Pottathil, B. A. Lazazzera, and others, "The extracellular Phr peptide-Rap phosphatase signaling circuit of Bacillus subtilis," *Frontiers in Bioscience*, vol. 8, pp. 32–45, 2003.
[69] S. S. Branda, J. E. González-Pastor, S. Ben-Yehuda, R. Losick, and R. Kolter, "Fruiting body formation by Bacillus subtilis," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, p. 11621, 2001.
[70] B. J. Crespi, "The evolution of social behavior in microorganisms," *Trends in Ecology & Evolution*, vol. 16, no. 4, pp. 178–183, 2001.
[71] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, "An excitable gene regulatory circuit induces transient cellular differentiation," *Nature*, vol. 440, no. 7083, pp. 545-550, Mar. 2006.
[72] A. Chastanet and R. Losick, "Just-in-Time Control of Spo0A Synthesis in Bacillus subtilis by Multiple Regulatory Mechanisms," *Journal of Bacteriology*, vol. 193, no. 22, pp. 6366-6374, 2011.
[73] H. Dortay, N. Mehnert, L. Bürkle, T. Schmülling, and A. Heyl, "Analysis of protein interactions within the cytokinin-signaling pathway of Arabidopsis thaliana," *FEBS Journal*, vol. 273, no. 20, pp. 4631–4644, 2006.
[74] D. Schultz, P. G. Wolynes, E. B. Jacob, and J. N. Onuchic, "Deciding fate in adverse times: Sporulation and competence in Bacillus subtilis," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, p. 21027, 2009.
[75] 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*, vol. 79, no. 6, p. 1047, 1994.
[76] P. A. Cotter and A. M. Jones, "Phosphorelay control of virulence gene expression in Bordetella," *Trends in Microbiology*, vol. 11, no. 8, pp. 367-373, Aug. 2003.

A Expansion of w_k

Above, we give that the output for the bidirectional relay as equation 2-19, which includes efflux signal product weights w_α. This in turn was derived from the set of algebraic equations given as equation 2-18. The weights come directly from w_α.

There is no simple expression for w_α, so we provide the following constructive algorithm instead.

1. Given an efflux signal product $\boldsymbol{\lambda}^\alpha = \prod_{i=1}^{N} \lambda_i^{\alpha_i}$, we wish to compute the corresponding weight w_α
2. Special cases: $w_{(0,\ldots,0)} = 0$, $w_{(1,\ldots,1)}$=1
3. Draw a set of N collinear circles
4. If $\alpha_i = 1$, fill in the circle at node N. If $\alpha_i = 0$ leave the circle empty
5. For every empty circle, draw an arrow pointing to either its left or right neighbor subject to the rules below. Let the i[th] valid assignment for α be given by $A_i(\alpha)$. Note that for some α, there are multiple valid assignments.
    a. If an empty circle has an arrow pointing at it, its outward arrow must point in the direction opposite of the incoming arrow (forming a path)
    b. All arrows must point at a node (i.e. it may not point out of the graph)
6. The weight w_α is given by $\sum_{i=1}^{Q} F(A_i(\alpha))$, where Q is the number of valid assignments for α and F is given by

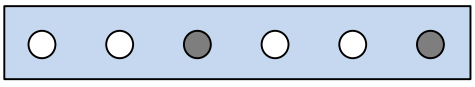
$$F\big(A(\alpha)\big) = \prod_{j\in R} \phi_j \prod_{k\in L} \beta_{k-1}$$

    As a reminder, ϕ_j and β_k are defined as the forward and backward transfer rates of each stage, and are defined in section 2.3.2.

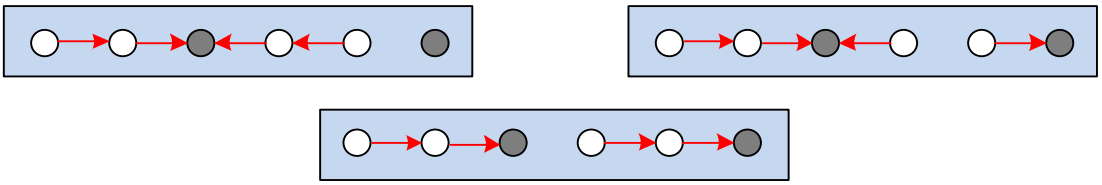
    In our definition for $F\big(A(\alpha)\big)$ above, R is the set of integer indices of those nodes which have an outward pointing arrow which points to the right, and L is the set of integer indices of those nodes which have an outward pointing arrow which points to the right.

Example

Consider the efflux signal product $\lambda_3\lambda_6$, where N=6. We wish to know $w_{(0,0,1,0,0,1)}$. We first draw our six circles and fill in circles 3 and 6.



Since nodes 1, 2, 4, and 5 are left unfilled, then each most have an outward pointing arrow. There are three valid arrow configurations shown below, given as $A_1(0,0,1,0,0,1)$, $A_2(0,0,1,0,0,1)$, and $A_3(0,0,1,0,0,1)$.



For $A_1(0,0,1,0,0,1)$, corresponding to the top left rectangle above, we have that $R = \{1,2\}$ and $L = \{4,5\}$. Thus $F\big(A_1(0,0,1,0,0,1)\big) = f_1 f_2 b_3 b_4$

Repeating this for A_2 ($R = \{2,3,5\}, L = \{4\}$) and A_3 ($R = \{1,2,4,5\}, L = \{\}$), we finally have that $w_{(0,0,1,0,0,1)} = \phi_1\phi_2\beta_3\beta_4 + \phi_1\phi_2\beta_3\phi_4 + \phi_1\phi_2\phi_3\phi_4$

Thus, if all of our phosphorelay domains lie on separate proteins, we have that

$$w_{(0,0,1,0,0,1)} = F_1F_2B_3B_4c_2c_3^2c_4 + F_1F_2B_3F_4c_2c_3^2c_5 + F_1F_2F_3F_4c_2c_3c_4c_5$$

As another sub-example, if our first four domains lie on the same protein, and the last two domains are on a second protein, then we have from our definitions of ϕ_i abd β_i that

$$w_{(0,0,1,0,0,1)} = F_1F_2B_3B_4c_4 + F_1F_2B_3F_4c_5 + F_1F_2F_3F_4c_5$$

B Full ODE for Complexing Model

```
d/dt(x1) = -x1*kappa+sigA*sigA_x1_transRate-
x1*x2P*x1_x2P_bindRate+x1_x2P*x1_x2P_unbindRate-
x1*Pi1*x1_Pi1_bindRate+x1_Pi1*x1_Pi1_unbindRate-
x1*x2*x1_x2_bindRate+x1_x2*x1_x2_unbindRate-(d+deg)*x1+L1*x1P
d/dt(x1P) = +x1*kappa-x1P*x2*x1P_x2_bindRate+x1P_x2*x1P_x2_unbindRate-
x1P*Pi1*x1P_Pi1_bindRate+x1P_Pi1*x1P_Pi1_unbindRate-(d+deg)*x1P-L1*x1P
d/dt(x1_x2P) = +x1_x2P_bindRate*x1*x2P-x1_x2P_unbindRate*x1_x2P-
x1_x2P_phosphotransferRate*x1_x2P+x1P_x2_phosphotransferRate*x1P_x2-(d+deg)*x1_x2P
d/dt(x1P_x2) = +x1P_x2_bindRate*x1P*x2-x1P_x2_unbindRate*x1P_x2-
x1P_x2_phosphotransferRate*x1P_x2+x1_x2P_phosphotransferRate*x1_x2P-(d+deg)*x1P_x2
d/dt(Pi1) = +sigA*sigA_Pi1_transRate-Pi1*x1P*x1P_Pi1_bindRate+x1P_Pi1*x1P_Pi1_unbindRate-
Pi1*x1*x1_Pi1_bindRate+x1_Pi1*x1_Pi1_unbindRate-(d+deg)*Pi1
d/dt(x1P_Pi1) = +x1P_Pi1_bindRate*x1P*Pi1-x1P_Pi1_unbindRate*x1P_Pi1-
x1P_Pi1_dephosphoRate*x1P_Pi1-(d+deg)*x1P_Pi1
d/dt(x1_Pi1) = +x1_Pi1_bindRate*x1*Pi1-
x1_Pi1_unbindRate*x1_Pi1+x1P_Pi1_dephosphoRate*x1P_Pi1-(d+deg)*x1_Pi1
d/dt(x1_x2) = +x1_x2_bindRate*x1*x2-x1_x2_unbindRate*x1_x2-(d+deg)*x1_x2
d/dt(x2) = +sigA*sigA_x2_transRate-x2*x1P*x1P_x2_bindRate+x1P_x2*x1P_x2_unbindRate-
x2*x3P*x2_x3P_bindRate+x2_x3P*x2_x3P_unbindRate-
x2*Pi2*x2_Pi2_bindRate+x2_Pi2*x2_Pi2_unbindRate-
x2*x1*x1_x2_bindRate+x1_x2*x1_x2_unbindRate-x2*x3*x2_x3_bindRate+x2_x3*x2_x3_unbindRate-
(d+deg)*x2+L2*x2P
d/dt(x2P) = -x2P*x1*x1_x2P_bindRate+x1_x2P*x1_x2P_unbindRate-
x2P*x3*x2P_x3_bindRate+x2P_x3*x2P_x3_unbindRate-
x2P*Pi2*x2P_Pi2_bindRate+x2P_Pi2*x2P_Pi2_unbindRate-(d+deg)*x2P-L2*x2P
d/dt(x2_x3P) = +x2_x3P_bindRate*x2*x3P-x2_x3P_unbindRate*x2_x3P-
x2_x3P_phosphotransferRate*x2_x3P+x2P_x3_phosphotransferRate*x2P_x3-(d+deg)*x2_x3P
d/dt(x2P_x3) = +x2P_x3_bindRate*x2P*x3-x2P_x3_unbindRate*x2P_x3-
x2P_x3_phosphotransferRate*x2P_x3+x2_x3P_phosphotransferRate*x2_x3P-(d+deg)*x2P_x3
d/dt(Pi2) = +sigA*sigA_Pi2_transRate-Pi2*x2P*x2P_Pi2_bindRate+x2P_Pi2*x2P_Pi2_unbindRate-
Pi2*x2*x2_Pi2_bindRate+x2_Pi2*x2_Pi2_unbindRate-(d+deg)*Pi2
```

```
d/dt(x2P_Pi2) = +x2P_Pi2_bindRate*x2P*Pi2-x2P_Pi2_unbindRate*x2P_Pi2-
x2P_Pi2_dephosphoRate*x2P_Pi2-(d+deg)*x2P_Pi2
d/dt(x2_Pi2) = +x2_Pi2_bindRate*x2*Pi2-
x2_Pi2_unbindRate*x2_Pi2+x2P_Pi2_dephosphoRate*x2P_Pi2-(d+deg)*x2_Pi2
d/dt(x2_x3) = +x2_x3_bindRate*x2*x3-x2_x3_unbindRate*x2_x3-(d+deg)*x2_x3
d/dt(x3) = +sigA*sigA_x3_transRate-x3*x2P*x2P_x3_bindRate+x2P_x3*x2P_x3_unbindRate-
x3*x4P*x3_x4P_bindRate+x3_x4P*x3_x4P_unbindRate-
x3*Pi3*x3_Pi3_bindRate+x3_Pi3*x3_Pi3_unbindRate-
x3*x2*x2_x3_bindRate+x2_x3*x2_x3_unbindRate-x3*x4*x3_x4_bindRate+x3_x4*x3_x4_unbindRate-
(d+deg)*x3+L3*x3P
d/dt(x3P) = -x3P*x2*x2_x3P_bindRate+x2_x3P*x2_x3P_unbindRate-
x3P*x4*x3P_x4_bindRate+x3P_x4*x3P_x4_unbindRate-
x3P*Pi3*x3P_Pi3_bindRate+x3P_Pi3*x3P_Pi3_unbindRate-(d+deg)*x3P-L3*x3P
d/dt(x3_x4P) = +x3_x4P_bindRate*x3*x4P-x3_x4P_unbindRate*x3_x4P-
x3_x4P_phosphotransferRate*x3_x4P+x3P_x4_phosphotransferRate*x3P_x4-(d+deg)*x3_x4P
d/dt(x3P_x4) = +x3P_x4_bindRate*x3P*x4-x3P_x4_unbindRate*x3P_x4-
x3P_x4_phosphotransferRate*x3P_x4+x3_x4P_phosphotransferRate*x3_x4P-(d+deg)*x3P_x4
d/dt(Pi3) = +sigA*sigA_Pi3_transRate-Pi3*x3P*x3P_Pi3_bindRate+x3P_Pi3*x3P_Pi3_unbindRate-
Pi3*x3*x3_Pi3_bindRate+x3_Pi3*x3_Pi3_unbindRate-(d+deg)*Pi3
d/dt(x3P_Pi3) = +x3P_Pi3_bindRate*x3P*Pi3-x3P_Pi3_unbindRate*x3P_Pi3-
x3P_Pi3_dephosphoRate*x3P_Pi3-(d+deg)*x3P_Pi3
d/dt(x3_Pi3) = +x3_Pi3_bindRate*x3*Pi3-
x3_Pi3_unbindRate*x3_Pi3+x3P_Pi3_dephosphoRate*x3P_Pi3-(d+deg)*x3_Pi3
d/dt(x3_x4) = +x3_x4_bindRate*x3*x4-x3_x4_unbindRate*x3_x4-(d+deg)*x3_x4
d/dt(x4) = +sigA*sigA_x4_transRate-x4*x3P*x3P_x4_bindRate+x3P_x4*x3P_x4_unbindRate-
x4*Pi4*x4_Pi4_bindRate+x4_Pi4*x4_Pi4_unbindRate-
x4*x3*x3_x4_bindRate+x3_x4*x3_x4_unbindRate-(d+deg)*x4+L4*x4P
d/dt(x4P) = -x4P*x3*x3_x4P_bindRate+x3_x4P*x3_x4P_unbindRate-
x4P*Pi4*x4P_Pi4_bindRate+x4P_Pi4*x4P_Pi4_unbindRate-(d+deg)*x4P-L4*x4P
d/dt(Pi4) = +sigA*sigA_Pi4_transRate-Pi4*x4P*x4P_Pi4_bindRate+x4P_Pi4*x4P_Pi4_unbindRate-
Pi4*x4*x4_Pi4_bindRate+x4_Pi4*x4_Pi4_unbindRate-(d+deg)*Pi4
d/dt(x4P_Pi4) = +x4P_Pi4_bindRate*x4P*Pi4-x4P_Pi4_unbindRate*x4P_Pi4-
x4P_Pi4_dephosphoRate*x4P_Pi4-(d+deg)*x4P_Pi4
d/dt(x4_Pi4) = +x4_Pi4_bindRate*x4*Pi4-
x4_Pi4_unbindRate*x4_Pi4+x4P_Pi4_dephosphoRate*x4P_Pi4-(d+deg)*x4_Pi4
d/dt(sigA) = 0
```