DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Quantitative Approaches to Cellular Information Processing and Metabolic Regulation

Citation

Dexter, Joseph Paul. 2018. Quantitative Approaches to Cellular Information Processing and Metabolic Regulation. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:41129219

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Quantitative approaches to cellular information processing and metabolic regulation

A DISSERTATION PRESENTED
BY
JOSEPH PAUL DEXTER
TO
THE COMMITTEE ON HIGHER DEGREES IN SYSTEMS BIOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
SYSTEMS BIOLOGY

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2018

©2018 – Joseph Paul Dexter
all rights reserved.

Dissertation advisors: Jeremy Gunawardena and Vamsi K. Mootha Joseph Paul Dexter

Quantitative approaches to cellular information processing and metabolic regulation

Abstract

Organisms of all levels of complexity must undertake complex information processing tasks. Diverse cellular and biochemical mechanisms are required to integrate multiple sources of information and to balance performance trade-offs, such as between speed and accuracy or robustness and fragility. This dissertation describes a series of quantitative analyses of cellular information processing, with particular attention given to the regulation of metabolism. Chapters 2 and 3 consider mechanisms for achieving concentration robustness in signal transduction. Chapter 2 develops a large compendium of reaction networks involving bifunctional enzymes, which are often positioned at key metabolic branch points and are experimentally associated with robust control. Using high-throughput algebraic analysis of this compendium, we demonstrate that bifunctional enzymes can implement five different forms of concentration robustness, and that the type of robustness is highly sensitive to biochemical details beyond bifunctionality. Chapter 3 identifies intermediate buffering in a three-component phospho-relay as a novel mechanism for concentration robustness and argues that such a mechanism accounts for robust inactivation of the yeast osmotic stress response. Chapter 4 reports an integrated computational and experimental analysis of production of the oncometabolite 2-hydroxyglutarate by mutant isocitrate dehydrogenase 1 (IDH1), which suggests that the clinically observed retention of a wild-type (WT) IDH1 allele in tumors is not due to a requirement for substrate channeling or substantial inter-subunit flux in WT/mutant IDH1 heterodimers. In Chapter 5 we examine the information processing capabilities of Ca^{2+}/calmodulin signaling and show that a straightforward equilibrium binding analysis can clarify longstanding questions about the control of smooth muscle contraction. Finally, Chapter 6 reports an experimental approach to investigate the limits of complex information processing in single cells. Resurrecting a classical body of literature on the behavior of unicellular organisms, we demonstrate that the giant ciliate *Stentor roeseli* engages in multi-step hierarchical sequences of avoidance behaviors. The *S. roeseli* avoidance response is distinct from other primitive forms of learning such as habituation and conditioning and is suggestive of complex decision-making by the organism. Throughout the dissertation, a common theme is the use of mathematical modeling to link biochemical form to physiological function and to generate experimentally testable predictions that are independent of hard-to-measure parameter values.

Contents

Listing of figures

To my parents.

Acknowledgments

Every chapter of this dissertation reports highly collaborative, interdisciplinary research. I therefore owe an enormous amount to each of my co-authors, especially Tathagata Dasgupta, Megan McClean, Sudhakaran Prabakaran, Yasemin Sancak, Matt Vander Heiden, and Pat Ward, without whom little to none of this work would have been possible. It is a particular pleasure to thank both of my advisors, Jeremy Gunawardena and Vamsi Mootha. Their keen insights and advice have shaped every aspect of my research, and I am tremendously grateful for the intellectual freedom they gave me as a Ph.D. student.

I have had the pleasure of being a member of both the Mootha and Gunawardena labs for the past four years. Although very different from each other, both environments have influenced my development as a scientist in countless positive ways. In addition to those listed above, I thank John Biddle, David Croll, Mohan Malleshaiah, Chris Nam, and Felix Wong for innumerable stimulating conversations and years of camaraderie. In addition, it was great fun to go through graduate school with the 11 other members of my Systems Biology cohort – Brendan Colón, Ivana Cvijovic, Charlie Fulco, Anna Green, Jonathan Gootenberg, Ed Hendel, Rebecca Herbst, Luis Hernandez-Nunez, Hector Medina Abraca, Naren Tallapragada, and Shirlee Wohl.

I have benefited from many helpful suggestions from the members of my dissertation advisory committee, Aldebaran Hofer, Mike Springer, and Matt Vander Heiden (again), and from John Higgins, who graciously agreed to serve as a final reader. I would also like to thank Joe Henrich for his terrific informal mentorship and for introducing me to the field of cultural evolution, and Tim Mitchison and Andrew Murray for so capably directing the Systems Biology Ph.D. program. Additionally, the administrative acumen of Kathy Buhl, Liz Pomerantz, Sam Reed, and Bryanna Vacanti has made my life in graduate school vastly smoother.

I received critical financial support for my Ph.D. research in the form of a National Science Foundation Graduate Research Fellowship (Grant No. GE1144152).

In 2014 I co-founded the Quantitative Criticism Lab (QCL), which has expanded into a surprisingly large initiative focused on research at the intersection of computation, biology, and the humanities. Although I do not discuss any of this work in my dissertation, it has been a major part of my intellectual life during graduate school. I am extremely grateful to everyone who has contributed to QCL's research, including Lizzy Adams, T.J. Bolt, Jorge Bonilla Lopez, James Brofos, Adriana Casarez, Tathagata Dasgupta (again), Alvin Deng, Jeff Flynt, Zen Grether, Ayelet Haimson Lushkov, Sally Heiter, Krithika Iyer, Ajay Kannan, Teddy Katz, M. Ski Krieger, Daniel Michael, Prathik Naidu, Leonard Neidorf, James Patterson, Max Rabinovich, Eric Rincon, Lea Schroeder, Ariane Schwartz, Nilesh Tripuraneni, and Michelle Yakubek, and to Joe Farrell, Stephen Hinds, and Dan Rockmore for their always trenchant advice. The development of QCL would have been impossible without initial seed money from the Neukom Institute for Computational Science and subsequent support from the American Council of Learned Societies, the Andrew W. Mellon Foundation, and the National Endowment for the Humanities. Finally, I thank my co-founder Pramit Chaudhuri for being a truly extraordinary friend and collaborator.

Many faculty and other mentors at Princeton have contributed invaluably to my development as a researcher, including David Botstein, Eva-Maria Collins, Kathy Crown, Jill Dolan, Denis Feeney, Andrew Feldher, Brooke Holmes, Megan McClean (again), and Tom Muir. Similarly, I am deeply appreciative of the unwavering support of numerous mentors from high school (or before), including Dylan Bate, Tom Charlton, Tom Ferenc, Sharon Jonynas, Angela Hutchins, the late Alice O'Hearn, Bill Parker, Jonathan Rice, and Kevin Surma.

My experience in graduate school has been enriched by the support of many friends in Boston, including (but absolutely not limited to!) Christina Chang, Apple Chawanthayatham, David Croll (again), Tathagata Dasgupta (yet again!), Hannah Diamond-Lowe, Delaney Granizo-Mackenzie, Vivian Guo, Mike Jin, Alec Lai, David Levonian, Satabhisha Mukhopadhyay, Iulia Neagu, Amol Punjabi, and Adam Sealfon, and others further away, including Maite Ballestero, Brianna Beehler, Chelle Beehler, Molly Bream, Jeffrey Cattel, Jonathan Cattel, Andy Charman, Pramit Chaudhuri (again), Ayelet Haimson Lushkov (again), Sam Jonynas, Zane Li, Elaine Liew, Tom Martin, Nic Orsi, Annie Ouyang, Heather Putnam, Rachel Putnam, Katie Svec, Nilesh Tripuraneni (again), and Vinay Tripuraneni.

It remains to thank my family for absolutely everything. I will always have fond memories of time spent in Boston with Sarah and Laura, and of every Cape May. My greatest debt is, of course, to my parents, to whom this dissertation is dedicated – thank you for 26 years of love and support.

1

Introduction

Information processing at the level of individual cells is an essential task for all organisms, regardless of complexity. Single-celled organisms must be able to sense their surrounding environment, transmit the information gained to different parts of the cell, and ultimately mount appropriate responses to external changes or threats. Multicellular organisms must accomplish all of the above and ensure appropriate cell-to-cell communication and coordination, with dire consequences for any breakdowns[27,227]. This dissertation presents a series of quantitative, systems-level analyses of diverse problems in cellular information processing, focusing especially on the regulation of energy production and consumption (metabolism).

Chapters 2 and 3 examine biochemical mechanisms by which an information processing network can be "robust," or insensitive to perturbation. Chapter 2 focuses on bifunctional enzymes, Chapter 3 on three-component post-translational modification relays. Chapter 4 considers heterodimerization and substrate channeling in the context of metabolic regulation in cancer, and Chapter 5 examines signal transduction involving Ca^{2+}/calmodulin (CaM). Finally, Chapter 6 describes an experimental approach to probe the information processing capabilities of unicellular eukaryotes focused on avoidance behaviors of the giant ciliate *Stentor roeseli*. Detailed introductions to each of these topics are provided in the individual chapters. The remainder of this general introduction surveys a few areas of broad interest or particular importance to the dissertation.

1.1 Biological robustness

Robustness has many meanings in biology. In the broadest sense, a biological system can be called robust if it is insensitive to external perturbations, such as changes in the environment or infectious disease, or to internal perturbations, such as genetic mutation or stochastic fluctuations in protein concentrations[114,141,226]. As suggested by these example perturbations, robustness can operate on vastly different timescales depending on the specific biological context, from millisecond or faster to evolutionary. Beyond straightforward mechanisms such as functional redundancy of genes or pathways[226,116,108], the information processing mechanisms that underlie robust responsiveness remain incompletely understood.

Given the polysemous nature of the term, it is important to develop precise quantitative definitions when analyzing the robustness of a specific biological process, or when a proposing a "design principle" that accounts for robustness in multiple related systems[190]. ("I know it when I see it" may be sufficient for obscenity, but not for systems biology.) Chapters 1 and 2 of this dissertation are concerned with a type of internal insensitivity referred to as concentration robustness, in which some quantity involving the concentrations of one or more species (such as proteins or metabolites) in a biochemical network remains constant at steady state. In the simplest case, termed absolute concentration robustness (ACR), the concentration of a single species is fixed[191,189]. The concept of ACR has been widely influential, both in systems biology and in other areas such as chemical reaction network theory[188,7].

Bifunctional enzymes catalyzing opposing post-translational modifications (such as phosphorylation and dephosphorylation of a single residue on a protein) are an important class of "paradoxical" biological components that bring antagonistic processes into close proximity[82]. Bifunctional enzymes are widely distributed, especially in bacterial signaling and metabolism but also in mammalian glycolysis (6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase) and inositol pyrophosphate signaling (5-InsP7 kinase/InsP8 phosphatase)[43,72]. It has been known for decades that some systems containing bifunctional enzymes exhibit experimental concentration robustness[186], leading to early proposals by LaPorte and Koshland and Russo and Silhavy that this robustness might be due specifically to bifunctionality[122,178]. Beginning with Batchelor and Goulian's analysis of the EnzV/OmpR two-component osmosensor in *Escherichia coli*[15], there have been numerous attempts to examine the bifunctionality/robustness association using mathematical or computational modeling, which has led to a general appreciation that the additional conservation law imposed by bifunctionality (i.e., that the total concentration of modifying enzyme must equal the total concentration of demodifying enzyme) is the basis for concentration robustness in many systems[189,48,43]. In Chapter 2, we examine a large compendium of bifunctional enzyme mechanisms and demonstrate that five different forms of concentration robustness can arise depending on the specific biochemical assumptions.

Despite this intensive attention, however, it is clear that bifunctionality is only a small part of the (concentration) robustness story. According to some estimates, more than 1,000 human metabolites may be maintained a robust concentration[51], many by mechanisms that remain to be elucidated. The primary result of Chapter 3 is the identification of a novel robustness mechanism - buffering by an intermediate component in a three-component phospho-relay - which explains aspects of the robust regulation of the *Saccharomyces cerevisiae* osmotic stress response. One implication of this result is that *E. coli* and yeast use contrasting biochemical mechanisms to achieve a common functional outcome of robust osmoregulation.

1.2 METABOLIC REGULATION

Organisms must integrate their metabolic networks with information processing capabilities to ensure efficient use of nutrients and appropriate response to environmental changes. Accordingly, several chapters of this dissertation involve quantitative analysis of metabolic regulation, with a particular interest in linking biochemical form and systems-level function. Bifunctional enzymes of the sort considered in Chapter 2 are often positioned at metabolic branch points. A classic example is isocitrate dehydrogenase kinase/phosphatase (IDHKP) in *E. coli* central carbon metabolism[121]. This bifunctional enzyme regulates the activity of isocitrate dehydrogenase (IDH) and therefore determines the partitioning of flux between the full TCA cycle and the anapleurotic glyoxylate bypass. The IDH/IDHKP system has been shown both theoretically and experimentally to exhibit concentration robustness[122,192,48], which may be critical for the adaptability of *E. coli* to growth on two-carbon substrates. Two IDHs very different from the *E. coli* enzyme, human isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2), are the subject of Chapter 4. Point mutations in these dimeric, $NADP^+$-dependent enzymes cause increased production of the "oncometabolite" 2-hydroxyglutarate (2-HG) in multiple cancers. In less than a decade, understanding of dysregulated IDH metabolism in cancer has progressed from identification of initial mutants by sequencing of patient-derived tumor samples to the credentialing of 2-HG as a biomarker for disease burden and development of an FDA-approved small-molecular inhibitor of mutant IDH2 for treatment of acute myeloid leukemia[158,235,21,44,237]. Despite this rapid progress, interesting mechanistic puzzles remain, one of which (the possibility of substrate channeling through the IDH1 heterodimer) is explored in Chapter 4. Finally, Chapter 5 considers the regulation of smooth muscle contraction, an energy-intensive process, by the multifunctional Ca^{2+}-binding protein CaM.

1.3 Limits of information processing by a single cell

Unicellular organisms are capable of remarkable responses to environmental changes. The *S. cerevisae* osmotic stress response, discussed in detail in Chapter 3, is a prime example. On timescales of just a few minutes, yeast activate the high-osmolarity glycerol (HOG) pathway, ultimately leading to production of the compatible solute glycerol and restoration of turgor pressure[88]. Yet single cells can undertake even more complicated information processing than is required for such feats of adaptability. If repeatedly forced to swim though a narrow glass tube, the unicellular ciliate *Paramecium* will escape faster and faster with each successive trial[8]. Another ciliate called *Spirostomum* can habituate to repeated tapping, such that after a while mild taps fail to elicit a response, and neuronal PC12 cells can similarly become habituated to certain chemical and electrical stimuli[9,145]. Although complex behavior is usually associated with animals that have nervous systems, these results demonstrate that elementary forms of learning occur even in single cells. Chapter 6 experimentally investigates the limits of unicellualr information processing by examining complex behavior in the giant ciliate *Stentor roeseli*. Using video microscopy and quantitative analysis, we resolve a longstanding controversy about *Stentor* behavior and characterize a hierarchical sequence of avoidance behaviors of unprecedented complexity for a single-cell eukaryote.

1.4 METHODOLOGICAL CONSIDERATIONS

Quantitative models involve parameters. In an informal sense, the number of parameters in a model tends to increase with level of detail - "thick," biologically rich models involve more parameters (and states) than "thin," abstract ones - and with scope (e.g., enzyme vs. pathway vs. genome-scale network)[74]. For the biochemical network models that are discussed throughout Chapters 2-5, these parameters are the rate constants for each reaction in the network.

Typical strategies for biological parameter estimation involve taking numerical values from the experimental literature or fitting the model to data. Both approaches have limitations. In the case of a mass-action biochemical model, values may be needed for the kinetics of intermediate reactions that are inaccessible to measurement. Moreover, it is not unusual for biochemical measurements to differ between papers by an order of magnitude or more. As described in Chapter 4, such a discrepancy involving mutant IDH1 prompted us to undertake our own experimental kinetic analysis as a prerequisite to further mathematical modeling. Finally, this approach to parameter estimation almost invariably requires extrapolation from *in vitro* data, even if the model is intended to describe *in vivo* physiology.

Fitting has its own perils. It can be all too easy to "overfit" a detailed biochemical model, so that it describes a particular dataset well but generalizes poorly. Furthermore, given the often limited availability of experimental data in mechanistic systems biology, opportunities for validation of a model independent of the data to which it was fit may be scarce. These concerns are exemplified by the famous quip, attributed to John von Neumann, that “with four parameters I can fit an elephant, and with five I can make him wiggle his trunk.” Although von Neumann’s joke is widely quoted in the systems biology literature[74,162,156,13], numerical simulation of dynamical models with fitted parameter values remains a stock-in-trade of the field.

A common theme of Chapters 2-5 of this dissertation is the use of exact mathematical analysis to circumvent the need for estimation of numerous parameters. In contrast to numerical simulation, analytical calculations can be undertaken with parameters treated symbolically, and therefore often lead to predictions or experimental strategies that are parameter-independent. In certain cases, such as the analysis of the Sln1-Ypd1-Ssk1 phospho-relay described in Chapter 3, *ad hoc* algebraic manipulations guided by biochemical intuition are sufficient. In other instances, more systematic mathematical machinery is required.

We make use of two such systematic approaches. Our analysis of bifunctional enzyme mechanisms in Chapter 2 relies on the method of Gröbner bases to calculate polynomial “invariants.” Steady-state mass-action analysis of a biochemical reaction network gives rise to a system of non-linear polynomial equations. In some instances, it is possible to combine the polynomials so as to eliminate a subset of variables, yielding a simplified expression in the remaining variables (an invariant). Gröbner bases provide a systematic way to do this elimination for non-linear polynomials and thus can be viewed as a generalization of Gaussian elimination for linear systems. As will be seen in Chapter 2, the algebraic structure of an invariant can often reveal a great deal about a system’s biochemical and functional characteristics. The mathematical details of Gröbner basis calculations are discussed in detail in Refs.[39,139], the latter of which is concerned specifically with their application to biochemistry.

A longstanding focus of research in the Gunawardena group has been to develop the linear framework, a graph-based approach to timescale separation in biochemical systems. Full mathematical details of the approach are described in numerous previous publications[75,147,3,55]. In Chapter 5, we show that several related models of Ca^{2+}/calmodulin signaling, which have been studied in the literature by numerical simulation, can be solved analytically using the linear framework, and we use an equilibrium binding analysis to propose parameter-independent model discrimination criteria.

2

Invariants reveal multiple forms of robustness in bifunctional enzyme systems

AUTHOR LIST

Joseph P. Dexter,* Tathagata Dasgupta,* and Jeremy Gunawardena

* J.P.D. and T.D. contributed equally.

CITATION

This chapter was published previously as: J.P. Dexter, T. Dasgupta, and J. Gunawardena, "Invariants reveal multiple forms of robustness in bifunctional enzyme systems" *Integrative Biology* 7 (2015) 883-894

CHAPTER SUMMARY

Experimental and theoretical studies have suggested that bifunctional enzymes catalyzing opposing modification and demodification reactions can confer steady-state concentration robustness to their substrates. However, the types of robustness and the biochemical basis for them have remained elusive. Here we report a systematic study of the most general biochemical reaction network for a bifunctional enzyme acting on a substrate with one modification site, along with eleven sub-networks with more specialized biochemical assumptions. We exploit ideas from computational algebraic geometry, introduced in previous work, to find a polynomial expression (an invariant) between the steady state concentrations of the modified and unmodified substrate for each network. We use these invariants to identify five classes of robust behavior: robust upper bounds on concentration, robust two-sided bounds on concentration ratio, hybrid robustness, absolute concentration robustness (ACR), and robust concentration ratio. This analysis demonstrates that robustness can take a variety of forms and that the type of robustness is sensitive to many biochemical details, with small changes in biochemistry leading to very different steady-state behaviors. In particular, we find that the widely-studied ACR requires highly specialized assumptions in addition to bifunctionality. An unexpected result is that the robust bounds derived from invariants are strictly tighter than those derived by *ad hoc* manipulation of the underlying differential equations, confirming the value of invariants as a tool to gain insight into biochemical reaction networks. Furthermore, invariants yield multiple experimentally testable predictions and illuminate new strategies for inferring enzymatic mechanisms from steady-state measurements.

2.1 INTRODUCTION

Cellular systems have a remarkable capacity to function correctly even as internal and external conditions change[114]. Understanding the mechanistic basis for this robustness has emerged as a major topic of interest in systems biology[114,4,123,190]. Such efforts have been hampered by the difficulty of developing precise, quantitative definitions of the various forms of biological robustness, and the lack of shared molecular or biochemical features between robust systems.

The study of enzymatic bifunctionality, however, has led to considerable progress in understanding a particular kind of robustness. Bifunctional enzymes can be divided into two broad categories, those that catalyze consecutive steps in a metabolic pathway and those that play a more regulatory role by catalyzing opposing modification and demodification of a single substrate. Little is currently known about the systems-level advantages conferred by the first group of bifunctional enzymes, and there is considerable need for further experimental and theoretical work on such enzymes. Regulatory bifunctional enzymes, in contrast, have been the subject of extensive study from a systems-level perspective and have led to new insights about robust biological control.

Experiments in bacteria and plants have revealed concentration robustness in several systems with bifunctional enzymes[186,122,15,152]. A key output of the system, usually the concentration of the substrate in a particular state of modification, was found to remain approximately constant despite changes in the expression levels of system components, including the substrate itself. In many cases, such robustness was maintained even in the face of 10-fold or greater changes in expression of system components. From consideration of these experiments, LaPorte and Koshland, working with the isocitrate dehydrogenase regulatory system, and Russo and Silhavy, working with the EnvZ/OmpR osmosensor, both in *Escherichia coli*, made early attempts to develop a mechanistic, quantitative understanding of the concentration robustness conferred by bifunctional enzymes[122,178]. Their work has been extended by a number of detailed theoretical studies of bifunctional enzyme systems[15,191,192,189,83,84,112,48,43,208]. In each of these studies, a specific biochemical reaction network was proposed to describe the system of interest, and mathematical calculations or numerical simulations were used to explain how concentration robustness might emerge from the reaction network. Other forms of "robustness," such as sensitivity to fluctuations in parameter values, have been considered in different contexts[123,205,173,126].

Regulatory bifunctional enzymes can be subdivided into four classes according to the active site configuration of the bifunctional enzyme and the quaternary structure of the substrate (Fig. 2.1). Bifunctional enzymes can either have separate, non-interacting active sites for each catalytic activity (e.g., a kinase domain and a distinct phosphatase domain) or a shared active site that can catalyze both activities[240,43,48]. Furthermore, a bifunctional enzyme can modify either a monomeric or multimeric substrate.

Multimeric substrates give rise to multi-site and avidity effects that require special attention. For instance, Hart *et al.* developed models of the nitrogen assimilation system in *E. coli* and the carbon fixation pathway in C_4 plants[83,84]. In both systems, they showed that mathematical robustness arises because the bifunctional enzymes have separate active sites and can therefore form ternary complexes with two units of the respective substrates. Similar analyses have been performed of the bifunctional uridylyltransferase/uridylyl-removing enzyme (UTase/UR)[220,208], also involved in nitrogen assimilation, for which there is now extensive experimental evidence of robustness[104,103]. In contrast, the isocitrate dehydrogenase (IDH) regulatory system in *E. coli* contains a bifunctional enzyme with a shared active site that acts on a dimeric substrate. In recent work we showed that, although the corresponding biochemical network does not exhibit mathematical robustness, there is a realistic parameter regime in which robustness is expected[48].

Our focus in this paper is on monomeric substrates modified by bifunctional enzymes that have either separate or shared active sites. Several such systems have been studied mathematically[15,191,192,189], of which the EnvZ/OmpR two-component osmosensor is the paradigmatic example[15,191]. There is

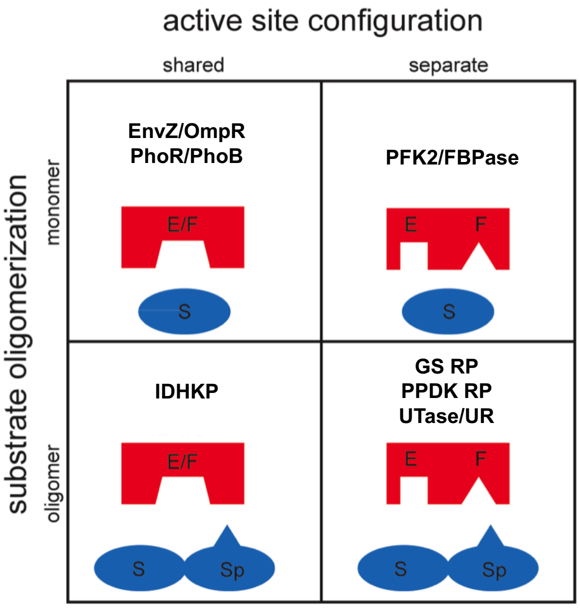


Figure 2.1: Classification of bifunctional enzyme systems by biochemical features. Bifunctional enzymes that catalyze opposing post-translational modifications can be sub-divided into four classes according to the active site configuration on the enzyme and the quaternary structure of the substrate. Known examples and a cartoon diagram are provided in each quadrant of the diagram for the appropriate system. This paper focuses on the robust behavior of systems in the two top quadrants (i.e., monomeric substrates regulated by a bifunctional enzyme).

strong experimental evidence for robustness both in EnvZ/OmpR and in other two-component systems such as PhoR/PhoB, which senses phosphorous levels[15,63,207]. No systematic characterization of robustness, however, has been undertaken for these systems[82]. To address this limitation, we developed and analyzed an extensive compendium of reaction networks involving a bifunctional enzyme (with either shared or separate active sites) acting on a monomeric substrate.

Our analysis identified five types of robust behavior: "robust upper bounds on concentration," in which the concentration of the modified or unmodified substrate form is held below a bound set only by the parameter values (i.e., the bound is independent of the total amount of substrate and enzyme); "robust two-sided bounds on the concentration ratio," in which the substrate concentration ratio is held between an upper bound and a lower bound determined only by the parameters;

"hybrid robustness," in which the same network exhibits a robust upper bound in concentration for certain parameter values and robust two-sided bounds on concentration ratio for other parameter values; "absolute concentration robustness" (ACR), in which the concentration of a substrate form is fixed at a single value set by the parameters[189]; and "robust concentration ratio," in which the substrate concentration ratio depends only on the parameters. The five behaviors are all variants of concentration robustness–some quantity involving the concentrations of the different species in the networks is held constant at steady state. We find that the type of concentration robustness is highly dependent on subtle but significant differences in biochemical mechanism. For instance, whether an enzyme-substrate ternary complex can form, and whether the reactions involving the ternary complex are symmetric or asymmetric, has a major impact on the robustness observed. Of particular interest, we demonstrate that ACR, which has dominated recent thinking about enzymatic bifunctionality, requires restrictive biochemical assumptions, and that more complex forms of robustness arise even with minimal assumptions beyond bifunctionality. Our results delineate the complexity of biochemical concentration robustness and suggest experiments that might distinguish one form of robustness from another.

The reaction networks that we study here all involve the interconversion of unmodified (S) and modified (S_P) substrate by a bifunctional regulatory enzyme (E). Assuming mass-action kinetics at steady state yields systems of non-linear polynomial equations in the concentrations of the free enzyme and substrate forms and various enzyme-substrate complexes. Our results are based on mathematical analysis of these equations with the parameters treated symbolically, thereby avoiding numerical simulations that require advance knowledge of parameter values. We use techniques from computational algebraic geometry, which we introduced in previous work, to systematically eliminate variables and to derive polynomial "invariants" involving only $[S]$ and $[S_P]$. (Throughout the paper, the notation $[X]$ signifies the concentration of X.) These invariants provide the essential insights into robust behavior. Interestingly, we find that invariants yield more accurate bounds than *ad hoc* manipulation of the underlying differential equations.

2.2 METHODS

2.2.1 REACTION NETWORK ASSUMPTIONS

Our networks apply to the modification and demodification of any small molecule or protein in which the modification is by attachment of a small-molecule moiety (phosphate, methyl, acetyl, etc.)[169]. Polypeptide modifications, such as ubiquitination and SUMOylation, involve more complicated biochemistry and are not modeled by our networks. As is standard, the modification donor (such as ATP in the case of phosphorylation) is assumed to be kept at constant concentration by background cellular processes and is therefore ignored as a dynamical variable. Each catalytic domain of the bifunctional enzyme is assumed to follow a standard Michaelis-Menten reaction scheme, which takes the form

$$E + S \rightleftharpoons ES \rightarrow E + S_P \tag{2.1}$$

for the forward modification, with reversible formation of an enzyme-substrate complex (ES) and irreversible formation of product. The limitations of these assumptions are reviewed further in the Discussion.

The most general reaction network (network 0) is shown as biochemical reactions in Fig. 2.2A and as a directed graph on the corresponding "complexes" (in the sense of Chemical Reaction Network Theory[112]) in Fig. 2.2B. The complexes are enumerated in Fig. 2.2C. Subsequent networks in Figs. 4 and 5 are shown as sub-graphs of the graph in Fig. 2.2B with the missing vertices and edges faded. In the text, the rate constants for mass-action kinetics are given in the notation $k_{i \to j}$ for the reaction corresponding to the edge $i \to j$ in the graph on the complexes. The two catalytic activities are not assumed to be independent of each other, so that, for instance, $k_{2\to 3} \neq k_{6\to 7}$.

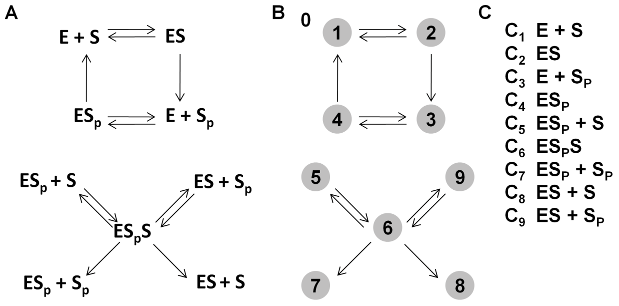


Figure 2.2: Full reaction network for a bifunctional enzyme with separate active sites acting on a monomeric substrate. **A** Biochemical reaction network. **B** Directed graph describing the reaction network in A, in which the nodes correspond to the "complexes"[112], as listed to the right. We used the notation $k_{i \to j}$ for the rate constant for the reaction corresponding to the edge $i \to j$ in the graph. **C** List of complexes in the reaction network.

2.2.2 ALGEBRAIC CALCULATIONS

All conclusions in the paper are based on algebraic analysis of the various chemical reaction networks at steady-state, assuming mass-action kinetics. We make no use of the Michaelis-Menten rate formula or any other form of approximation. Algebraic calculations were either done manually or using Mathematica 9.0 (Wolfram). An accompanying Mathematica notebook containing many of the calculations is provided as Supporting Material. Different numbering and notation had to be used in the notebook because of Mathematica conventions. The differences are described in tables in the notebook for each network considered in the paper, and these tables should be consulted when comparing expressions between the paper and the notebook. Gröbner basis calculations undertaken in Mathematica were performed as described previously by Manrai and Gunawardena[139].

2.3 RESULTS

2.3.1 ANALYSIS OF NETWORK 0

AD HOC DERIVATION.

We begin with consideration of the full network (network 0), which includes all possible reactions involving the bifunctional enzyme and the two substrate forms (Fig. 2.2). In the full network, the order of formation of the ternary complex is random, and E can act as either a modifying or demodifying enzyme when bound in any complex. As such, network 0 is the most general model possible of a bifunctional enzyme acting on a monomeric substrate and is completely symmetric. We find that either $[S]$ or $[S_P]$ may be robustly bounded depending on the relative values of the catalytic rate constants for the ternary complex ES_PS.

At steady state, the rate of formation of S_P from S must be balanced by the rate of formation of S from S_P, which leads to the following relationship between the three complexes in the network:

$$k_{2\to3}[ES] + k_{6\to7}[ES_PS] = k_{4\to1}[ES_P] + k_{6\to8}[ES_PS]. \tag{2.2}$$

Solving for the ternary complex in this expression yields

$$[ES_PS] = \frac{k_{4\to1}[ES_P] - k_{2\to3}[ES]}{k_{6\to7} - k_{6\to8}}. \tag{2.3}$$

Here we ignore the possibility that $k_{6\to7} = k_{6\to8}$, since such fine-tuning of rate constants is unlikely to occur naturally. Production and consumption of the ternary complex must also balance at steady state, leading to the following expression:

$$[ES_PS] = \frac{k_{5\to6}[S][ES_P] + k_{9\to6}[S_P][ES]}{k_{6\to5} + k_{6\to9} + k_{6\to7} + k_{6\to8}}. \tag{2.4}$$

Subtracting Eq. (2.3) from Eq. (2.4) eliminates the ternary complex and gives

$$A[ES_P] + B[ES] = 0, \tag{2.5}$$

where the coefficients A and B are

$$A = \frac{k_{5\to6}}{k_{6\to5} + k_{6\to9} + k_{6\to7} + k_{6\to8}}[S] - \frac{k_{4\to1}}{k_{6\to7} - k_{6\to8}} \tag{2.6}$$

and

$$B = \frac{k_{9\to6}}{k_{6\to5} + k_{6\to9} + k_{6\to7} + k_{6\to8}}[S_P] + \frac{k_{2\to3}}{k_{6\to7} - k_{6\to8}}. \tag{2.7}$$

There are two possibilities for the relative rates of modification and demodification in the ternary complex. Either the modification rate is greater and $k_{6\to8} < k_{6\to7}$, or the demodification rate is greater and $k_{6\to7} < k_{6\to8}$. If the modification rate is greater, $B > 0$, so to satisfy Eq. (2.5) we must have $A < 0$. This implies the following bound on $[S]$:

$$[S] < \frac{k_{4\to1}(k_{6\to5} + k_{6\to9} + k_{6\to7} + k_{6\to8})}{k_{5\to6}(k_{6\to7} - k_{6\to8})}. \tag{2.8}$$

When $k_{6\to7} < k_{6\to8}$, by an analogous argument we have the following bound on $[S_P]$:

$$[S_P] < \frac{k_{2\to3}(k_{6\to5} + k_{6\to9} + k_{6\to7} + k_{6\to8})}{k_{9\to6}(k_{6\to8} - k_{6\to7})}. \tag{2.9}$$

Thus, when the modification rate exceeds the demodification rate, $[S]$ is robustly held below a threshold that is determined only by the rate constants, whereas $[S_P]$ is bounded when the demodification rate is greater.

IMPROVED UPPER BOUNDS USING INVARIANTS.

The *ad hoc* argument above is straightforward and follows directly from the underlying differential equations. We turn now to an alternative method that is less direct but reveals that the upper bounds in Eqs. (2.8) and (2.9) are not optimal.

The method of Gröbner bases for polynomial systems, which can be thought of as a generalization of Gaussian elimination for linear systems, can be used to obtain the simplest possible algebraic expression between specified variables that holds at steady state. We have previously used Gröbner bases to study multisite protein phosphorylation and robustness in *E. coli* isocitrate dehydrogenase regulation[139,48], and readers interested in the mathematical details of the approach should consult the paper by Manrai and Gunawardena. From a Gröbner basis calculation on the polynomial system describing network 0 (the details of which are provided in the accompanying Mathematica notebook), it is possible to obtain the following polynomial expression in the steady-state concentrations of S and S_P:

$$\begin{aligned}(a_1[S]^2[S_P] + a_2[S][S_P]^2 + a_3[S]^2 + a_4[S_P]^2 \\ +a_5[S][S_P] + a_6[S] + a_7[S_P])[E] = 0,\end{aligned} \tag{2.10}$$

where $a_1, \cdots, a_7$ are algebraic combinations of the rate constants given in the Mathematica notebook. At this point, we restrict attention to those steady states in which each species has a positive steady-state concentration, which we call positive steady states. Since then $[E] \neq 0$, we can divide by $[E]$ in the expression above to obtain a polynomial in just the two substrate forms:

$$\begin{aligned}a_1[S]^2[S_P] + a_2[S][S_P]^2 + a_3[S]^2 + a_4[S_P]^2 \\ +a_5[S][S_P] + a_6[S] + a_7[S_P] = 0.\end{aligned} \tag{2.11}$$

Eq. (2.11) is an example of what we have called in previous work an “invariant”–a polynomial expression on selected variables that holds in any positive steady state[139]. This cubic invariant can be used to obtain robust upper bounds on $[S]$ or $[S_P]$ by the following argument.

The invariant defines an algebraic curve, which is a locus of points in the $([S], [S_P])$ plane that satisfy the invariant. An initial way to get a sense of the shape of such a curve is to consider its asymptotes (the tangent lines at infinity), which are given by the highest-degree terms in the polynomial. For Eq. (2.11) the highest-degree terms are

$$a_1[S]^2[S_P] + a_2[S][S_P]^2 = 0, \tag{2.12}$$

which can be factored into the following linear factors:

$$[S][S_P](a_1[S] + a_2[S_p]). \tag{2.13}$$

These factors determine the slopes of the possible asymptotes: there can be a vertical asymptote ($[S] = c_1$), a horizontal asymptote ($[S_P] = c_2$), or a diagonal asymptote ($a_1[S] + a_2[S_P] = c_3$). Lower-order terms in the invariant must be taken into account to characterize the asymptotes fully. This quick calculation, however, is helpful to provide a sense of the possibilities. As we will see, the vertical asymptote sets the robust upper bound for $[S]$, while the horizontal asymptote sets the bound for $[S_P]$. The existence of horizontal or vertical asymptotes for the invariant is therefore an indication of robust upper bounds. (The diagonal asymptote may be indicative of a robust ratio, which is not relevant for this network but becomes so for others.)

We now rigorously show the existence of robust upper bounds using a classical technique to study a curve by examination of its intersections with horizontal and vertical lines. If we take $[S]$ to be constant, which corresponds to the intersection of the curve with a vertical line in the $([S], [S_P])$

plane, then the invariant in Eq. (2.11) reduces to a quadratic equation in $[S_P]$:

$$A[S_P]^2 + B[S_P] + C = 0, \tag{2.14}$$

where $A = a_2[S] + a_4$, $B = a_1[S]^2 + a_5[S] + a_7$, and $C = a_3[S]^2 + a_6[S]$. The solutions of a quadratic equation are given by its discriminant, but it is easier to consider the normalized constant term C/A. If this term is negative, the quadratic equation must have one positive and one negative solution. Inspection of the expressions for $a_1, \cdots, a_7$ in the supplementary Mathematica notebook shows that $a_3, a_6 > 0$, $a_4, a_7 < 0$, a_5 is indeterminate in sign, and the sign of both a_1 and a_2 is the same as that of $k_{6\to 7} - k_{6\to 8}$. Hence, if $k_{6\to 7} - k_{6\to 8} > 0$, we have $a_1, a_2, a_3, a_6 > 0$ and $a_4, a_7 < 0$. For any positive steady state (for which $[S], [S_P] > 0$), we have that $C/A < 0$ if, and only if, $0 < [S] < -a_4/a_2$. In this case, Eq. (2.14) has only one positive solution for $[S_p]$.

To confirm that this expression sets an upper bound for $[S]$ it is necessary to show that no part of the curve in the positive quadrant has $[S] > -a_4/a_2$. In this region, the normalized constant term C/A is positive and equation Eq. (2.14) can have two positive solutions, two negative solutions, or two complex-conjugate solutions. We need to rule out the two positive solutions. We consider the intersection of the curve with a horizontal line, so that $[S_P]$ is taken to be constant. This procedure yields a quadratic equation in $[S]$:

$$A'[S]^2 + B'[S] + C' = 0, \tag{2.15}$$

in which $A' = a_1[S_P] + a_3$, $B' = a_2[S_P]^2 + a_5[S_P] + a_6$, and $C' = a_4[S_P]^2 + a_7[S_P]$. Assuming again that $k_{6\to 8} < k_{6\to 7}$, the normalized constant term C'/A' is negative for all $[S_P] > 0$. It follows that a horizontal line intersects the curve in at most one point in the positive quadrant. Since we already know that there is such a point with $0 < [S] < -a_4/a_2$, there cannot also be another point with $[S] > -a_4/a_2$. Hence, the curve in the positive quadrant is restricted to $0 < [S] < -a_4/a_2$, which sets an upper bound for $[S]$. Expanding the expressions for a_4 and a_2 given in the Mathematica notebook, we find that

$$[S] < -\frac{a_4}{a_2} = \frac{k_{4\to 1}(k_{6\to 5} + k_{6\to 7})}{k_{5\to 6}(k_{6\to 7} - k_{6\to 8})}. \tag{2.16}$$

This upper bound is strictly tighter than Eq. (2.8). In particular, the numerator of the bound in Eq. (2.16) does not contain the term $k_{6\to 9} + k_{6\to 8}$, which appears in the numerator of the bound obtained from *ad hoc* calculations. We return to this surprising discrepancy in the Discussion.

Fig. 2.3 shows a plot of the invariant in the $([S], [S_P])$ plane with all parameters set to 1 except $k_{6\to 7}$, which is set to 2 to ensure that $k_{6\to 8} < k_{6\to 7}$. The resulting curve has three disjoint lobes (blue curves). The branch that lies in the positive quadrant is confined to the left of the line $[S] = -a_4/a_2$ (red line). The upper bound in Eq. (2.8) is shown as the green line, revealing the striking discrepancy between the two bounds. The bound obtained from the invariant describes a vertical asymptote to the curve and is thus the best possible upper bound on $[S]$.

An analogous argument under the assumption that $k_{6\to 7} < k_{6\to 8}$ reveals a horizontal asymptote that gives the optimal upper bound for $[S_P]$:

$$[S_P] < -\frac{a_3}{a_1} = \frac{k_{2\to 3}(k_{6\to 8} + k_{6\to 9})}{k_{9\to 6}(k_{6\to 8} - k_{6\to 7})}. \tag{2.17}$$

As before, this bound is strictly tighter than Eq. (2.9). We see from this analysis that network 0 exhibits different behavior depending on the choice of certain parameter values.

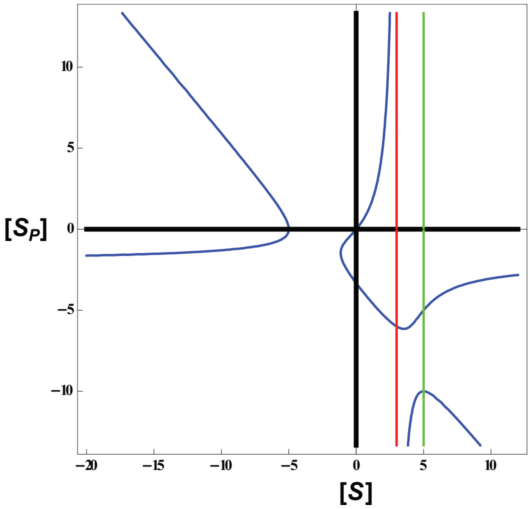


Figure 2.3: Asymptotic behavior of the cubic invariant. The cubic invariant describes a curve with three distinct lobes. Only portions of the second lobe correspond to positive $[S]$ and $[S_P]$, and the vertical asymptote in the first quadrant corresponds to the robust upper bound on $[S]$. The vertical red line indicates the bound calculated from the invariant in Eq. 2.16, and the vertical green line indicates the looser bound calculated in Eq. 2.8 by using the *ad hoc* procedure. To generate the plot, all parameters except $k_{6\to7}$ were set to 1; $k_{6\to7}$ was set to 2 to ensure that $[S]$ rather than $[S_P]$ is bounded.

2.3.2 OTHER SUB-NETWORKS WITH ROBUST UPPER BOUNDS (NETWORKS I, II, IV, AND VIII)

Fig. 2.4 shows eight related sub-networks that can be constructed by eliminating one or more reactions involving the ternary complex from the full network. Each of these networks contains all possible reactions involving binary enzyme-substrate complexes (nodes 1-4), but makes assumptions about ternary complex formation and catalytic activity. There are eight networks that meet these requirements because there are three possible binding orders for the ternary complex (random, ordered with S then S_P binding to E, ordered with S_P then S binding) and three possible catalytic activities (modification and demodification, modification only, demodification only). One combination (random binding, modification and demodification activity) is the full network.

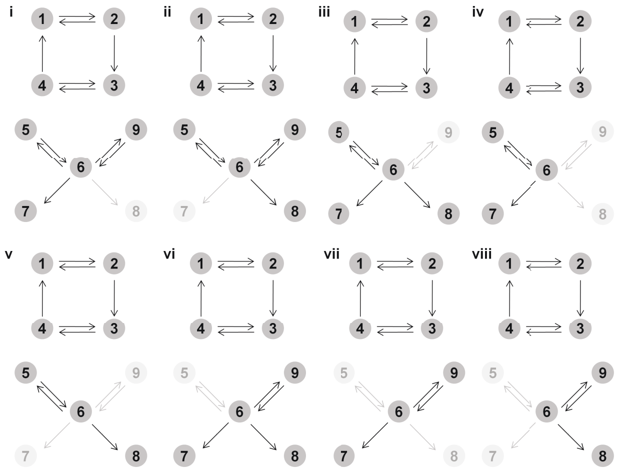


Figure 2.4: Sub-networks of the full network. Reaction networks are shown for the eight sub-networks of the full network that can be obtained by eliminating one or two reactions involving the ternary complex. Excluded reactions are shown in light gray; the actual reactions in the sub-network are black. Nodes are labeled as in Fig. 2.2.

The eight networks can be divided into symmetric pairs by reflection through the ternary complex (node 6). As such, networks i and ii, iii and vi, iv and viii, and v and vii in Fig. 2.4 are related by interchange of the roles of S and S_P. The algebraic analysis of one member of each pair can be applied readily to the other member. It is thus sufficient to explain the analysis of networks i, iii, iv, and v in detail. The results for all networks are summarized in Tables 2.1 and 2.2.

network	robustness	conditions	invariant
0	$[S] < \frac{k_{4\to1}(k_{6\to5}+k_{6\to7})}{k_{5\to6}(k_{6\to7}-k_{6\to8})}$	$k_{6\to8} < k_{6\to7}$	cubic
0	$[S_P] < \frac{k_{2\to3}(k_{6\to8}+k_{6\to9})}{k_{9\to6}(k_{6\to8}-k_{6\to7})}$	$k_{6\to7} < k_{6\to8}$	cubic
i	$[S] < \frac{k_{4\to1}(k_{6\to5}+k_{6\to7})}{k_{5\to6}k_{6\to7}}$	none	cubic
ii	$[S_P] < \frac{k_{2\to3}(k_{6\to8}+k_{6\to9})}{k_{6\to8}k_{9\to6}}$	none	cubic
iii	$[S] < \frac{k_{4\to1}(k_{2\to1}+k_{2\to3})(k_{6\to5}+k_{6\to7}+k_{6\to8})}{k_{5\to6}(k_{6\to7}(k_{2\to1}+k_{2\to3})-k_{2\to1}k_{6\to8})}$	$k_{2\to1}k_{6\to8} < k_{6\to7}(k_{2\to1}+k_{2\to3})$	quadratic
iii	$\frac{k_{1\to2}k_{2\to3}k_{6\to8}}{k_{3\to4}(k_{2\to1}k_{6\to8}-k_{6\to7}(k_{2\to1}+k_{2\to3}))} < \frac{[S_P]}{[S]}$ $< \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{6\to7}(k_{2\to1}+k_{2\to3}) < k_{2\to1}k_{6\to8}$ and $k_{4\to1}k_{6\to8}(k_{2\to1}+k_{2\to3}) <$ $(k_{4\to1}+k_{4\to3})(k_{2\to1}k_{6\to8}-k_{6\to7}(k_{2\to1}+k_{2\to3}))$	quadratic
iii	$\frac{k_{1\to2}k_{2\to3}k_{6\to8}}{k_{3\to4}(k_{2\to1}k_{6\to8}-k_{6\to7}(k_{2\to1}+k_{2\to3}))} > \frac{[S_P]}{[S]}$ $> \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{6\to7}(k_{2\to1}+k_{2\to3}) < k_{2\to1}k_{6\to8}$ and $(k_{4\to1}+k_{4\to3})(k_{2\to1}k_{6\to8}-k_{6\to7}(k_{2\to1}+k_{2\to3})) <$ $k_{4\to1}k_{6\to8}(k_{2\to1}+k_{2\to3})$	quadratic
iv	$[S] < \frac{k_{4\to1}(k_{6\to5}+k_{6\to7})}{k_{5\to6}k_{6\to7}}$	none	quadratic
v	$\frac{k_{1\to2}k_{2\to3}}{k_{2\to1}k_{3\to4}} < \frac{[S_P]}{[S]} < \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{4\to1}(k_{2\to1}+k_{2\to3}) < k_{2\to1}(k_{4\to1}+k_{4\to3})$	quadratic
v	$\frac{k_{1\to2}k_{2\to3}}{k_{2\to1}k_{3\to4}} > \frac{[S_P]}{[S]} > \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{2\to1}(k_{4\to1}+k_{4\to3}) < k_{4\to1}(k_{2\to1}+k_{2\to3})$	quadratic

Table 2.1: Summary of robust behavior in the networks (Part I). Table 1 gives the robust upper bound or robust two-sided bound on concentration ratio for networks 0 through v (numbered as in Figs. 2 and 4), along with the corresponding parameter conditions that must be satisfied. The degree of the underlying invariant for each network is also listed. Only the tighter bound is listed for networks for which the *ad hoc* derivation and the invariant yield different results. As indicated in the Methods, the numbering and notation for parameters differs between the paper and the accompanying Mathematica notebook. The tables in the notebook should be consulted to ensure correct translation between the notebook and the paper.

network	robustness	conditions	invariant
vi	$[S_P] < \frac{k_{2\to3}(k_{4\to1}+k_{4\to3})(k_{6\to7}+k_{6\to8}+k_{6\to9})}{k_{9\to6}(k_{6\to8}(k_{4\to1}+k_{4\to3})-k_{4\to3}k_{6\to7})}$	$k_{4\to3}k_{6\to7} < k_{6\to8}(k_{4\to1}+k_{4\to3})$	quadratic
vi	$\frac{k_{1\to2}(k_{4\to3}k_{6\to7}-k_{6\to8}(k_{4\to1}+k_{4\to3}))}{k_{4\to1}k_{3\to4}k_{6\to7}} < \frac{[S_P]}{[S]}$ $< \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{6\to8}(k_{4\to1}+k_{4\to3}) < k_{4\to3}k_{6\to7}$ and $(k_{2\to1}+k_{2\to3})(k_{4\to3}k_{6\to7}-k_{6\to8}(k_{4\to1}+k_{4\to3})) <$ $k_{2\to3}k_{6\to7}(k_{4\to1}+k_{4\to3})$	quadratic
vi	$\frac{k_{1\to2}(k_{4\to3}k_{6\to7}-k_{6\to8}(k_{4\to1}+k_{4\to3}))}{k_{4\to1}k_{3\to4}k_{6\to7}} > \frac{[S_P]}{[S]}$ $> \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{6\to8}(k_{4\to1}+k_{4\to3}) < k_{4\to3}k_{6\to7}$ and $k_{2\to3}k_{6\to7}(k_{4\to1}+k_{4\to3}) <$ $(k_{2\to1}+k_{2\to3})(k_{4\to3}k_{6\to7}-k_{6\to8}(k_{4\to1}+k_{4\to3}))$	quadratic
vii	$\frac{k_{1\to2}k_{4\to3}}{k_{3\to4}k_{4\to1}} < \frac{[S_P]}{[S]} < \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{4\to3}(k_{2\to1}+k_{2\to3}) < k_{2\to3}(k_{4\to1}+k_{4\to3})$	quadratic
vii	$\frac{k_{1\to2}k_{4\to3}}{k_{3\to4}k_{4\to1}} > \frac{[S_P]}{[S]} > \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	$k_{2\to3}(k_{4\to1}+k_{4\to3}) < k_{4\to3}(k_{2\to1}+k_{2\to3})$	quadratic
viii	$[S_P] < \frac{k_{2\to3}(k_{6\to8}+k_{6\to9})}{k_{6\to8}k_{9\to6}}$	none	quadratic
ix	$[S] = \frac{k_{4\to1}(k_{6\to5}+k_{6\to7})}{k_{5\to6}k_{6\to7}}$	none	linear
x	$[S_P] = \frac{k_{2\to3}(k_{6\to9}+k_{6\to8})}{k_{9\to6}k_{6\to8}}$	none	linear
xi	$\frac{[S_P]}{[S]} = \frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}$	none	linear

Table 2.2: Summary of robust behavior in the networks (Part II). Table 2 gives the robust upper bound, robust two-sided bound on concentration ratio, ACR constant, or robust concentration ratio for networks vi through xi (numbered as in Figs. 4 and 5), along with the corresponding parameter conditions that must be satisfied.

Network i (and the symmetric network ii) have random-order binding to the ternary complex. These are the only two networks aside from the full network that yield cubic invariants. The remaining networks in Fig. 2.4 all have quadratic invariants. The details of the invariants are provided in the accompanying Mathematica notebook.

The cubic invariant for network i has a similar structure to that of Eq. (2.11) (the invariant for network 0), but the pattern of signs is simpler. Using the same notation for the coefficients as in Eq. (2.11), we see from inspection of the notebook that $a_1, a_2, a_3, a_6 > 0$, $a_4, a_7 < 0$, and a_5 is indeterminate. It follows by a similar analysis that $[S]$ is robustly bounded above by $-a_4/a_2$:

$$[S] < -\frac{a_4}{a_2} = \frac{k_{4\to1}(k_{6\to5} + k_{6\to7})}{k_{5\to6}k_{6\to7}}. \tag{2.18}$$

This bound, however, holds for all parameter values. In contrast, network ii has a robust upper bound on $[S_P]$, which also holds irrespective of the values of the parameters.

Networks iii, iv, and v all have quadratic invariants, whose analysis is simpler than for the cubic invariants from networks 0, i, and ii. We turn first to network iv, which shows behavior similar to that of network i but for different reasons. Inspection of the quadratic invariant for network iv indicates that the highest-degree term is $[S][S_P]$, suggesting the possible existence of a vertical or horizontal asymptote associated with robust upper bounds. The invariant can be rewritten in the form

$$[S_P] = \frac{a_8[S]}{a_9 + a_{10}[S]}, \tag{2.19}$$

where $a_8, a_9 > 0$ and $a_{10} < 0$. The denominator of this expression becomes infinite when $[S] = -a_9/a_{10} > 0$, revealing a vertical asymptote for positive values of $[S]$. For both $[S]$ and $[S_P]$ to be positive, $[S]$ must be confined to the left of that asymptote, so that $0 < [S] < -a_9/a_{10}$. As such, $[S]$ is robustly bounded above:

$$[S] < \frac{k_{4\to 1}(k_{6\to 5} + k_{6\to 7})}{k_{5\to 6}k_{6\to 7}}. \tag{2.20}$$

The remaining networks iii and v exhibit different behaviors, which are studied in the next two sections.

2.3.3 Robust two-sided bounds on the concentration ratio (Networks v and vii)

Network v has a quadratic invariant, whose highest-degree terms are $[S]^2$ and $[S][S_P]$. This suggests the possible existence of a vertical asymptote and a diagonal asymptote, the former of which may have an accompanying robust upper bound. A vertical asymptote does exist, but in this case it does not imply an upper bound. Rearrangement of the quadratic invariant yields the following expression:

$$\frac{[S_P]}{[S]} = \frac{a_{11} + a_{12}[S]}{a_{13} + a_{14}[S]}, \tag{2.21}$$

where a_{11}, a_{12}, a_{13}, and a_{14}, which are given in the supplemental notebook, are all positive. The denominator of the expression on the right of Eq. (2.21) becomes infinite at $[S] = -a_{13}/a_{14} < 0$, revealing a vertical asymptote. This asymptote plays no physiological role, however, because it occurs at a negative value of $[S]$.

The expression on the right of Eq. (2.21) is defined and positive for all non-negative value of $[S]$. It has the value a_{11}/a_{13} when $[S] = 0$ and tends to the value a_{12}/a_{14} as $[S]$ gets large. From calculation of the derivative, it is clear that the expression decreases monotonically as $[S]$ increases if $a_{12}/a_{14} < a_{11}/a_{13}$, whereas the expression increases monotonically if $a_{11}/a_{13} < a_{12}/a_{14}$. In either case, the concentration ratio $[S_P]/[S]$ is trapped between two-sided robust upper bounds.

Expanding the expressions for a_{11}, a_{12}, a_{13}, and a_{14} we find that, if

$$k_{4\to1}(k_{2\to1} + k_{2\to3}) < k_{2\to1}(k_{4\to1} + k_{4\to3}), \tag{2.22}$$

then

$$\begin{gathered}\frac{a_{11}}{a_{13}} = \frac{k_{1\to2}k_{2\to3}}{k_{2\to1}k_{3\to4}} < \frac{[S_P]}{[S]} < \\ \frac{k_{1\to2}k_{2\to3}(k_{4\to1} + k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1} + k_{2\to3})} = \frac{a_{12}}{a_{14}}.\end{gathered} \tag{2.23}$$

Similarly, if

$$k_{2\to1}(k_{4\to1} + k_{4\to3}) < k_{4\to1}(k_{2\to1} + k_{2\to3}), \tag{2.24}$$

then

$$\begin{gathered}\frac{a_{11}}{a_{13}} = \frac{k_{1\to2}k_{2\to3}}{k_{2\to1}k_{3\to4}} > \frac{[S_P]}{[S]} > \\ \frac{k_{1\to2}k_{2\to3}(k_{4\to1} + k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1} + k_{2\to3})} = \frac{a_{12}}{a_{14}}.\end{gathered} \tag{2.25}$$

These results suggest that certain biochemical combinations are unsuitable for implementing a robust upper bound on one of the substrate forms. In particular, robust upper bounds are lost when the only substrate form produced from the ternary complex is required to bind second in the formation of the ternary complex (e.g., in network v only S_P is produced from the ternary complex, but the binding order is S then S_P). Interestingly, we find that such conditions do lead to bounds on the concentration ratio of the substrate modification forms, a type of robustness that has not been noted previously. As such, this asymmetry appears to have important consequences for robustness.

2.3.4 Hybrid robustness (Networks iii and vi)

We turn now to network iii. Inspection of its invariant shows that it resembles that for network v, analyzed in the previous section. It has the same terms, and the invariant can be reorganized into the same form as Eq. (2.21). There is, however, one significant difference. We have $a_{11}, a_{12}, a_{13} > 0$, but a_{14} is of indeterminate sign. We must therefore distinguish two possibilities. If $a_{14} < 0$, there is a vertical asymptote at a positive value of $[S]$ and hence a robust upper bound with $0 < [S] < -a_{13}/a_{14}$. In this parameter region, network iii behaves like network iv. If $a_{14} > 0$, however, the vertical asymptote occurs at a negative, unphysiological value of $[S]$. Network iii then behaves like network v and exhibits robust two-sided bounds on the concentration ratio. (We ignore the case $a_{14} = 0$, again on the grounds that such fine tuning is unlikely to occur.) Network iii and (by symmetry) network vi therefore exhibit hybrid robustness, with different robust behaviors in different regions of the parameter space. The precise details of the parameter regions and the bounds are given in Tables 2.1 and 2.2.

2.3.5 Absolute concentration robustness (Networks ix and x)

The full network and the eight sub-networks in Fig. 2.4 all contain a complete set of reactions involving the binary enzyme-substrate complexes. We now consider the two networks in Fig. 2.5A ,in which only one binary enzyme-substrate complex is formed and, accordingly, the ternary complex is formed in only one way. These networks are symmetric under interchange of S and S_P, and it is only necessary to analyze one of them. The Gröbner basis (or a simple *ad-hoc* calculation) indicates that network ix has a linear invariant of the form

$$[S] = \frac{k_{4\to1}(k_{6\to5} + k_{6\to7})}{k_{5\to6}k_{6\to7}}. \tag{2.26}$$

Eq. (2.26) shows ACR in S, which can also be deduced using the Shinar-Feinberg Theorem[189]. The deficiency of the network is one, and a pair of non-terminal nodes (4 and 5) differ by only the single species S. The network thus exhibits ACR in S.

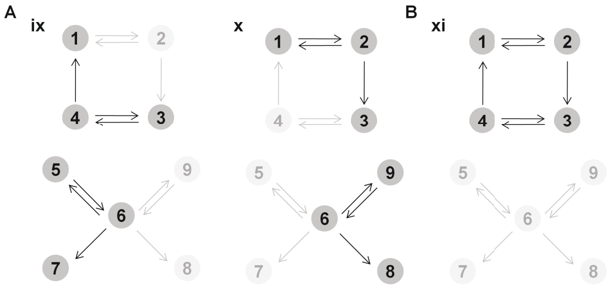


Figure 2.5: Reaction networks leading to ACR and robust ratios. **A** Reaction networks that exhibit ACR in S or S_P. **B** Reaction network involving only binary complexes that exhibits a robust ratio of the substrate forms. This network is the "full network" for a bifunctional enzyme with a shared active site acting on a monomeric substrate. Excluded reactions are shown in light gray; the actual reactions in the sub-network are black. Nodes are labeled as in Fig. 2.2.

We note that very small changes in reaction network structure can mean the difference between ACR and a robust upper bound. For instance, the only difference between network ix (which has ACR) and network iv (which has a robust upper bound) is the absence of node 2.

2.3.6 ROBUST CONCENTRATION RATIO (NETWORK XI)

Up to this point, our focus has been on bifunctional enzymes that have separate, non-interacting active sites that are far enough apart to permit the simultaneous binding of two substrate molecules. Many regulatory bifunctional enzymes, however, have only a single active site for both activities or have active sites that are too close together to permit simultaneous binding[240,10]. Such enzymes are not able to form a ternary complex when acting on a monomeric substrate and therefore give rise to a full network that contains only four nodes (Fig. 2.5B). As shown in the accompanying Mathematica notebook, the following linear invariant in $[S]$ and $[S_P]$ can be derived using a Gröbner calculation:

$$\begin{aligned} &k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})[S] \\ &-k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})[S_P]=0. \end{aligned} \tag{2.27}$$

Rearrangement of Eq. (2.27) yields

$$\frac{[S_P]}{[S]}=\frac{k_{1\to2}k_{2\to3}(k_{4\to1}+k_{4\to3})}{k_{3\to4}k_{4\to1}(k_{2\to1}+k_{2\to3})}. \tag{2.28}$$

Eq. (2.28) indicates that the relative amounts of modified and unmodified substrate will stay the same in all positive steady states. This ratio is a form of robustness that has not been noted before and could indicate a previously unrecognized advantage conferred by certain bifunctional enzyme systems.

2.3.7 Absence of ACR in networks other than ix and x

It is logically possible that a network might exhibit ACR in addition to another form of robustness such as an upper bound on concentration. We therefore checked that there is no possibility of ACR in any of the networks in the compendium besides ix and x. We find that, if a network has an invariant involving both S and S_P, then neither of these variables can exhibit ACR. For instance, if S exhibits ACR, then substituting for the absolute value of $[S]$ in the invariant imposes a polynomial constraint on $[S_P]$, which in turn can take only a finite number of values. However, the total amount of substrate in the system can be made arbitrarily large. Since the concentrations of enzyme-substrate complexes are limited by the total amount of enzyme, the only way for this to happen is if either $[S]$ or $[S_P]$ becomes arbitrarily large. This contradiction shows that S cannot exhibit ACR. Accordingly, neither the full network nor any of the networks in Fig. 2.4 exhibit ACR in either variable.

2.3.8 Experimental tests

To the best of our knowledge, there are no existing experimental demonstrations of a robust substrate ratio in a bifunctional enzyme system. As such, identification and characterization of bifunctional enzyme systems satisfying the assumptions of network xi should be of particular future interest. The two-component EnvZ/OmpR osmoregulatory system in *E. coli* does meet the basic assumptions of quadrant II (shared active site, monomeric substrate) but has been shown experimentally to exhibit ACR[15]. Imposition of additional assumptions on network xi, including initial autophosphorylation of EnvZ, can lead to networks with ACR[15,191].

The framework developed here also offers the attractive possibility of distinguishing bifunctional enzyme mechanisms by measurements undertaken at steady state, without use of advanced biochemical or structural techniques and without the need for fitting parameter values to data. Here we outline a candidate experimental strategy. Our suggested approach involves measurement of $[S]$ and $[S_P]$ following incubation *in vitro* of a purified bifunctional enzyme with its substrate, under conditions in which the system can reach a steady state[139], and testing whether the appropriate invariant is satisfied. Quantification of $[S]$ and $[S_P]$ can be accomplished by western blotting with antibodies specific for each substrate form or by mass spectrometry with an exogenous heavy-isotope or internal peptide standard[143,67,203]. Mass spectrometry has become a standard method for detection of post-translational modifications and has been used to quantify modification form distributions in several systems, including CDK1/2 and ERK[138,143,168]. Because the invariants found here do no depend on the total amounts of enzyme or substrate, these can be chosen arbitrarily over a wide range, without the need for high precision in setting up the reaction conditions. This feature also permits the substrate to be taken in sufficient excess over the enzyme so that dissociation of the enzyme-substrate complexes, which is difficult to prevent under normal experimental conditions, does not introduce much error into the measurement of $[S]$ and $[S_P]$. An advantage of bifunctional enzymes for such "systems biochemistry" is that it is not necessary to incubate two competing enzymes. Although straightforward in principle, simultaneous incubation can lead to technical complications in practice because of unexpected interactions between the two enzymes *in vitro*.

The statistical issues that arise in testing invariants against experimental data have been addressed by Harrington *et al.*[81], who built upon our previous work[139]. Their technique involves transformation of model variables so that the locus of points satisfying the invariant lies on a hyperplane regardless of parameter values. The transformed input data, which is obtained from the measured $[S]$ and $[S_P]$ values, can then be deemed consistent or inconsistent with a given steady-state invariant depending on distance from this hyperplane. Such parameter-independent model assessment enables confirmation or rejection of robustness even in systems for which measurements of rate constants have not been made.

A natural first experiment would be to undertake such measurements for a well-studied system that satisfies the assumptions of the full model (network 0), such as the mammalian 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase involved in glycolytic regulation, and confirm that Eq. (2.11) is satisfied. It might eventually be possible to make mechanistic predictions about poorly characterized bifunctional enzyme systems by measuring $[S]$ and $[S_P]$ and determining which invariant in the compendium holds. For instance, measurement of a robust ratio could suggest that a regulatory enzyme has a shared active site even in the absence of detailed structural data.

2.4 Discussion

Modification/demodification bifunctional enzymes are found in diverse cellular networks. Many of these networks regulate core transitions, such as flux partitioning at a metabolic branch point, which require coherent regulation. As such, bifunctional enzymes are used to maintain robust control of network output. Our analysis has confirmed that bifunctionality confers robustness to total enzyme levels regardless of mechanistic details. As we have discussed previously[234,43], this robustness can have important physiological consequences, including preventing incoherent behavior in tissues due to cell-to-cell variations in protein expression levels[193].

We have also identified several further insights, including that it is an oversimplification to think of biochemical concentration robustness as a single concept. Rather, it appears that multiple types of robust control can be implemented by biochemical reaction networks involving bifunctional enzymes. We have identified the mechanistic requirements for achieving five different types of robustness. The invariant involving $[S]$ and $[S_P]$ provides a general means to determine the type of robustness that a network exhibits. Furthermore, the type of robustness is highly dependent on the biochemical mechanism. Adding or removing just a single node from a reaction network can radically change the type of robustness. Additionally, a single network may exhibit different types of robustness depending on the numerical values of its parameters, as in networks iii and vi. We now consider several extensions and implications of the analysis, including the possibility of engineering biochemical robustness, treatment of non-Michaelis-Menten reaction mechanisms, and the advantages of adopting an "invariant-centric" viewpoint in theoretical biochemistry.

Engineered robustness. Section 3.8 outlines a minimal set of experiments to validate the core predictions of the invariants. Once this validation has been achieved, it might be possible to use our results to design modified or synthetic networks that exhibit a particular type of robustness. For instance, modification of the *E. coli* glyoxylate bypass regulatory machinery could yield a novel system exhibiting a robust substrate ratio. Reversible phosphorylation of IDH by the bifunctional isocitrate dehydrogenase kinase/phosphate (IDHKP) partitions flux between the glyoxylate bypass and the full TCA cycle[122,48]. *E. coli* IDHKP has a single active site for both its kinase and phosphatase activities[240], and IDH forms an obligate homodimer[95], which places it outside the scope of the networks in our compendium. However, mutation of one of the IDH phosphorylation sites, which potentially could be achieved by constructing a tandem dimer of one wild-type and one mutant subunit[29], would make it a functional monomer. This modified system would be described by network xi and therefore exhibit a robust ratio (i.e., constant $[IDH]/[IDH_P]$), whereas our previous steady-state analysis predicted different robustness for the wild-type network[48]. Biochemical analysis of the mutant system would thus provide a stringent test of a key prediction from the compendium.

Treatment of non-Michaelis-Menten reaction networks. We have assumed here that both forward and reverse modification follow the Michaelis-Menten reaction scheme. This is the standard assumption in the literature and is nearly universally followed. As we have pointed out elsewhere[234,43,77], however, if the enzyme-substrate complex can release product, on thermodynamic grounds the product must be able to rebind the enzyme. Michaelis and Menten could assume a simplified reaction scheme because they measured reaction rates in the absence of product[76]. In the context of modification and demodification cycles, each enzyme may encounter substantial amounts of product with some attendant degree of rebinding. A more realistic scheme is

$$E + S \rightleftharpoons ES \rightarrow EP \rightleftharpoons E + P, \tag{2.29}$$

which allows for product rebinding while still being irreversible overall, as would be expected under many physiological conditions for forward post-translational modification and reverse demodification. In previous work, we have developed mathematical methods for analyzing systems with such physiologically-realistic reaction schemes, allowing for multiple intermediates and branching pathways[234,43,77]. For a bifunctional enzyme system, there is still an invariant between $[S]$ and $[S_P]$ even when modification and demodification take place with arbitrarily complicated reactions schemes. This invariant, however, can become very complicated, with its degree increasing with the complexity of the reaction mechanism. In particular, the invariant may no longer be cubic, and analysis of its shape by the methods used for network o becomes more difficult. The analysis presented here, despite being restricted to the classical Michaelis-Menten scheme, should be considered an essential first step towards the technically more challenging problem of dealing with more realistic biochemistry. We stress, moreover, that analysis of reaction schemes beyond Michaelis-Menten is still in its infancy, and the vast majority of studies in the literature continue to assume Michaelis-Menten kinetics.

Invariants in biochemical analysis. Our analysis of the reaction network compendium demonstrates that invariants can be a helpful tool for studying biochemical systems at steady state. For each of the twelve networks, the invariant in $[S]$ and $[S_P]$ reveals the kind of robustness that the network exhibits. Cubic invariants are found for networks 0, i, and ii, in which the ternary complex is formed by random-order binding. Quadratic invariants are found for networks iii to viii, in which the ternary complex is formed by ordered binding. Linear invariants are found for networks ix and x, in which only one of the binary complexes is formed, and for network xi, in which no ternary complex is formed. Invariants involving both $[S]$ and $[S_P]$ rule out the possibility of ACR but allow for robust upper bounds on concentration (networks 0, i, ii, iv and viii), robust two-sided bounds on concentration ratio (networks v and vii), hybrid behavior between these two types of robustness (networks iii and vi), or a robust concentration ratio (network xi). ACR arises only for the two networks in which the invariant involves just one of the substrate forms (networks ix and x).

ACR has been widely discussed in the literature following the pioneering insights of Shinar and Feinberg[190]. We found that asymmetry in the reactions involving the binary complexes is required for ACR. In the two networks with ACR (ix and x in Fig. 2.5A), the bifunctional enzyme forms a ternary complex but can only form a binary complex with S or S_P, not both. To the best of our knowledge, no known biochemical networks exhibit these characteristics. It is possible, however, that such a network structure could be implemented if each active site on the bifunctional enzyme had a massively different affinity for the substrate depending on whether the other site was already occupied. This scenario would be analogous to the avidity effect observed for bifunctional enzymes acting on multimeric substrates[83]. It might also be possible to modulate relative affinity artificially with a small molecule, suggesting the possibility that drugs could be developed to promote or interfere with robust control.

Despite the limited number of networks that give rise to ACR, ACR is a useful conceptual framework for understanding one common type of biological robustness. To that end, demonstrations of approximate ACR (i.e., when the steady-state concentration of a variable is only approximately, not exactly, constant) have been made in several networks, including network iv[192]. To show approximate ACR, Shinar *et al.* assumed that the substrate is present in large excess over the bifunctional enzyme, $S_T \gg E_T$, and that total substrate is large compared to certain aggregated parameters. (They proposed network iv as a model of the IDH/IDHKP system, for which there is experimental evidence that $S_T \gg E_T$[23,151].) In our context, approximate ACR can be obtained by substituting

$$[S_P] \approx S_T - [S], \tag{2.30}$$

which holds when enzyme-substrate complexes can be ignored ($S_T \gg E_T$), into Eq. (2.19). It is then possible to show using a Taylor expansion that $[S]$ is approximately constant when S_T is sufficiently large. The invariants we derived can therefore be used as starting points for investigating approximate ACR in many specific systems.

In addition to providing a systematic mathematical language for classifying the different types of robustness, invariants also yield tighter bounds on the robustness than we were able to obtain from *ad hoc* algebraic manipulation of the underlying differential equations (Fig. 2.3). We were puzzled by this unexpected discrepancy. Our best explanation for it is that the method of Gröbner bases distills all the algebraic information that is present in the underlying equations and yields the simplest possible algebraic relationship between $[S]$ and $[S_P]$. It appears that *ad hoc* manipulation of the equations, while intuitively reasonable in terms of our biochemical understanding, may not take into account all the information in the system as a whole. It is possible that some sequence of manipulations not motivated by biochemical understanding might yield the optimal bounds. We were unable to find any such sequence, however, and the ability to pick out the variables of interest (in this case $[S]$ and $[S_P]$) and to eliminate all others using Gröbner bases provides a powerful capability for analyzing biological systems at steady state.

The difficulty with using invariants is calculating them. The method of Gröbner bases can always be used in principle, but in practice it becomes computationally infeasible for networks that are only slightly more complicated than the ones studied here. This is particularly so when parameters are treated symbolically, as we have done. In other work, we have developed efficient methods for generating a restricted class of invariants[112]. For modification and demodification systems, such as the Goldbeter-Koshland loop, invariants can be calculated using the so-called linear framework[75], which allows invariants to be written down irrespective of the complexity of the enzyme reaction mechanisms[234,43]. Software packages designed specifically for computational algebraic geometry, such as Singular and Macaulay2[46,70], have also enabled efficient calculation of invariants for more complicated polynomial systems. Development of new methods for calculating invariants remains an important challenge. Efficient methods could allow for steady-state algebraic analysis of entire metabolic or signaling networks, complementing existing probabilistic landscape approaches[126]. An "invariant-centric" viewpoint should enable a deeper investigation of robustness in biochemical networks than has been possible previously.

2.5 Acknowledgments

We thank Yasemin Sancak and Sudhakaran Prabakaran for comments on the manuscript, and Jeremy Owen, David Croll, and Tom Muir for scientific discussions. This work was supported by National Science Foundation Grant 0856285 (to J.G.). J.P.D. was supported in part by a National

Science Foundation Graduate Research Fellowship.

3

Robust network structure of the Sln1-Ypd1-Ssk1 three-component phospho-relay prevents unintended activation of the HOG MAPK pathway in *Saccharomyces cerevisiae*

AUTHOR LIST

Joseph P. Dexter, Ping Xu, Jeremy Gunawardena, and Megan N. McClean

CITATION

This chapter was published previously as: J.P. Dexter, P. Xu, J. Gunawardena, and M.N. McClean, "Robust network structure of the Sln1-Ypd1-Ssk1 three-component phospho-relay prevents unintended activation of the HOG MAPK pathway in *Saccharomyces cerevisiae*," *BMC Systems Biology* 9 (2015) 17

CHAPTER SUMMARY

Background:The yeast *Saccharomyces cerevisiae* relies on the high-osmolarity glycerol (HOG) signaling pathway to respond to increases in external osmolarity. The HOG pathway is rapidly activated under conditions of elevated osmolarity and regulates transcriptional and metabolic changes within the cell. Under normal growth conditions, however, a three-component phospho-relay consisting of the histidine kinase Sln1, the transfer protein Ypd1, and the response regulator Ssk1 represses HOG pathway activity by phosphorylation of Ssk1. This inhibition of the HOG pathway is essential for cellular fitness in normal osmolarity. Nevertheless, the extent to and mechanisms by which inhibition is robust to fluctuations in the concentrations of the phospho-relay components has received little attention.

Results: We established that the Sln1-Ypd1-Ssk1 phospho-relay is robust—it is able to maintain inhibition of the HOG pathway even after significant changes in the levels of its three components. We then developed a biochemically realistic mathematical model of the phospho-relay, which suggested that robustness is due to buffering by a large excess pool of Ypd1. We confirmed experimentally that depletion of the Ypd1 pool results in inappropriate activation of the HOG pathway.

Conclusions: We identified buffering by an intermediate component in excess as a novel mechanism through which a phospho-relay can achieve robustness. This buffering requires multiple components and is therefore unavailable to two-component systems, suggesting one important advantage of multi-component relays.

3.1 BACKGROUND

The high-osmolarity glycerol (HOG) pathway (Fig. 3.1) of the budding yeast *Saccharomyces cerevisiae* mediates cellular response to increased external osmolarity[88,180]. A key component of the HOG pathway is a mitogen-activated protein (MAP) kinase cascade. Within the kinase cascade, the MAP3Ks Ssk2 and Ssk22 phosphorylate the MAP2K Pbs2, which in turn phosphorylates the MAP kinase Hog1. Phospho-Hog1 then regulates transcriptional and metabolic changes that increase production and accumulation of the compatible solute glycerol. Mounting a rapid response to increased osmolarity is essential to yeast survival [87,133]. Accordingly, *S. cerevisiae* can activate the HOG pathway within one minute of experiencing an osmotic shock[167]. Yeast can also effectively respond to rapid periodic oscillations (with frequencies up to 0.0046 Hz) between low and high external osmolyte concentrations[87].

Despite its importance during periods of increased osmolarity, unintended activation of the HOG pathway during growth in normal osmolarity conditions is severely deleterious [136,117]. The Sln1-Ypd1-Ssk1 three-component phospho-relay is responsible for maintaining inactivation of the HOG pathway under normal conditions. This three-component phospho-relay is a variant of the two-component signaling systems used by many prokaryotes for osmoregulation, chemotaxis, and other key cellular processes. Sln1 is active *in vivo* as a membrane-bound dimer[136]. Under normal osmolarity conditions, Sln1 autophosphorylates on a histidine residue and then irreversibly transfers the phosphate to an aspartate in its response regulator (RR) domain (Fig. 3.2A). Aspartate-phosphorylated Sln1 binds to the histidine-containing phospho-transfer (HPt) protein Ypd1 and reversibly transfers its phosphate to Ypd1 (Fig. 3.2B). Finally, phospho-Ypd1 transfers its phosphate to dimeric Ssk1, preventing it from interacting with Ssk2 (or the functionally redundant Ssk22) and inhibiting HOG pathway activity[167,135,92]. The sequence of phosphate transfers in the three-component relay is summarized in Fig. 3.2A. In response to osmotic shock, the phospho-relay is

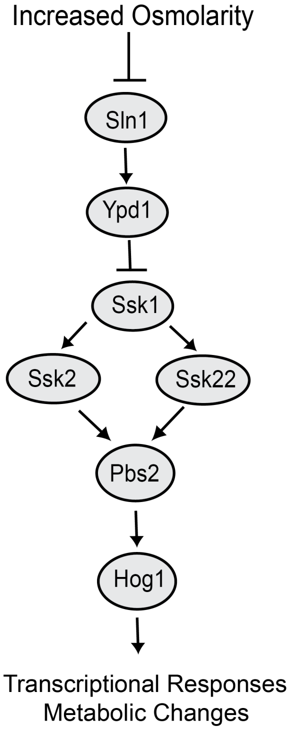


Figure 3.1: Structure of the Sln1-Ypd1-Ssk1 relay and the HOG MAPK pathway. Under normal growth conditions, Sln1 activates Ypd1, which in turn inhibits Ssk1. Increased osmolarity inhibits Sln1, resulting in activation of Ssk1 and of the MAP kinase Hog1 through the upstream MAP3Ks Ssk2 and Ssk22 and the MAP2K Pbs2. Activated Hog1 regulates transcriptional and metabolic changes within the cell that increase production and retention of glycerol.

inactivated, and Ssk1 is rapidly dephosphorylated through an as-yet unknown mechanism. Unphosphorylated Ssk1 then activates Ssk2 and Ssk22, leading to induction of the HOG pathway[88,180]. It is thus the essential controller of HOG pathway activity, and variations in its concentration could compromise fitness.

There is limited existing experimental evidence that the phospho-relay is able to maintain robust phosphorylation of Ssk1 and inactivation of the HOG pathway despite changes in the levels of some pathway components[117,92,200]. We undertook a comprehensive characterization of the sensitivity of HOG pathway activation to changes in the expression levels of the phospho-relay proteins Sln1, Ypd1, and Ssk1. We systematically under- and overexpressed the three proteins using the GEV artificial induction system, which allows for rapid and nearly gratuitous induced expression of individual yeast genes[146]. We found that the phospho-relay maintains inactivation of the HOG pathway even after moderate perturbation of Sln1, Ypd1, and Ssk1.

We developed a detailed, biochemically realistic mathematical model of the HOG pathway three-component phospho-relay to elucidate the mechanism underlying this robustness (Fig. 3.2C). Our model incorporates extensive structural and mechanistic information about the phospho-relay and considers nearly all possible interactions between the three relay proteins. We used mass-action kinetics and algebraic calculations to characterize the steady-state behavior of the model. Steady-state algebraic models are a useful alternative to existing computational models of the HOG pathway for understanding robust behavior[87,117,115]. Unlike numerical simulations[87,117,115], algebraic manipulations can be done without ever assigning specific values to the parameters (i.e., the rate constants in the reaction network), many of which are difficult or impossible to measure experimentally[74]. This advantage enabled us to design and analyze a more biochemically realistic model. A steady-state approximation is appropriate because previous studies have shown that activation of the HOG pathway does not vary under normal growth conditions[87,133,115,90,89].

Our steady-state analysis predicted that relative levels of dephosphorylated Ssk1 depend solely on

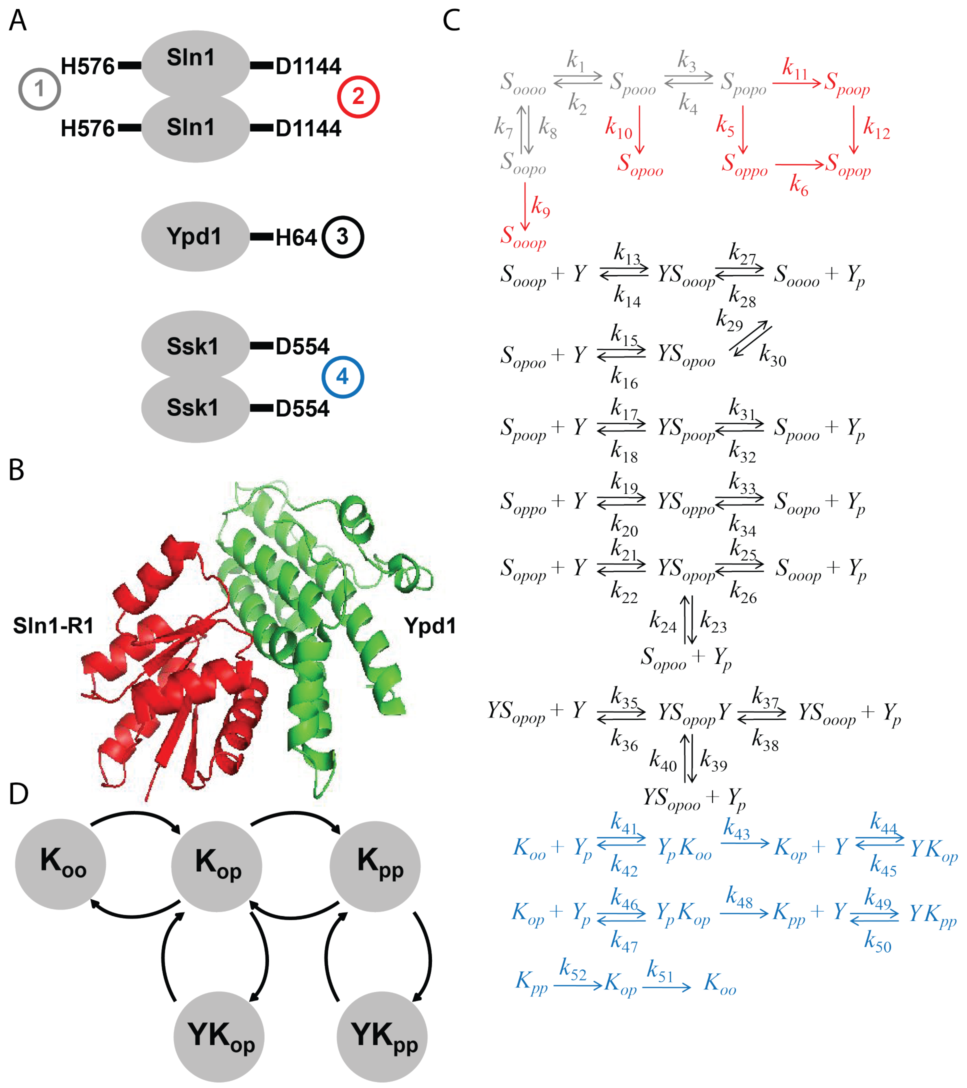


Figure 3.2: Biochemically realistic model of the HOG pathway three-component phospho-relay in *S. cerevisiae*. **A** Cartoon diagram of the phospho-relay indicating the quaternary structure of each component and the four relevant phosphorylation sites. Phospho-transfer proceeds from Sln1 (after autophosphorylation on H576 and transfer to D1144) to Ypd1 to Ssk1, as indicated by the numbers in circles. **B** Crystal structure of Ypd1 (green) in complex with the response regulator domain of Sln1 (Sln1-R1, red). Drawn from data presented in Ref.[232]. **C** Reaction network diagram describing our model of the phospho-relay. The network includes nearly all possible interactions between the three proteins subject to the biochemical assumptions outlined in the main text. For clarity, the reaction network is color-coded to indicate the groups of reactions involved in each phosphorylation event. S denotes Sln1, Y denotes Ypd1, and K denotes Ssk1. Phosphorylated residues are denoted by p, unphosphorylated residues by o. **D** Directed graph describing the subnetwork involving phosphorylation and dephosphorylation of Ssk1. The graph contains four loops that are connected as a branched tree.

Ypd1 levels and that robustness is achieved by maintaining Ypd1 in large excess. We experimentally tested this prediction by perturbing protein expression levels so as to deplete this buffering pool of Ypd1. All such perturbations compromised the ability of the phospho-relay to inhibit the HOG pathway, leading to hyperactivation in normal osmolarity conditions. The presence of a large buffering pool of an intermediate phospho-relay component is a previously underappreciated mechanism for robustness and suggests a possible advantage of a three-component relay over a two-component system.

3.2 Results and Discussion

3.2.1 Inhibition of HOG pathway signaling is robust to moderate overexpression of phospho-relay components.

Inappropriate HOG pathway activation during normal osmolarity growth unnecessarily alters transcription and metabolism[117,92,200]. To assess the robustness of HOG pathway inhibition by the three-component phospho-relay, we created strains capable of overexpressing Sln1, Ypd1, and Ssk1 in response to β-estradiol. For these overexpression experiments, we used diploid strains homozygous for the GEV artificial transcription factor[146]. GEV consists of the Gal4 DNA-binding domain, the estrogen receptor, and the VP16 activation domain. Upon treatment with the hormone β-estradiol, GEV rapidly translocates to the nucleus, where it activates transcription from promoters containing the Gal4 DNA-binding target sequence. The GEV system enables rapid induction of individual yeast genes with limited off-target effects[146]. To make a given phospho-relay gene GEV-inducible, we placed it under the control of the *GAL1* promoter (SLN1/P$_{GAL1}$-SLN1, YPD1/P$_{GAL1}$-YPD1, and SSK1/P$_{GAL1}$-SSK1), as described previously[146].

Inappropriate activation of the HOG pathway is known to cause a growth defect[117,92,200]. We therefore measured growth of these GEV strains after induction with β-estradiol at a range of concentrations to screen for HOG pathway hyperactivity. A strain carrying an inducible allele of Pbs2 (P_{GAL1}-PBS2/PBS2) was used as a control because Pbs2 overexpression is known to cause severe growth defects from inappropriate activation of the HOG pathway[117]. As shown in Fig. 3.3A, strains overexpressing Pbs2 exhibited a measurable growth defect, while strains overexpressing components of the phospho-relay (Sln1, Ypd1, and Ssk1) showed no significant change.

We also assayed for Hog1 phosphorylation following GEV induction of phospho-relay components to obtain direct evidence that moderate overexpression does not cause HOG pathway hyperactivation. We used overexpression of Pbs2 and Ssk2 (also known to cause hyperactivation of the HOG pathway[117]) as positive controls. After 30 minutes of GEV induction, there was no detectable increase in Hog1 phosphorylation in strains overexpressing phospho-relay components (Fig. 3.3B). In contrast, Pbs2 and Ssk2 overexpression caused phosphorylation of Hog1. Interestingly, overexpression of the Ssk2 homolog Ssk22 had the strongest effect on Hog1 phosphorylation.

We then constructed diploid strains with a single inducible copy of the HOG phospho-relay gene of interest and a single P_{STL1}-YFP reporter to assay for HOG pathway transcriptional activity in response to overexpression of relay components. Stl1 is a glycerol/H^{+} symporter whose expression is strongly upregulated in response to osmotic shock[61]. We overexpressed all three relay components, Pbs2, and Ssk22. Here we used Ssk22 as a control instead of Ssk2 because it showed a strong effect on Hog1 phosphorylation in the previous experiment. After 120 minutes of induction with 10 μM β-estradiol, Ssk22 and Pbs2 overexpression led to HOG-dependent transcription from the *STL1* promoter, as indicated by an increase in YFP fluorescence (Fig. 3.4). Over the same period of time, overexpression of the phospho-relay components (Sln1, Ypd1, and Ssk1) caused almost no transcription from the *STL1* promoter. After 19 hours, overexpression of Ssk1 and Sln1 did increase expression of *YFP* from the P_{STL1}-YFP reporter. These effects on longer time scales may have been due to factors beyond Ssk1 and Sln1 overexpression, however, as there was a population expressing YFP even in the control strain at 19 hours.

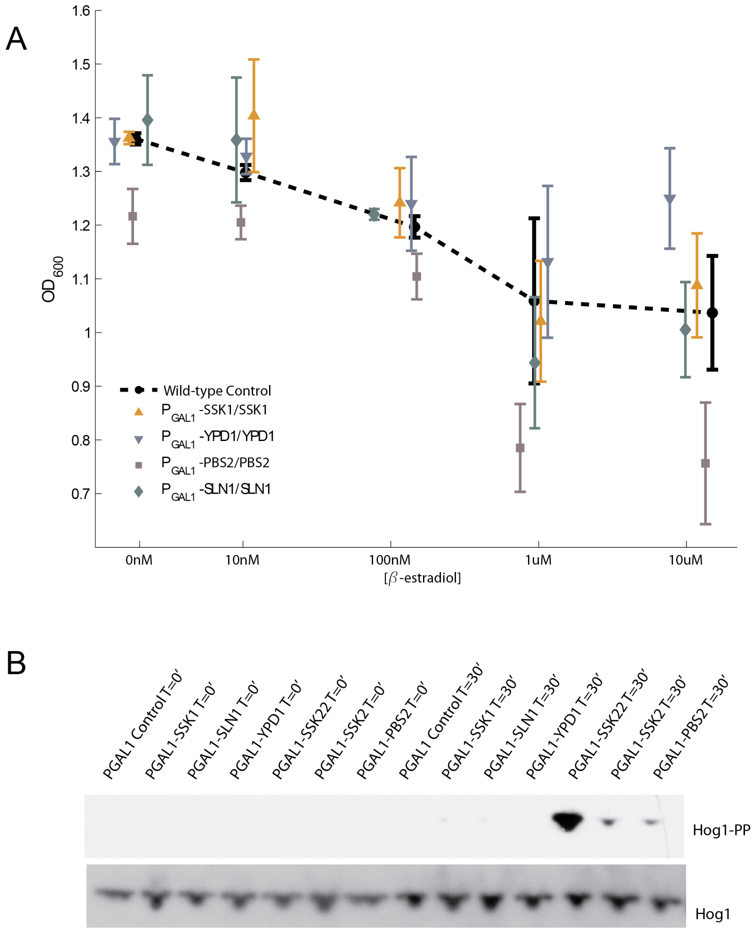


Figure 3.3: Moderate overexpression of phospho-relay components does not activate the HOG pathway. **A** Homozygous GEV diploid strains with a single inducible copy of a phospho-relay gene were grown to saturation in different concentrations of β-estradiol. The optical density (OD_{600}) after 13 hours of growth is plotted. Overexpression of Pbs2 caused a growth defect at higher concentrations of β-estradiol, while no significant growth defects were observed following overexpression of relay components (beyond the defect in the wild-type due to overexpression of Gal4 target genes) . Each point represents the mean and standard deviation over four replicates. **B** We assayed for Hog1 phosphorylation using an antibody specific to doubly phosphorylated Hog1 to confirm that moderate overexpression of phospho-relay components does not lead to activation of the HOG pathway. Overexpression of phospho-relay components using a saturating dose (10 μM) of β-estradiol did not lead to Hog1 phosphorylation. Overexpression of the positive controls Ssk22, Ssk2, and Pbs2, however, led to clear upregulation of Hog1 phosphorylation after 30 minutes of induction. Total Hog1 is shown as a loading control.

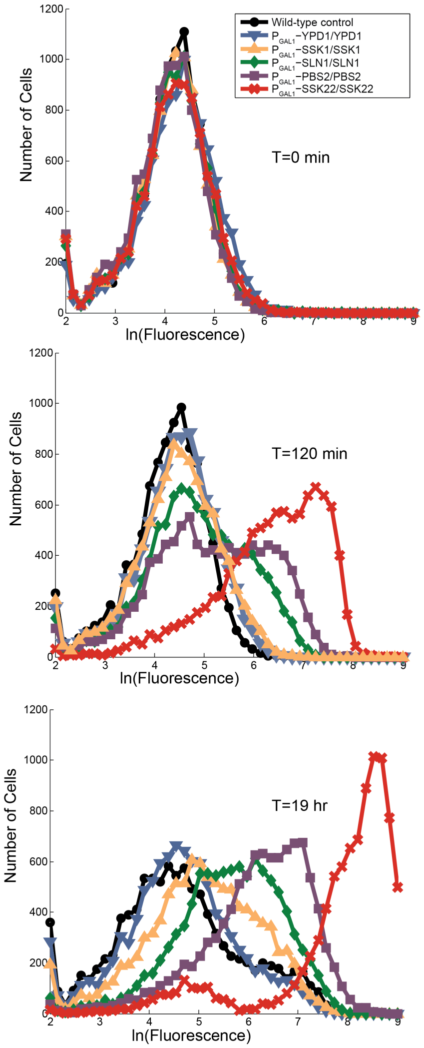


Figure 3.4: Moderate overexpression of phospho-relay components causes mild activation of HOG pathway transcriptional targets on long timescales. *STL1* is a gene induced in response to active Hog1. We replaced one copy of the *STL1* gene with yEVenus in homozygous GEV strains to create a P$_{STL1}$-yEVenus transcriptional reporter of HOG pathway activity. We used flow cytometry to analyze yEVenus expression in cells at the start of the experiment and at 2 hours and 19 hours after induction with 10 μM β-estradiol. Pbs2 and Ssk22 overexpression caused the strongest activation of the HOG pathway transcriptional reporter. Cell counts are plotted against the natural log of their fluorescence signal [a.u.].

3.2.2 Construction of a biochemically realistic model of the HOG pathway phospho-relay.

Growth, phosphorylation, and transcriptional measurements of HOG pathway activity all indicated that HOG pathway activation is robust to fluctuations in the Sln1-Ypd1-Ssk1 phospho-relay components. These results prompted us to investigate the mechanistic basis of this robustness using a biochemical model. The reaction network underlying our model (Fig. 3.2C) has 37 nodes and involves 13 species. In this section we discuss the biochemical justification for key assumptions in the model.

There exist high-resolution crystal structures of Ypd1 alone and in complex with the Sln1 receiver domain(Fig. 3.2B)[198,232]. Genetic and biochemical evidence suggest that Sln1 forms an obligate homodimer and that Ypd1 can interact with either half of the dimer[136,214], which implies that formation of a Ypd1-Sln1(dimer)-Ypd1 ternary complex is possible. Accordingly, we include Sln1 in the reaction network as a dimer with four relevant phosphorylation sites. The Sln1 dimer is referred to as $S_{H1D1H2D2}$, where H1 and D1 denote the phosphorylatable histidine and aspartate residues in one half of the dimer and H2 and D2 denote the corresponding residues in the other half. The dimer is allowed to autophosphorylate on either histidine residue. Phospho-transfer from the histidine to the aspartate in the RR domain is treated as irreversible. A reverse reaction is included with all histidine autophosphorylation steps to account for possible hydrolysis of the phosphate prior to transfer[180,100]. Each half of the dimer is assumed to be independent from the other. Coincident phosphorylation events (e.g., S_{OOOO} forming S_{POPO} in one step) are therefore considered to be unlikely and are excluded from the model. Following these assumptions, the model includes nine different forms of free Sln1 that are interconverted as shown in the reaction network.

In the second leg of the phospho-relay, any Sln1 phospho-form with at least one phosphorylated aspartate is allowed to reversibly associate with unphosphorylated monomeric Ypd1 (Y) to form a series of binary complexes (YS_{OOOP}, YS_{OPOO}, YS_{POOP}, YS_{OPPO}, and YS_{OPOP}). Sln1-Ypd1 phospho-transfer has been shown to be reversible[98]. As such, all reactions that produce phospho-Ypd (Y_P) are treated as reversible. No assumption is made about which half Ypd1 binds to in the YS_{OPOP} binary complex because the halves of the Sln1 dimer are considered to be indistinguishable. Accordingly, the YS_{OPOP} complex is allowed to form from either $Y_P + S_{OOOP}$ or $Y_P + S_{OPOO}$. Additionally, YS_{OPOP} can bind to a second Y molecule to form a ternary complex ($YS_{OPOP}Y$) that produces Y_P using either of the Sln1 phospho-aspartate residues.

Phospho-Ypd1 then binds to and phosphorylates Ssk1, which is modeled as a dimer with two phosphorylation sites (K_{OO})[92]. The two monophosphorylated forms of Ssk1, which are known to be fully inactive[92], are assumed to be identical (i.e., $K_{PO} = K_{OP}$). Y_P can form a complex with K_{OO}, leading to the production of K_{OP}, and in turn Y_P can bind to K_{OP} and transfer a phosphate to produce K_{PP}. The network includes one further interaction between Ypd1 and Ssk1 deduced from kinetic data. The half-life of phospho-Ssk1 *in vitro* has been measured to be dramatically different with and without the presence of Ypd1 (over 40 hours vs. 13 minutes, respectively), suggesting that Ypd1 binds to K_{OP} and K_{PP} to prevent hydrolysis of the phosphate[100,99]. Accordingly, the reaction network includes the reversible formation of dead-end complexes between Ypd1 and phospho-Ssk1. Finally, unstabilized K_{PP} and K_{OP} are allowed to lose phosphates via spontaneous hydrolysis. Inclusion of these hydrolysis reactions ensures that there is a complete cycle for Ssk1 modification/demodification and that the system can reach a stable steady state. We emphasize, however, that spontaneous hydrolysis of complexed Ssk1 is likely not the mechanism for rapid dephosphorylation of large quantities of Ssk1 in response to osmotic shock. The mechanism for this rapid activation remains unknown but is irrelevant for our model, which is restricted to yeast growing in steady-state normal osmolarity conditions.

3.2.3 DERIVATION OF THE KEY INVARIANT.

Robust inactivation of the HOG pathway requires that only a small fraction of total Ssk1 (K_T) be in the active (K_{OO}) modification form at steady state. The goal of this section is to derive a simple steady-state expression (an invariant) for the ratio of active to total Ssk1. We find an invariant of the form

$$\frac{[K_{OO}]}{K_T} = \frac{1}{1 + \frac{1}{\gamma}[\alpha[Y_P] + \beta[Y_P]^2 + [Y](\alpha'[Y_P] + \beta'[Y_P]^2)]}, \tag{3.1}$$

where the coefficients are combinations of the rate constants. In this section we derive Eq. 3.1, and in the following section we discuss experimental tests of its predictions.

The subnetwork involving Ssk1 contains four loops, which are linked in a branched tree (Fig. 3.2D). It is a general feature of such networks that, at steady state, each individual loop is at steady state, irrespective of any other loops in which the components participate[73,3].

If each individual loop is at steady state, then the forward flux through each loop must be balanced by the backward flux, which yields the following four equations:

$$\begin{aligned} k_{51}[K_{OP}] &= k_{43}[Y_P K_{OO}] \\ k_{52}[K_{PP}] &= k_{48}[Y_P K_{OP}] \\ k_{45}[Y K_{OP}] &= k_{44}[Y][K_{OP}] \\ k_{50}[Y K_{PP}] &= k_{49}[Y][K_{PP}]. \end{aligned} \tag{3.2}$$

Because the intermediate complexes $Y_P K_{OO}$ and $Y_P K_{OP}$ are also at steady state, we can write

$$\begin{aligned} k_{43}[Y_P K_{OO}] &= k_{41}[Y_P][K_{OO}] - k_{42}[Y_P K_{OO}] \\ k_{48}[Y_P K_{OP}] &= k_{46}[Y_P][K_{OP}] - k_{47}[Y_P K_{OP}]. \end{aligned}$$

from which we deduce that

$$\begin{aligned} [Y_P K_{OO}] &= \tfrac{k_{41}}{k_{42}+k_{43}}[Y_P][K_{OO}] \\ [Y_P K_{OP}] &= \tfrac{k_{46}}{k_{47}+k_{48}}[Y_P][K_{OP}]. \end{aligned} \tag{3.3}$$

Substituting Eq. 3.3 into Eq. 3.2, we obtain expressions for K_{OP}, K_{PP}, YK_{OP}, and YK_{PP} in terms of K_{OO}, Y, and Y_P. From Eq. 3.3 we already have expressions for $Y_P K_{OO}$ and $Y_P K_{OP}$ in terms of K_{OO} and Y_P. As such, we are able to calculate the total amount of Ssk1 (K_T) in terms of just K_{OO}, Y, and Y_P. We have

$$K_T = [K_{OO}] + [K_{OP}] + [K_{PP}] + [Y_P K_{OO}] + [Y_P K_{OP}] + [YK_{OP}] + [YK_{PP}].$$

Substituting for the individual terms, we obtain

$$K_T = [K_{OO}]\left[1 + \frac{1}{\gamma}\left(\alpha[Y_P] + \beta[Y_P]^2 + \left(\alpha'[Y_P] + \beta'[Y_P]^2\right)[Y]\right)\right],$$

where

$$\begin{aligned} \alpha &= k_{41}k_{45}k_{50}k_{52}(k_{43}+k_{51})(k_{47}+k_{48}) \\ \beta &= k_{41}k_{43}k_{45}k_{46}k_{50}(k_{48}+k_{52}) \\ \alpha' &= k_{41}k_{43}k_{44}k_{50}k_{52}(k_{47}+k_{48}) \\ \beta' &= k_{41}k_{43}k_{45}k_{46}k_{48}k_{49} \\ \gamma &= k_{45}k_{50}k_{51}k_{52}(k_{42}+k_{43})(k_{47}+k_{48}). \end{aligned}$$

The relative concentration of K_{OO} is thus given by the invariant in Eq. 3.1.

The tree of loops structure of the Ssk1 network has an important consequence. It implies that the steady-state ratio of active to total Ssk1 is independent of the upstream biochemistry (i.e., the mechanistic details of the various Sln1 and Ypd1 reactions) as long as some process exists to generate positive levels of Y and Y_P. In that case, the ratio will always be given by Eq. 3.1, although the numerical value will of course differ depending on steady-state concentrations of Y and Y_P. The implications of this result, including its suggestion that robustness in the HOG pathway is independent of putative Sln1 bifunctionality, are considered in the conclusion.

3.2.4 BREAKDOWN OF ROBUSTNESS DUE TO DEPLETION OF THE YPD1 POOL.

The invariant derived from our mathematical model (Eq. 3.1) suggests that Ypd1 levels are critical to robustness. The denominator of the invariant is quadratic in the concentration of free Y_P and linear in the concentration of free Y. Provided that the upstream network favors production of Y_P over Y and that there is substantially more Ypd1 than Ssk1, the denominator of Eq. 3.1 will be large, and the relative concentration of K_{OO} will be maintained at a low level. This situation allows for considerable under- or overexpression of pathway components without spurious activation of the HOG pathway, in agreement with our experimental findings.

The invariant predicts that massive overexpression of Ypd1 should not cause phosphorylation of Hog1. In fact, additional Ypd1 would drive the $\frac{[K_{OO}]}{K_T}$ ratio even closer to zero, lowering the amount of unphophorylated Ssk1 required for HOG pathway activation. In contrast, Ypd1 underexpression should increase the ratio, potentially compromising fitness due to inappropriate activation of the HOG pathway. Similarly, massive overexpression of Ssk1 should deplete free Y and Y_P due to increased levels of the four intermediate complexes ($Y_P K_{OO}$, $Y_P K_{OP}$, $Y K_{OP}$, and $Y K_{PP}$). Under the assumption of tight binding between Ypd1 and Ssk1 in each of these complexes, which is well-supported by existing kinetic data[100,99], very little free Ypd1 will be present at steady state if there is much more Ssk1 than Ypd1. As such, Eq. 3.1 predicts that the ratio will be higher following massive overexpression of Ssk1 than under wild-type conditions.

We experimentally validated these three predictions. The GEV system can achieve at most a 10-

fold increase in protein expression from a single inducible allele[146]. We created haploid GEV yeast strains carrying high-copy 2μ plasmids with a GEV-inducible allele (P$_{GAL1}$-GENE) of a gene of interest, which allowed us to test the model prediction that massive overexpression of Ssk1, but not of Ypd1, should lead to inappropriate HOG pathway activation. These high-copy yeast plasmids are estimated to be present at 15-50 copies per cell[219,111,54,33].

We measured the growth of these strains in different concentrations of β-estradiol to assay for growth defects that might be due to HOG pathway hyperactivation. Massive overexpression of both Sln1 and Ssk1 caused a growth defect over a range of β-estradiol concentrations, but the strain with Ypd1 overexpressed grew as well as a wild-type strain carrying only the empty vector (P$_{GAL1}$ 2μ scURA3) (Fig. 3.5A). Examination of growth over a finer range of β-estradiol concentrations indicated that overexpression of Ssk1 caused a more severe growth defect than overexpression of Sln1 (Fig. 3.5B). These defects were also visible on solid media (Fig. 3.5C). As such, extreme overexpression of Ssk1 compromises fitness.

We again assayed phospho-Hog1 levels to check if overexpression of Sln1 and Ssk1 causes activation of the HOG pathway in normal osmolarity conditions (Fig. 3.6). We measured Hog1 phosphorylation levels after GEV-induction of relay components and of the positive controls Pbs2 and Ssk22 from a multi-copy plasmid. Overexpression of Pbs2, Ssk22, and Ssk1 caused a significant change in Hog1 phosphorylation after 30 minutes (Fig. 3.6B). Although some Hog1 phosphorylation was observed after overexpression of Sln1, the increase was insignificant. Interestingly, overexpression of Ypd1 did not cause an increase in the level of phosphorylated Hog1. In fact, levels of Hog1 phosphorylation were reduced in the Ypd1 overexpression strain ($p = 0.0246$, two-way ANOVA).

We could not perform underexpression experiments in a ypd1Δ background because deletion of Ypd1 is lethal. To underexpress Ypd1, we instead sporulated the diploid strain (P$_{GAL1}$-YPD1/YPD1) containing a wild-type copy of *YPD1* and a single copy under the control of P$_{GAL1}$ onto media containing 10 nM β-estradiol. We reasoned that 10 nM β-estradiol would give sufficient expression

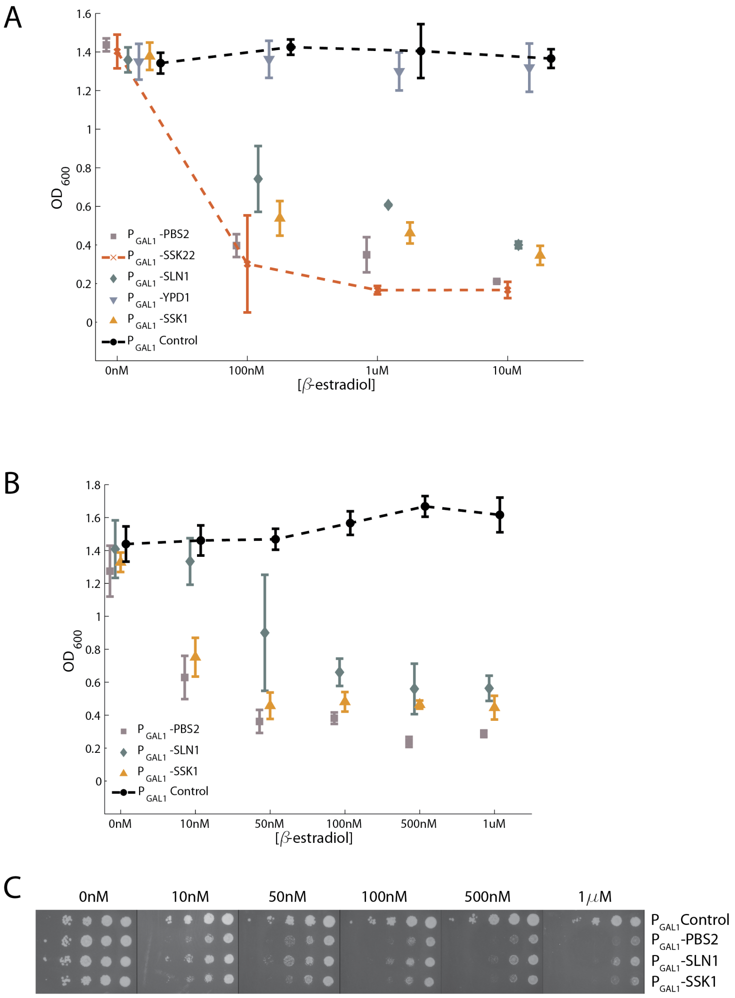


Figure 3.5: Massive overexpression of phospho-relay components leads to growth defects. **A** Haploid GEV strains carrying a high-copy plasmid with an inducible HOG pathway gene were grown in different concentrations of β-estradiol. The OD$_{600}$ after 36 hours of growth is plotted as a function of β-estradiol concentration. Each point represents the mean and standard deviation of four replicates. Overexpression of Sln1 and Ssk1 (but not Ypd1) caused a growth defect. **B** The same strains were grown over a finer titration of β-estradiol concentrations. The OD$_{600}$ after 36 hours is plotted. At this resolution, it is clear that the growth defect from Ssk1 overexpression is more severe than the growth defect from Sln1 overexpression at low β-estradiol concentrations. **C** The same strains were frogged onto plates containing different concentrations of β-estradiol. Massive overexpression of Sln1 and Ssk1 again caused a growth defect comparable to that from overexpression of Pbs2. In all experiments, the parent strain carrying the empty vector plasmid [$2\mu\ P_{GAL1}$ scURA3] was used as a negative control.

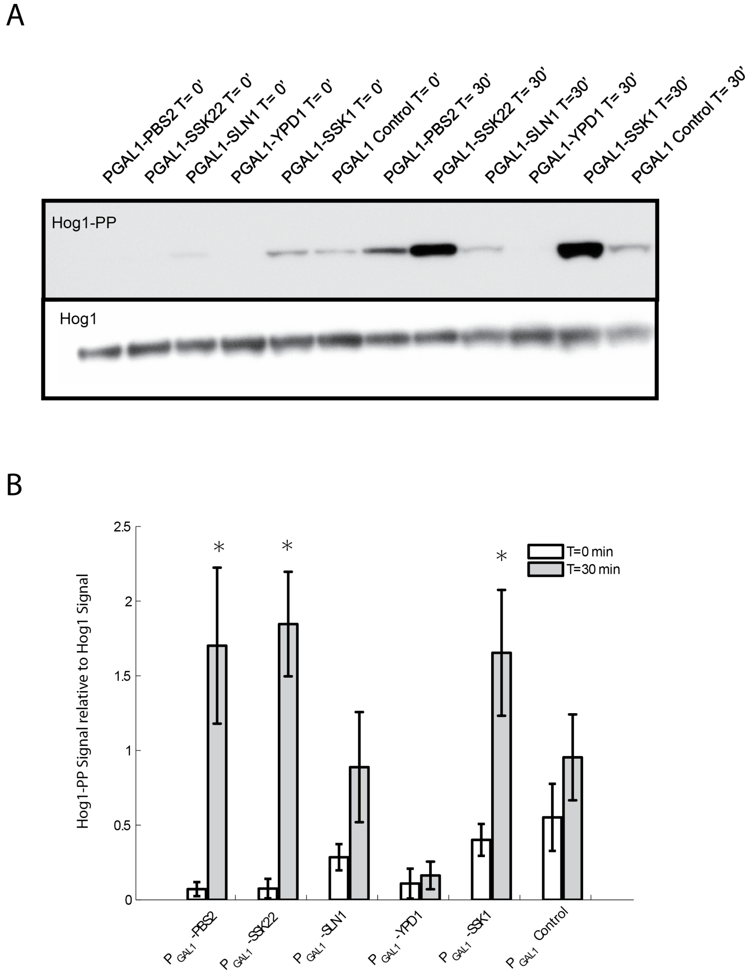


Figure 3.6: Growth defects following massive overexpression of phospho-relay components are due to activation of the HOG pathway. **A** We assayed for Hog1 phosphorylation after overexpression of relay components (Sln1, Ypd1, Ssk1) and positive controls (Pbs2, Ssk22). The parental strain carrying the empty plasmid vector was used as a negative control. **B** We quantified the amount of phosphorylated Hog1 (relative to Hog1) in five biological replicates of this experiment. Error bars represent the standard error. Pbs2, Ssk22, and Ssk1 caused a significant ($*$) change in Hog1 phosphorylation levels after overexpression for 30 minutes ($p = 0.0395, 0.0096$, and 0.0224, respectively; paired t-test). Hog1 phosphorylation levels were also significantly lower in the Ypd1 overexpression strain ($p = 0.0246$, two-way ANOVA).

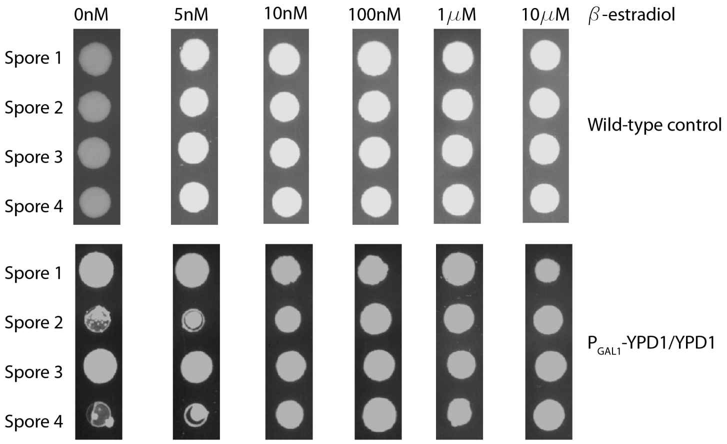


Figure 3.7: Underexpression of Ypd1 causes a severe growth defect. A diploid homozygous GEV strain carrying one inducible allele of Ypd1 (YPD1/P_{GAL1}-YPD1) was sporulated onto 10 nM β-estradiol, and individual spores were frogged onto plates containing different concentrations of β-estradiol. At low concentrations of β-estradiol, spores carrying the inducible allele (P_{GAL1}-YPD1, row 2 and row 4) exhibited a growth defect due lower levels of Ypd1 and hyperactivation of the HOG pathway. Spores carrying the wild-type allele (row 1 and row 3) showed no growth defect.

of Ypd1 for cell growth, which was confirmed by observation of four viable spores. We then grew these spores on media containing a range of β-estradiol concentrations (Fig. 3.7). Ypd1 underexpression caused a clear growth defect compared to the wild-type control on 0 nM and 5 nM β-estradiol, indicating that Ypd1 underexpression is toxic. In contrast, underexpression of Sln1 and Ssk1 did not cause a growth defect (Supplemental Fig. B.S1).

3.2.5 The growth defect following Sln1 overexpression is only partially due to HOG pathway activation.

As discussed above, we observed a slight but non-significant increase in Hog1 phosphorylation following massive overexpression of Sln1 (Fig. 3.6). We therefore investigated whether the growth defect in response to Sln1 overexpression is only partially due to HOG pathway activation by creating yeast strains null for SSK1. Activation of the HOG cascade through the Sln1 branch requires Ssk1, so in ssk1Δ strains it is not possible for Sln1 overexpression to activate the HOG pathway.

Overexpression of Sln1 from a 2μ plasmid using GEV was still detrimental to growth in the ssk1Δ strain (Fig. 3.8), indicating that the growth defect due to Sln1 overexpression is only partially due to HOG pathway activation. This result held both for growth in liquid cultures (Fig. 3.8A) and on solid media (Fig. 3.8B). It is consistent with Sln1 overexpression causing a smaller effect on Hog1 phosphorylation levels (Fig. 3.6).

3.3 CONCLUSIONS

Robustness of the Sln1-Ypd1-Ssk1 phospho-relay is essential to prevent spurious activation of the HOG pathway, which severely compromises yeast fitness. We established that the phospho-relay is robust to perturbations in the concentrations of the three relay components. A theoretical analysis suggested that a large pool of the intermediate component Ypd1 can buffer fluctuations in other pathway components to maintain robustness. This suggestion was consistent with earlier published measurements indicating that Ypd1 is at least 5 times more abundant than Ssk1 at normal expression levels[117,68]. Although Ypd1 may also bind to the protein Skn7, combined levels of Ssk1 and Skn7 have been measured to be below total Ypd1 levels[68]. Our subsequent experiments confirmed that depletion of this buffering pool of Ypd1 leads to inappropriate activation of the HOG pathway.

The differential expression of Ypd1 and Ssk1 enables phosphorylation of excess Ssk1 and stabilization of the new phospho-Ssk1, buffering HOG pathway activation to fluctuations in Ssk1 levels. This novel mechanism of robustness suggests an advantage of a three-component architecture over a two-component one. In particular, the implementation of an analogous buffering strategy in a two-component system would be difficult because it would require expressing the sensor histidine kinase at very high levels. This situation might lead to imprecise sensing and various other off-target effects. In contrast, the use of an intermediate transfer protein enables robust buffering with both the sen-

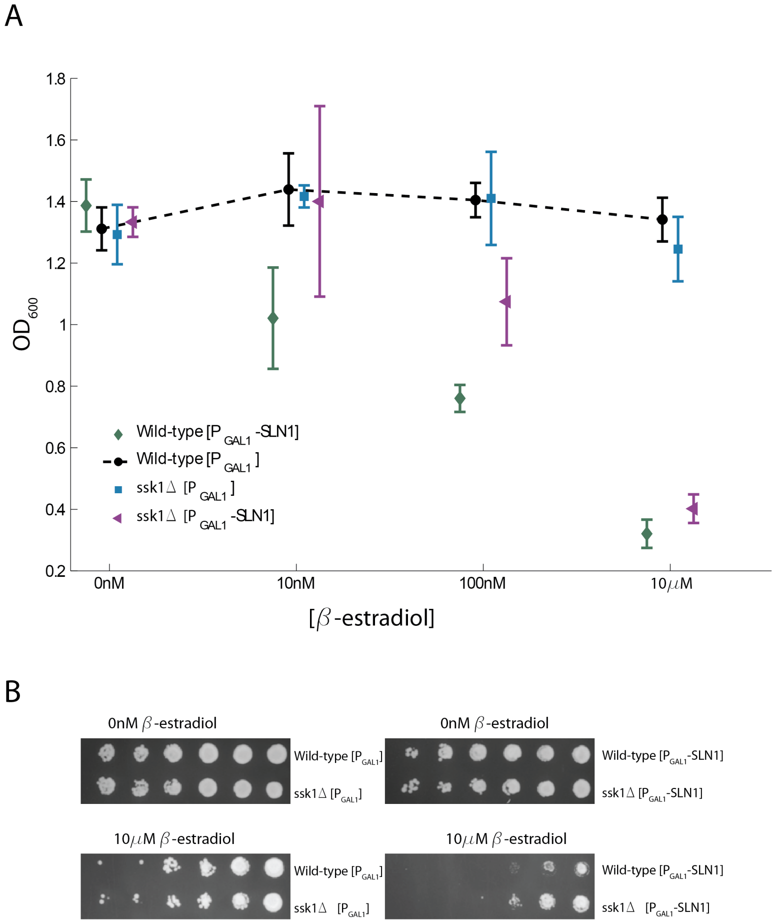


Figure 3.8: Growth defects following Sln1 overexpression are not completely due to HOG pathway activation. **A** We assayed growth in wild-type and ssk1Δ strains overexpressing Sln1 in response to β-estradiol or carrying an empty P$_{GAL1}$ vector control. Deletion of Ssk1, which prevented HOG pathway activation by Sln1, partially alleviated the growth defect due to Sln1 overexpression. **B** Growth of cells on plates containing 10 μM β-estradiol indicated that ssk1Δ reduced the toxicity of Sln1 overexpression.

sor and response regulator expressed at comparable levels. Our work has thus identified a potential mechanism for circumventing a trade-off between efficient sensing and robust control. There are other possible advantages for a three-component architecture, including combinatorial control of response regulators by sensor proteins through a common phosphotransfer protein or segregation of sensing and activation functions between the nucleus and cytoplasm. Intriguingly, deletion of YPD1 has recently been shown to cause constitutive activation of the HOG pathway in *Candida albicans*, suggesting that its buffering capacity might also be important in this organism[142].

Robustness in real biological systems is necessarily approximate and apt to be compromised at extreme expression levels of cellular components. In many systems, however, it has proven difficult to characterize where robustness breaks down and to reconcile such results with mathematical models, which often predict exact robustness[189]. Our combined theoretical and experimental results specify a single condition (Ypd1 in large excess) for robust regulation of the HOG pathway.

The link between bifunctionality and robustness is well-established[189,83,82,48,43], and it is known that bifunctionality of EnvZ is essential to robustness in *Escherichia coli* osmoregulation[15,191]. As such, it is intriguing that our model suggests that robustness in *S. cerevisiae* osmoregulation is not dependent on bifunctionality of Sln1. It is important to emphasize, however, that bifunctionality would not compromise robustness. Rather, the model indicates that any upstream process that produces non-zero levels of Y and Y_P should enable the same fundamental behavior predicted by Eq. 3.1. The possibility that Sln1 exhibits phosphatase activity warrants further experimental investigation.

3.4 METHODS

3.4.1 YEAST STRAINS AND MEDIA.

All yeast strains used in this study are listed in Supplemental Table B.S1. The homozygous GEV diploid strain, which served as the wild-type background strain for all diploid overexpression experiments, was created by mating haploid GEV strains yMM598 and yMM1101[146] and picking zygotes to create yMM1104. As described previously[146], diploid yeast strains capable of overexpressing the desired HOG pathway protein from a single locus (yMM1263, yMM1272, yMM1259) were created by transforming[69] the homozygous GEV diploid strain yMM1104 with the KanMX-P$_{GAL1}$ cassette amplified from yMM1100 genomic DNA using appropriate oligonucleotide pairs. Transformants were verified by colony PCR and sequencing.

Yeast strains containing 2μ plasmids for massive overexpression of pathway components (yMM1313-yMM1318) were constructed using recombination-mediated plasmid construction[132,153] to generate the overexpression plasmids pMM330-pMM334 *in vivo* (as described below). Positive transformants were selected for and maintained on SC-Ura media[69].

Yeast strains containing the P$_{STL1}$-YFP reporter of HOG pathway activity and one estradiol-inducible allele of a HOG pathway gene (yMM1296, yMM1298, yMM1300, yMM1301, yMM1304, yMM1305) were constructed by transforming the heterozygous diploid GEV yeast strains (yMM1104, yMM1272, yMM1264, yMM1259 yMM1286, yMM1287) already containing an inducible allele with the product of PCRing yECitrine-HphMX off plasmid pMM280 using appropriate oligonucleotides. The oligonucleotides contained homology such that the *STL1* ORF was replaced with the yECitrine-HphMX cassette. Transformants were verified by colony PCR, and expression of YFP in 1M sorbitol was assayed.

Yeast strains null for the SSK1 gene (ssk1Δ) were created by deleting the *SSK1* ORF using appropriate oligonucleotide pairs to amplify KanMX from pMM131 and transforming it into yMM630.

Transformants were selected for drug resistance and verified by colony PCR and sequencing.

Standard yeast media was used as noted. Low fluorescence yeast media was prepared as described previously[187].

3.4.2 Plasmid construction.

All plasmids used in this study are listed in Supplemental Table B.S2. Plasmid pMM329 (P_{GAL1} scURA3 2μ) was constructed by PCR of the native *GAL1* promoter from genomic DNA prepared from yMM1100 using appropriate primers. This promoter was ligated in pMM12 between the restriction sites KpnI and XhoI (scURA3 2μ) [194]. The resulting plasmid served as a template to create a series of overexpression plasmids with different HOG pathway genes under the control of the *GAL1* promoter using yeast recombination-mediated plasmid construction [132,153]. Appropriate primer pairs were used to amplify Pbs2, Ssk22, Sln1, Ypd1, and Ssk1, respectively, from yMM1100 genomic DNA. These PCR products were then co-transformed with pMM329 linearized with XhoI and SalI. The primer pairs used to amplify the HOG pathway genes contained homology with the pMM329 backbone such that the gene of interest was integrated after P_{GAL1} to create a P_{GAL1}-HOGGENE plasmid (pMM330-pMM334). Positive transformants in which the plasmid had been repaired were selected for on SC-Ura media. Plasmids were purified from these transformants and verified by sequencing.

3.4.3 Overexpression growth experiments on solid media.

Yeast strains were grown overnight to saturation in appropriate media (YPD or SC-Ura media to maintain plasmids). These saturated cultures were serially diluted in 10-fold increments and frogged onto YPD or SC-URA plates containing 0 nM, 10 nM, 100 nM, 1 μM, or 10 μM β-estradiol (Tocris Biosciences). These plates were incubated at 30°C for two days before imaging.

3.4.4 OVEREXPRESSION GROWTH EXPERIMENTS USING PLATE READER.

Yeast strains were grown overnight to saturation in YPD or SC-Ura media. In the morning, each strain was diluted 1:2000 into 200 μL of the same media containing 0 nM, 100 nM, 1 μM, or 10 μM of β-estradiol in a well of a 96-well flat-bottom plate (Costar). Each strain/estradiol combination was run in four replicates on the same plate. Growth curves were generated using a Synergy H1 microplate reader (BioTek). Cells were grown at 30°C with continuous, double-orbital (555 cpm) shaking, and OD_{600} was measured every 20 minutes. Growth rates were calculated from the growth curves using spline-fits determined with the R package grofit[109]. OD was plotted at the time points indicated in the figure legends.

3.4.5 UNDEREXPRESSION EXPERIMENTS BY DIPLOID SPORULATION.

Diploid GEV strains yMM1104 (control), yMM1259 (SLN1/KanMX-P_{GAL1}-SLN1), yMM1263 (SSK1/KanMX-P_{GAL1}-SSK1), yMM1264 (SSK1/KanMX-P_{GAL1}-SSK1), and yMM1272 (YPD1/KanMXrev-P_{GAL1}-YPD1) were sporulated in 1% potassium acetate for 3 days and dissected onto YPD plates containing 10 nM β-estradiol. Two spores from each tetrad contained the wild-type HOG pathway gene (Sln1, Ypd1, or Ssk1), while the other contained the same gene under the control of the *GAL1* promoter (KanMX-P_{GAL1}-SSK1, KanMX-P_{GAL1}-SLN1, KanMX-P_{GAL1}-YPD1). After all spores had grown to a sufficient size, they were diluted into YPD and frogged onto YPD plates containing 0 nM, 5 nM, 10 nM, 100 nM, 1 μM, or 10 μM β-estradiol. Spores were allowed to grow at 30°C for 2 days prior to imaging.

3.4.6 P_{STL1}-YEVENUS INDUCTION AND FLOW CYTOMETRY.

We created diploid GEV strains that carried both an inducible HOG gene under the control of P_{GAL1} promoter and a HOG pathway transcriptional reporter (P_{STL1}-yEVenus) to assay for downstream transcriptional activation in response to overexpression of various HOG pathway proteins. Strains were grown with agitation in low fluorescence media at 30°C to mid-log (Klett 80), at which point $200\ \mu$l of cell culture was sampled for flow cytometry by adding it to $800\ \mu$l of cold PBS + 0.1% Tween 20 stored at 4 °C. Each culture was induced by adding β-estradiol to a final concentration of 10 μM. Cultures were sampled for flow cytometry after induction with β-estradiol at $T = 2$ hours and $T = 19$ hours. Fluorescence was analyzed by flow cytometry on a BD LSRII Multi-Laser Analyzer with HTS (BD Biosciences).

3.4.7 Preparation of protein extracts.

We measured levels of phosphorylated Hog1 following both moderate and massive overexpression of pathway components. For the moderate overexpression experiments, diploid GEV yeast strains yMM1104, yMM1263, yMM1259, yMM1272, and yMM1287, which each contained one estradiol-inducible copy of a HOG pathway gene, were grown to mid-log (Klett 80) in YPD at 30°C with shaking. To assess the effect of massive overexpression of HOG pathway proteins, yeast strains containing the P_{GAL1}-HOGGENE scURA3 2μ overexpression plasmids (yMM1313-yMM1318) were grown in SC-Ura media to mid-log (Klett 80) at 30°C with shaking. For all strains, at $T = 0$ expression of the gene of interest was induced by addition of 10 μM β-estradiol (final concentration). At indicated timepoints, 1.5ml of culture was sampled.

Protein was prepped from samples immediately after each time point. Each sample was centrifuged (1320 RPM) and the supernatant aspirated. The resulting cell pellet was resuspended in $100\ \mu$l of 1X sample buffer (Invitrogen) with β-mercaptoethanol (final concentration of 10%), protease inhibitor (Roche), and phosphatase inhibitor (Fisher Scientific). Samples were heated at 95°C for 5 minutes, vortexed for 2 minutes, and then rapidly frozen in liquid nitrogen and stored at -20°C.

3.4.8 WESTERN BLOTTING.

Prior to western blotting, samples were thawed and centrifuged at 1320 RPM for 5 minutes. They were run on 4-10% Bis-Tris gels (Invitrogen) and transferred to PVDF membranes (Invitrogen) by electrophoresis at 13 V for 4 hours. Membranes were blocked for 1 hour at room temperature with agitation in 1X TBS, 0.1% Tween-20, and 5% milk. Membranes were then probed with primary antibody overnight at 4°C.

The following antibodies were used to detect phosphorylated Hog1, total Hog1, and actin, respectively: anti-phospho-p38 MAPK rabbit monoclonal antibodies (Cell Signaling Technology #9215), anti-c-myc goat polyclonal antibodies (Santa Cruz Biotechnology sc-6815), and anti-β-actin antibody (Abcam ab8224). All primary antibodies were diluted 1:1000 in 1X TBS, 0.1% Tween-20, and 5% milk. Following incubation with the primary antibody, membranes were washed (4×5 minutes) with TBST (1X TBS, 1% Tween-20) and then incubated for 1 hour at room temperature with the appropriate secondary antibody conjugated to HRP (anti-rabbit IgG (Cell Signaling #7074, 1:5000 dilution), rabbit anti-goat IgG (Santa Cruz sc-2768, 1:5000 dilution), and anti-mouse IgG (Abcam ab97023, 1:20000 dilution), respectively). All secondary antibodies were diluted in 5% milk, 1X TBS, and 1% Tween-20. After incubation with the secondary antibody, membranes were washed 4×5 minutes with 5% milk, 1X TBS, and 1% Tween-20.

Western blots were quantified using chemiluminescence. Membranes were developed using the Pierce Supersignal Femto kit following the manufacturer's protocol. Chemiluminescence was quantified using HyBlot CL Autoradiography film. Developed film was scanned on a Epson Perfection 4490 Photo Scanner in transmission mode. Protein levels were quantified by densitometry using the Gel Analysis plug-in in ImageJ[184].

Each membrane was probed for phospho-Hog1, beta-actin, and total Hog1 (in that order). After the chemiluminescence assay but before re-blocking, the membrane was stripped by washing in stripping buffer (2×10 minutes), phosphate-buffered saline (2×10 minutes) and 1X TBS + 1% Tween-20 (2×5 minutes). Stripping buffer contained 15 g glycine, 1 g SDS, and 10 mL Tween 20 in 1 L ultrapure water, with the pH adjusted to 2.2 using concentrated HCl.

3.5 Acknowledgements

We thank Michael Laub, Tom Muir, and Tathagata Dasgupta for helpful discussions, and Tina DeCoste and Christi O'Donnell of the Princeton Flow Cytometry Resource Facility for assistance with flow cytometry. This work was supported by NSF grant 0856285 to JG and a Lewis-Sigler Fellowship from Princeton University to MNM. JPD was supported in part by an NSF Graduate Research Fellowship. Megan Nicole McClean, Ph.D., holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

4

Lack of evidence for substrate channeling or flux between wild-type and mutant isocitrate dehydrogenase to produce 2-HG

AUTHOR LIST

Joseph P. Dexter,* Patrick S. Ward,* Tathagata Dasgupta, Aaron M. Hosios, Jeremy Gunawardena, and Matthew G. Vander Heiden

* J.P.D. and P.S.W. contributed equally.

CHAPTER SUMMARY

Monoallelic point mutations in the cytosolic, NADP$^+$-dependent enzyme isocitrate dehydrogenase 1 (IDH1) cause increased production of the oncometabolite 2-hydroxyglutarate (2-HG) in multiple cancers. In the vast majority of IDH1 mutant tumors, there is retention of one wild-type (WT) IDH1 allele. Several groups have proposed that retention of this WT allele is pro-tumorigenic by facilitating substrate channeling through a WT/mutant IDH1 heterodimer, with the WT subunit generating a local supply of α-ketoglutarate and NADPH that can then be consumed by the mutant subunit to produce 2-HG[165,222]. Here we demonstrate that the increase in 2-HG production observed in IDH1 mutant cells with co-expression of WT IDH1 is indeed associated with the formation of WT/mutant IDH1 heterodimers. From analysis of a crystal structure of the WT/R132H IDH1 heterodimer and *in vitro* kinetic parameters for 2-HG production, however, we calculate that substrate channeling between subunits is biophysically implausible. We also find that putative carbon substrate flux between WT and mutant IDH1 is inconsistent with the results of isotope tracing experiments in cancer cells harboring an endogenous, monoallelic IDH1 mutation. Finally, we use a mathematical model of WT/mutant IDH1 heterodimers to estimate that the NADPH/NADP$^+$ ratio is higher in the cytosol than in the mitochondrion. These findings argue against substrate supply being limiting for 2-HG production by cytosolic mutant IDH1, and suggest that the retention of a WT allele in IDH1 mutant tumors is not due to a requirement for carbon or cofactor flux between WT and mutant IDH1.

4.1 Introduction

The metabolic enzymes isocitrate dehydrogenase 1 (IDH1) and 2 (IDH2) are mutated in numerous cancers, including glioma, acute myeloid leukemia (AML), chondrosarcoma, and cholangiocarcinoma[158,235,140,223,5,22]. These mutations confer a neomorphic activity for the NADPH-dependent reduction of α-ketoglutarate (α-KG) to R(-)-2-hydroxyglutarate (2-HG), an oncometabolite that is believed to competitively inhibit a range of α-KG-dependent enzymes involved in epigenetic regulation[42,71,223]. A small-molecule inhibitor of mutant IDH2 was recently approved for treatment of IDH2-mutant AML, and other inhibitors of both mutant IDH1 and IDH2 are in clinical trials for cancer[237,204]. 2-HG is also an established biomarker of disease burden in multiple IDH mutant malignancies[21,44].

With few exceptions, mutations in IDH1 are monoallelic with retention of a wild-type (WT) allele. Explanations for retention of the WT allele in tumors remain controversial. One possible reason is that mutant IDH might deplete cellular NADPH levels[128,66]. In this context, retention of a WT IDH allele could mitigate the depletion of NADPH, minimizing toxic accumulation of ROS or impairment of biosynthetic capacity. As the kinetics of the NADPH-consuming reaction of mutant IDH proteins are slow relative to the NADPH-producing reactions of WT IDH and of other enzymes[42,223,71], however, 2-HG production may not be sufficient to deplete NADPH except in engineered cell systems with mutant IDH1 overexpressed.

Alternatively, a retained WT subunit could be critical for optimum 2-HG production by a mutant subunit as part of a WT/mutant IDH heterodimer. Ward *et al.* showed previously that for the cytosolic isoform IDH1, but not the mitochondrial isoform IDH2, retention of the WT allele in cells is indeed important for maximizing 2-HG production by the mutant[222]. Others have disputed these findings based on experiments performed with transfected cells in culture[175]. There exist, however, rare human glioblastoma samples with IDH1 mutations in which the WT IDH1 allele is eventually lost. In these patient-derived samples, there is a nearly 10-fold decrease in intra-tumor 2-HG concomitant with loss of the WT allele[105]. *In vitro* assays of α-KG and NADPH consumption in IDH mutant cell lysates have also suggested synergy between the WT and mutant IDH activities[42].

Several explanations have been proposed for the dependency of mutant IDH1 activity on WT IDH1. It is possible that the enzymatic activity of WT/mutant heterodimers is intrinsically superior to that of mutant/mutant homodimers, as suggested by some studies of recombinant IDH1 proteins[24,25]. Others, however, have found no advantage for the recombinant IDH1 heterodimer[124]. Alternatively, the WT-dependence might be due to substrate channeling or some other form of cooperativity between the WT and mutant subunits, as previously proposed[222,12]. Pietrak *et al.* demonstrated that both the WT activity for oxidative decarboxylation of isocitrate and the mutant activity for reduction of α-KG to 2-HG remain intact in a WT/mutant IDH1 heterodimer, suggesting that channeling could be involved[165]. The dependence of 2-HG production on WT IDH1 might also reflect differences in substrate availability in the cytosol and mitochondria, as levels of α-KG and NADPH may differ between compartments.

Here we report a lack of evidence for substrate channeling or metabolic flux between WT and mutant IDH1. We first considered the possibility that α-KG and NADPH are produced and maintained in a protected pool around the heterodimer by the WT subunit. On the basis of existing structural evidence and *in vitro* IDH1 enzyme kinetics, we calculate that such channeling is biophysically implausible, even for densely packed IDH1 heterodimers. After ruling out substrate channeling, we then characterized the extent of all carbon flux through WT IDH1. Using stable isotope tracing, we show that 2-HG is predominantly derived from glutamine in IDH1 mutant cancer cells, with no need to invoke flux of α-KG between subunits. Finally, through quantitative modeling of IDH1 WT/mutant heterodimers, we estimate the mitochondrial to cytosolic NADPH/NADP$^+$ ratio, which suggests that NADPH is unlikely to be limiting for the production of 2-HG by mutant IDH1 in the cytosol.

4.2 RESULTS

4.2.1 CO-EXPRESSION OF WT AND MUTANT IDH1 INCREASES 2-HG PRODUCTION AND IS ASSOCIATED WITH THE FORMATION OF WT/MUTANT IDH1 HETERODIMERS

WT IDH1 in the cytosol can produce α-KG and NADPH. Mutant IDH1 consumes α-KG and NADPH, producing the oncometabolite 2-HG and regenerating NADP$^+$ (Fig. 4.1A). Most cancer-associated IDH1 mutations occur in monoallelic fashion, with retention of one WT IDH1 allele in the tumor cells. It was shown previously that co-expression of WT and R132H mutant IDH1 in cells results in greater 2-HG production than expression of equimolar amounts of mutant IDH1 alone[222]. In these prior experiments, however, the generation of WT/mutant IDH1 heterodimers was not formally demonstrated. Here we confirm previous findings that the co-expression of WT and mutant IDH1 increases 2-HG production (Fig. 4.1B). In HEK 293T cells already expressing endogenous WT IDH1, expression of increasing amounts of FLAG-tagged R132H mutant IDH1 resulted in increasing 2-HG production. Co-expression of Myc-tagged WT IDH1 further enhanced the production of 2-HG, consistent with previous findings[222]. In addition, we show that the exogenous co-expression of WT and mutant IDH1 resulting in enhanced 2-HG production is associated with the formation of WT/mutant IDH1 heterodimers in cells. 48 h following transfection with FLAG-tagged R132H mutant IDH1 or FLAG-tagged R132H mutant IDH1 and Myc-tagged WT IDH1, cells either were quenched rapidly in 80% MeOH for subsequent metabolite extraction, or lysed and subjected to FLAG-immunoprecipitation. Myc-tagged WT IDH1 was found to co-precipitate with FLAG-tagged R132H mutant IDH1 in cells in which they were co-expressed (Fig. 4.1B).

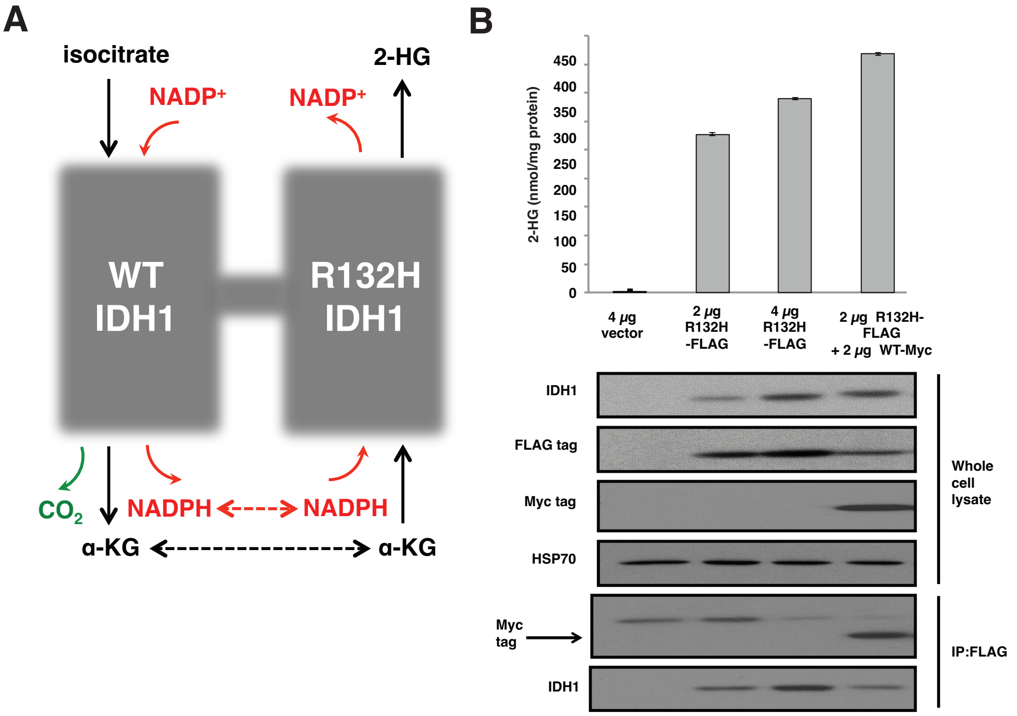


Figure 4.1: 2-HG production in the cytoplasm requires WT IDH1 activity. A Schematic of the WT and neomorphic reactions catalyzed by the WT/mutant IDH1 heterodimer. WT IDH1 catalyzes the NADP$^+$-dependent oxidative decarboxylation of isocitrate to α-KG. Mutant IDH1 reduces α-KG to the oncometabolite 2-HG. The dashed arrows indicate potential opportunities for substrate channeling between the WT and mutant subunits of the IDH1 heterodimer. B FLAG-tagged R132H mutant IDH1 (at various doses), FLAG-tagged R132H mutant IDH1 plus Myc-tagged WT IDH1, or empty vector were transfected into HEK-293T cells. Cells were harvested 48 h post-transfection and assayed for 2-HG accumulation by GC-MS (top) or protein expression by immunoprecipitation and western blot (bottom). Absolute 2-HG levels were determined by normalization of the 2-HG GC-MS signal intensity to that of a deuterated 2-HG internal standard and then normalized to total cellular protein. Error bars denote one standard deviation.

WT		R132H	
K_M, isocitrate (μM)	6.88 ± 0.97	K_M, α-KG (μM)	988 ± 77
K_M, NADP$^+$ (μM)	6.79 ± 0.31	K_M, NADPH (μM)	1.36 ± 0.39
k_{cat} (s^{-1})	3.10 ± 0.14	k_{cat} (s^{-1})	0.0666 ± 0.0016

Table 4.1: *In vitro* kinetics of WT and R132H IDH1. All uncertainty values are standard errors.

4.2.2 *IN VITRO* KINETIC CHARACTERIZATION OF IDH1

A number of recent studies have reported *in vitro* kinetic parameters for WT IDH1 and for various oncogenic mutants[42,236,165,171,25,12]. There is substantial variability in the values reported; for instance, measurements of k_{cat} for the WT enzyme range from 8.6 to 44,000 s^{-1}, and from 0.19 to 1,000 s^{-1} for R132H IDH1. To investigate this discrepancy and to obtain values for mathematical modeling, we measured the *in vitro* kinetics of recombinant WT and R132H IDH1 (Table 4.1). The parameters obtained are in close agreement with the prior values reported in two studies of the enzymatic mechanism of IDH1[165,171] and are used for all subsequent calculations.

4.2.3 SUBSTRATE CHANNELING IS BIOPHYSICALLY IMPLAUSIBLE

As discussed above, the neomorphic reaction catalyzed by mutant IDH1 requires α-KG and NADPH, both of which are produced in the WT reaction (Fig. 4.1A). As such, it has been suggested that cytoplasmic 2-HG synthesis might be enhanced by the local production of α-KG and NADPH in WT/mutant heterodimers, and that this effect might explain the dramatic reduction of 2-HG levels following loss of the WT enzyme[222]. As an initial test of this hypothesis, we performed a biophysical analysis of putative substrate channeling between the WT and mutant subunits.

A crystal structure of the IDH1 WT/R132H heterodimer at 3.2 Å resolution has been reported previously (Fig. 4.2A)[236]. The structure reveals that there is no physical association between the active sites of the WT and the mutant, suggesting that any interaction would be due to a local elevation of [α-KG] and [NADPH] around each heterodimer. We first examined diffusion of α-

KG and NADPH away from the WT active site to assess the feasibility of achieving local enrichment. From inspection of the crystal structure, we estimated that the straight-line distance between the two active sites is approximately 37.3 Å. Assuming $D_{\alpha KG} = 6.7 \times 10^{-10}$ m^2/s and $D_{NADPH} = 4.1 \times 10^{-10}$ m^2/s (see Experimental Procedures) and three-dimensional diffusion governed by

$$t = \frac{< x^2 >}{6D},$$

it would take just 3.5 ns and 5.7 ns for α-KG and NADPH, respectively, to travel the distance between the active sites. In 15 s, the time required for one turnover of the mutant enzyme, both substrates could travel over 3000 times further than the inter-site distance. This simple calculation suggests that the local and bulk cytoplasmic concentrations of α-KG and NADPH are likely equilibrated on catalytically relevant timescales. As our estimate for inter-site distance is based on a crystal structure, it does not account for possible dynamic changes in the conformation of the two subunits, which could bring the active sites into closer proximity. It is plausible that such shifts do occur in the IDH1 heterodimer, as certain inhibitors of the mutant enzyme are known to bind to an allosteric pocket on the WT subunit, and the activity of WT IDH1 is controlled in part by the highly flexible α10 regulatory element[233,231]. Any dynamic change, however, would have to reduce the straight-line distance by at least several orders of magnitude for local elevation of substrate concentration to be relevant.

Packing of enzymes into dense clusters has been suggested to enable channeling even in the absence of physical association of active sites. Castellana *et al.* reported a detailed theoretical analysis of such channeling by agglomeration[30]. Using their framework, we investigated whether $N = 4000$ IDH1 heterodimers confined in a sphere with radius $r = 40$ nm (i.e., at the dense-packing limit $n_{max} = 25$ mM) could efficiently channel α-KG and NADPH. Efficient channeling requires that

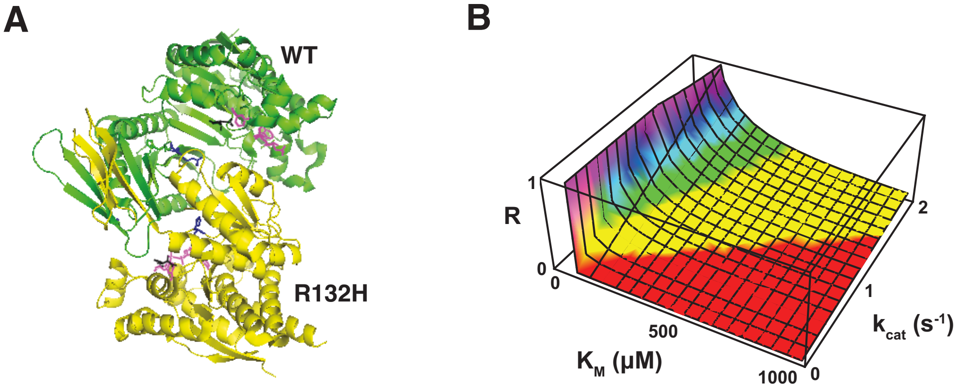


Figure 4.2: Biophysical evaluation of putative substrate channeling between WT and R132H IDH1 subunits. A Crystal structure of the WT/R132H IDH1 heterodimer. The WT subunit is colored green, the mutant yellow. Residue 132 is colored blue in both subunits. Isocitrate (black) and NADP(H) (red) are bound in the respective active sites. Drawn from data reported in [236]. B Channeling by enzyme agglomeration is infeasible for the IDH1 WT/R132H heterodimer. Plot of the ratio of the processing rate to escape rate as a function of K_M (for either α-KG or NADPH) and k_{cat} for $N = 4000$ heterodimers clustered in a sphere with radius $r = 40$ nm.

the processing rate

$$R_P = \frac{3\kappa N}{4\pi r^3}$$

exceed the escape rate

$$R_E = \frac{D}{r^2},$$

where κ is the catalytic efficiency of the enzyme. For this calculation, we do not need to make any assumptions about the relative positions of the WT and mutant active sites. The R_P/R_E ratio for carbamoyl phosphate synthetase and aspartate carbamoyl transferase in *Escherichia coli*, a system for which Castellana *et al.* obtained experimental evidence of channeling by agglomeration, is approximately 9.5[30]. In contrast, we calculated ratios of 4.0×10^{-8} and 4.7×10^{-3} for the processing of α-KG and NADPH, respectively, by the dense IDH1 cluster. Further analysis indicated that submicromolar K_M values would be required for both substrates to achieve parity given the relatively slow k_{cat} for R132H IDH1 (Fig. 4.2B).

4.2.4 2-HG production in IDH1 mutant cancer cells is predominantly from glutamine and can be explained without carbon flux between WT and mutant subunits

The most direct route of 2-HG production from glucose involves production of cytosolic α-KG by WT IDH1, followed by consumption of this α-KG by mutant IDH1 to generate 2-HG. In contrast, 2-HG production by mutant IDH1 from glutamine-derived α-KG can occur without production of α-KG by WT IDH (Fig. 4.3A). We grew HT1080 fibrosarcoma cells, which harbor an endogenous IDH1 mutation, either in glucose-free media supplemented with uniformly labeled glucose([U-^{13}C] glucose), or in glutamine-free media supplemented with uniformly labeled glutamine ([U-^{13}C] glutamine). As shown in Supplemental Fig. C.S1, [U-^{13}C] glucose can be converted via glycolysis to pyruvate with three ^{13}C atoms (M+3 pyruvate), which subsequently can enter the TCA cycle as M+2 acetyl-CoA, with further metabolism to M+2 citrate and M+2 isocitrate. Cytosolic pools of M+2 isocitrate can be converted by WT IDH1 to M+2 α-KG, which finally can be reduced to M+2 2-HG. When we followed the fate of glucose carbon by growing HT1080 cells in media with U-^{13}C glucose (and unlabelled glutamine), we found minimal incorporation of glucose-derived carbon into 2-HG (Fig. 4.3B). In contrast, [U-^{13}C] glutamine can be converted via glutaminolysis to fully labeled M+5 α-KG, which can then be reduced to fully labeled M+5 2-HG. Comparing 2-HG production from labeled glucose and labeled glutamine therefore provides information about the extent of carbon flux between the WT to mutant subunits in the IDH1 heterodimer (Fig. 4.3A). For HT1080 cells grown in media with U-^{13}C glutamine (and unlabelled glucose), we found that nearly the entire 2-HG pool was labeled with carbon derived from glutamine (Fig. 4.3B), suggesting minimal importance of inter-subunit flux.

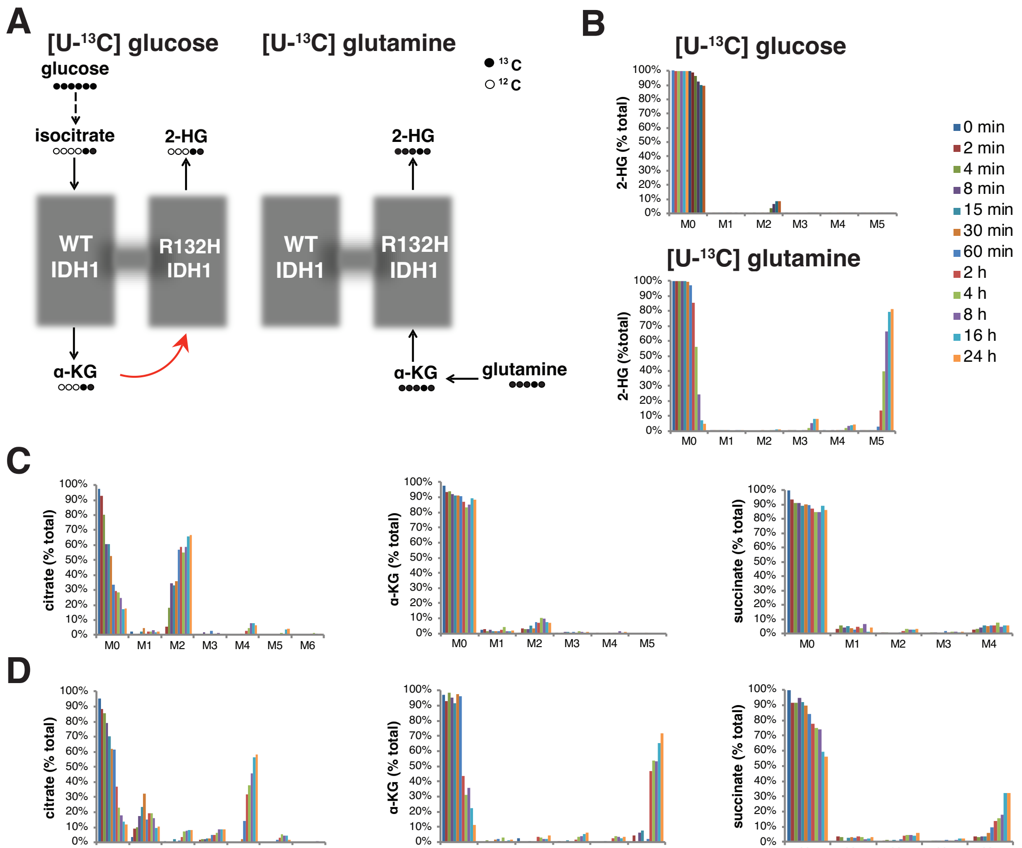


Figure 4.3: *In vivo* evaluation of 2-HG production kinetics. A Schematics of expected labeling when cells are grown on glucose uniformly labeled with ^{13}C (U-^{13}C glucose) or on glutamine uniformly labeled with ^{13}C (U-^{13}C glutamine). U-^{13}C glucose is converted to doubly labeled isocitrate, α-KG, and 2-HG. U-^{13}C glutamine is converted to fully labeled α-KG and 2-HG. Dashed arrows indicate that one or more intermediate reactions have been omitted from the diagram for clarity. The red arrow indicates possible flux of labeled α-KG derived from U-^{13}C glucose. B Kinetics of the labeling of 2-HG with ^{13}C over 24 h of growth on U-^{13}C glucose or U-^{13}C glutamine. M0 ... M6 refer to the number of labeled carbon atoms. C and D Kinetics of the labeling of citrate, α-KG, and succinate with ^{13}C over 24 h of growth on U-^{13}C glucose (C) and U-^{13}C glutamine (D). The color scheme is the same as in B.

4.2.5 Evidence for a branched TCA cycle in cancer cells harboring an IDH1 mutation

Further analysis of the labeling data from IDH1 mutant HT1080 cells revealed that traditional TCA cycle metabolites were not labeled in a pattern consistent with a complete TCA cycle. HT1080 cells labeled with [U-^{13}C] glucose had significant accumulation of M+2 citrate, as predicted (Fig. 4.3C). However, minimal label was found in the downstream TCA cycle metabolites α-KG or succinate, even after 24 h of labeling. In contrast, cells labeled with [U-^{13}C] glutamine exhibited robust production of fully labeled α-KG (M+5 α-KG), as predicted, and exhibited further oxidative metabolism through the TCA cycle to M+4 succinate and then to M+4 citrate (Fig. 4.3D). This M+4 citrate was not further metabolized to make a complete TCA cycle, as indicated by minimal accumulation of M+3 α-KG or M+2 succinate, even at later time points. Minimal accumulation of M+5 citrate was observed, consistent with minimal net flux from α-KG to citrate through reductive carboxylation. Overall, these data demonstrate minimal net flux from citrate to α-KG in these cells. Oxidative metabolism of either glucose or glutamine-derived carbon can generate citrate, but significant metabolism of citrate through aconitase and IDH does not appear to occur in these cells under standard culture conditions. Instead, most citrate can be metabolized to cytosolic acetyl-CoA and oxaloacetate by ATP citrate lyase (Supplemental Fig. C.S1).

4.2.6 Estimation of compartment-specific NADPH levels

In addition to converting isocitrate to α-KG, WT IDH1 also produces cytosolic NADPH, which could potentially be limiting for 2-HG production by mutant IDH1. Rigorous determination of compartment-specific NADPH levels and NADPH/NADP^{+} ratios, however, is technically challenging. We sought therefore to estimate [NADPH] in the cytosol and mitochondria using quantitative modeling of IDH heterodimers. To this end, we developed a biochemically realistic

parameter class	rate constants	value
k_{on}, NADP$^+$	k_1, k_8	1.00×10^8 μM^{-1}s^{-1}
k_{on}, isocitrate	k_3, k_5	1.00×10^8 μM^{-1}s^{-1}
k_{off}, NADP$^+$	k_2, k_9	676 s^{-1}
k_{off}, isocitrate	k_4, k_9	685 s^{-1}
k_{cat}, WT	k_7	3.10 s^{-1}
k_{on}, NADPH	k_{10}	1.00×10^8 μM^{-1}s^{-1}
k_{on}, α-KG	k_{12}	1.00×10^8 μM^{-1}s^{-1}
k_{off}, NADPH	k_{11}	136 s^{-1}
k_{off}, α-KG	k_{13}	9.88×10^4 s^{-1}
k_{cat}, R132H	k_{14}	0.0666 s^{-1}

Table 4.2: Kinetic parameters for reaction network in Fig. 4.4.

mathematical model of the IDH1 WT/mutant heterodimer. The model involves 11 species - IDH1 (denoted E), isocitrate (denoted I), α-KG, NADP$^+$, NADPH, 2-HG, and five enzyme-substrate complexes - and 16 reactions, each with a corresponding kinetic parameter. Fig. 4.4 shows the full reaction network diagram, and Table 4.2 lists numerical estimates for the rate constants (see Experimental Procedures for details of parameter estimation). Following recent experimental studies of the IDH1 enzyme mechanism[171], we model the WT reaction as random sequential, so that either isocitrate or NADP$^+$ may bind first to form an E·I·NADP$^+$ ternary complex (module 1), and the neomorphic reaction as ordered sequential, with NADPH and then α-KG binding to the enzyme to form an E·αKG·NADPH ternary complex (module 2). The production of 2-HG by the IDH1 heterodimer occurs *in vivo* as part of a larger metabolic network. To make the network an open system, we include reactions for the zero-order synthesis of isocitrate and first-order degradation of 2-HG by processes not explicitly modeled (module 3). For simplicity, any reactions involving α-KG, NADP$^+$, or NADPH that occur independent of the IDH1 heterodimer are excluded from the model.

Following mass-action kinetics, the reaction network in Fig. 4.4 gives rise to a system of 11 ordinary differential equations, which reduce at steady state to a polynomial system. The goal of this

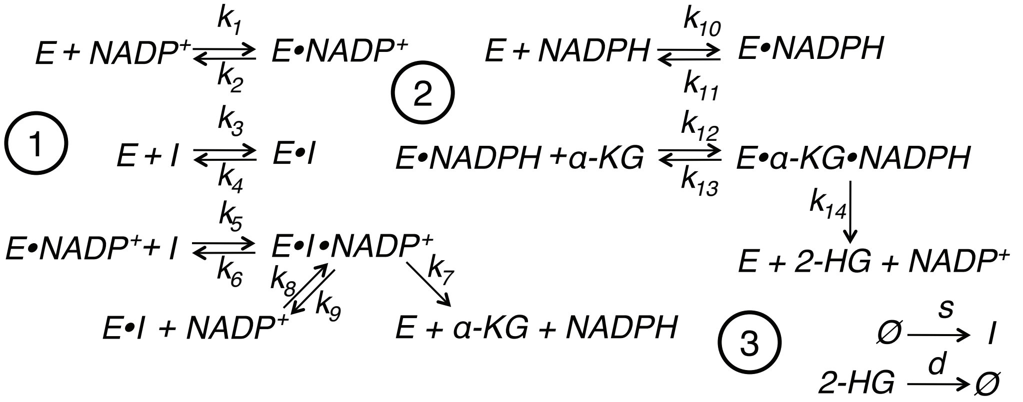


Figure 4.4: Biochemically realistic mathematical modeling of IDH1 heterodimers. Reaction network diagram for a detailed mechanistic model of WT IDH1 (1) and mutant IDH1 (2). The WT conversion of isocitrate to α-KG is assumed to follow a random-order, ternary complex mechanism. The mutant conversion of α-KG to 2-HG is assumed to follow an ordered sequential mechanism, with NADPH binding before α-KG. The reactions in group 3 reflect the synthesis of isocitrate and α-KG and degradation of 2-HG by processes not explicitly modeled. E denotes the WT/mutant IDH1 heterodimer, and I denotes isocitrate.

section is to derive a steady-state expression for [NADPH] in terms of just [I], [α-KG], [NADP$^+$], and the rate constants, which, in combination with literature values for metabolite concentrations, enables estimation of the ratio of mitochondrial to cytosolic NADPH ($[\text{NADPH}]_m/[\text{NADPH}]_c$). At steady state, production and consumption of each of the binary enzyme-substrate complexes involved in the WT reaction (E·NADP$^+$ and E·I) must balance, so that we have

$$k_1[\text{E}][\text{NADP}^+] + k_6[\text{E}\cdot\text{NADP}^+\cdot\text{I}] = (k_2 + k_5[I])\,[\text{E}\cdot\text{NADP}^+] \tag{4.1}$$

and

$$k_3[\text{E}][\text{I}] + k_9[\text{E}\cdot\text{NADP}^+\cdot\text{I}] = \left(k_4 + k_8[\text{NADP}^+]\right)[\text{E}\cdot\text{I}]. \tag{4.2}$$

Similarly, production and consumption of the ternary complex E·NADP$^+$·I must balance, so

$$k_5[\mathrm{I}][\mathrm{E} \cdot \mathrm{NADP}^+] + k_8[\mathrm{NADP}^+][\mathrm{E} \cdot \mathrm{I}] = (k_6 + k_7 + k_9)\,[\mathrm{E} \cdot \mathrm{NADP}^+ \cdot \mathrm{I}]. \tag{4.3}$$

Using Eqs. 4.1 and 4.2 in Eq. 4.3 yields an expression for the ternary complex in terms of [E], [I], and [NADP$^+$]:

$$[\mathrm{E} \cdot \mathrm{NADP}^+ \cdot \mathrm{I}] = \frac{[\mathrm{E}][\mathrm{I}][\mathrm{NADP}^+]\,(k_3 k_8\,(k_2 + k_5[\mathrm{I}]) + k_1 k_5\,(k_4 + k_8[\mathrm{NADP}^+]))}{k_5[\mathrm{I}]\,(k_4\,(k_7 + k_9) + k_7 k_8[\mathrm{NADP}^+]) + k_2\,(k_4\,(k_6 + k_7 + k_9) + (k_6 + k_7)\,k_8[\mathrm{NADP}^+])}. \tag{4.4}$$

The mutant reactions can be analyzed in an analogous fashion. From steady-state balance of the binary ([E·NADPH]) and ternary ([E·α-KG·NADPH]) complexes, we have

$$k_{10}[\mathrm{E}][\mathrm{NADPH}] + k_{13}[\mathrm{E} \cdot \alpha\text{-KG} \cdot \mathrm{NADPH}] = (k_{11} + k_{12}[\alpha\text{-KG}])\,[\mathrm{E} \cdot \mathrm{NADPH}]$$

and

$$k_{12}[\alpha\text{-KG}][\mathrm{E} \cdot \mathrm{NADPH}] = (k_{13} + k_{14})\,[\mathrm{E} \cdot \alpha\text{-KG} \cdot \mathrm{NADPH}],$$

which together give

$$[\mathrm{E} \cdot \mathrm{NADPH}] = \frac{(k_{10}(k_{13} + k_{14})\,[\mathrm{E}][\mathrm{NADPH}]}{k_{11}\,(k_{13} + k_{14}) + k_{12}k_{14}[\alpha\text{-KG}]}. \tag{4.5}$$

Finally, balance of production and consumption of [NADPH] gives

$$k_{11}[\mathrm{E} \cdot \mathrm{NADPH}] + k_7[\mathrm{E} \cdot \mathrm{NADP}^+ \cdot \mathrm{I}] = k_{10}[\mathrm{E}][\mathrm{NADPH}]. \tag{4.6}$$

Using Eqs. 4.4 and 4.5 in Eq. 4.6 yields

$$[\mathrm{NADPH}] = \frac{k_7[\mathrm{I}]\left(k_{11}\left(k_{13}+k_{14}\right)+k_{12}k_{14}[\alpha\text{-}\mathrm{KG}]\right)[\mathrm{NADP}^+]\left(k_3k_8\left(k_2+k_5[\mathrm{I}]\right)+k_1k_5\left(k_4+k_8[\mathrm{NADP}^+]\right)\right)}{k_{10}k_{12}k_{14}[\alpha\text{-}\mathrm{KG}]\left(k_5[\mathrm{I}]\left(k_4\left(k_7+k_9\right)+k_7k_8[\mathrm{NADP}^+]\right)+k_2\left(k_4\left(k_6+k_7+k_9\right)+\left(k_6+k_7\right)k_8[\mathrm{NADP}^+]\right)\right)}, \tag{4.7}$$

as desired. Eq. 4.7 can also be obtained directly from a Gröbner basis calculation (see Experimental Procedures for details). Crucially, the expression is independent of [E] and of the concentrations of the various enzyme-substrate complexes, which are inaccessible to measurement.

In contrast to IDH1, the enzymatic mechanism of the mitochondrial isoform IDH2 has not been charactered in detail. It is plausible, however, that WT/R172K IDH2 heterodimers, which are structurally analogous to WT/R132H IDH1, can be modeled by the reaction network in Fig. 4.4. Making this assumption, $[\mathrm{NADPH}]_m/[\mathrm{NADPH}]_c$ can be estimated from just Eq. 4.7 and mitochondrial and cytosolic values for [I], [α-KG], and [NADP$^+$]. Using a novel technique for absolute quantification of mitochondrial matrix metabolites, Chen *et al.* recently reported compartment-specific concentrations of cis-aconitate (25 μM matrix, 800 μM whole-cell), α-KG (150 μM, 1200 μM), and NADP$^+$ (30 μM, 75 μM) , but not of NADPH[31]. Using the cis-aconitate values as a proxy for [I] and assuming that cytosolic metabolite concentrations are well-approximated by the whole-cell measurements, we calculate $[\mathrm{NADPH}]_m/[\mathrm{NADPH}]_c = 0.10$. As Chen *et al.* reported $[\mathrm{NADP}^+]_m/[\mathrm{NADP}^+]_c = 0.40$, this estimate implies $([\mathrm{NADPH}]_m/[\mathrm{NADP}^+]_m)/([\mathrm{NADPH}]_c/[\mathrm{NADP}^+]_c)$ $= 0.25$.

4.3 DISCUSSION

A nearly universal feature of IDH1 mutant cancers is the retention of a WT IDH1 allele. This monoallelic nature of IDH1 mutations was initially difficult to reconcile with the observation that IDH mutations result in loss of function for WT activity (conversion of isocitrate to α-KG)[235]. A initial explanation for this paradox invoked a dominant negative mechanism, whereby a mutant IDH1 subunit would heterodimerize with and inhibit the activity of a WT IDH1 subunit. This mechanism was suggested to promote tumorigenesis due to decreased α-KG levels and a resultant increase in HIF-1α expression[238]. In assays of isocitrate-dependent NADPH production in lysates from HEK 293T cells co-expressing WT and mutant IDH1, however, no appreciable inhibition of WT enzyme activity was found[42]. In these same lysates, α-KG-dependent NADPH consumption consistent with a gain of function for a reductive activity was observed. This mutant IDH1 activity was also enhanced with co-expression of WT IDH1[42], suggesting that the mutant IDH1 activity facilitated a novel reaction rather than a mere reversal of the normal WT oxidative decarboxylation of isocitrate. This neomorphic activity of mutant IDH1 was found to be the NADPH-dependent reduction of α-KG to the oncometabolite 2-HG, which is a common feature of the vast majority of recurrent IDH1/2 mutations in cancer[42,71,223]. Explanations for the observed synergism between WT and mutant IDH1, however, are controversial.

It was subsequently confirmed that co-expression of WT and mutant IDH1 greatly enhanced 2-HG production compared to mutant IDH1 expression alone[222]. The importance of retained WT IDH1 to 2-HG production by mutant IDH1 was also demonstrated in studies of IDH1 mutant glioma samples exhibiting spontaneous loss of the WT allele[105]. Frequently invoked models to explain the retention of a WT IDH1 allele invovle either substrate channeling or bulk metabolic flux between WT and mutant subunits in an IDH1 heterodimer[222,12,165]

The data presented here argue against both of the above models. It remains critical, however, to understand how WT and mutant IDH synergize *in vivo* in human cancers. We present evidence that the primary carbon source for 2-HG production in a cancer cell line harboring an endogenous IDH1 mutation is glutamine. In a direct metabolic pathway involving glutamine → glutamate → α-KG → 2-HG, there is no need to invoke flux through a WT IDH subunit to account for 2-HG production. Consistent with our results, Salamanca *et al.* recently reported that the majority of 2-HG produced by IDH both in cultured cancer cells and *in vivo* is derived from glutamine, facilitating *in vivo* monitoring approaches for IDH mutant solid tumors with hyperpolarized [1-^{13}C] glutamine[181]. In contrast to us and to Salamanca *et al.*, Gelman *et al.* reported that both glucose and glutamine carbon contribute substantially to 2-HG production by mutant IDH1[66]. Gelman *et al.* relied on cells that were engineered to express the IDH1 mutation artificially, while we and Salamanca *et al.* performed metabolic tracing experiments in cell lines harboring naturally occurring IDH1 mutations at the endogenous locus, possibly accounting for the divergent results obtained.

^{13}C labeling approaches, however, do not address the source of required redox cofactors. It may still be possible that WT IDH1 is necessary for generating an optimum local supply of NADPH for 2-HG production by mutant IDH1 in the cytosol. We estimate here, however, that the ratio of NADPH/NADP^{+} is higher in the cytosol than in the mitochondrial matrix, which suggests that, if NADPH production by a WT IDH subunit were essential for a mutant IDH subunit to produce 2-HG, it would be at least as important for mitochondrial IDH2 as for cytosolic IDH1. Arguing against this prediction is the prior report that the WT IDH2 allele is dispensable for mutant IDH2-dependent 2-HG production in cells[222]. Several groups have also reported loss of the WT IDH2 allele in IDH2 mutant tumors, suggesting that WT IDH2 is dispensable *in vivo*[157,164]. We conclude that metabolic flux of either carbon substrate or redox cofactors between WT and mutant IDH does not contribute substantially to 2-HG production in cancer.

An alternative, not mutually exclusive, explanation for the retention of one WT IDH1 allele in IDH1 mutant tumors is that the WT/mutant IDH1 heterodimer may harbor a structural or biochemical advantage over mutant/mutant IDH1 homodimers *in vivo*. Initial studies with recombinant IDH proteins were not universally supportive of this hypothesis[124], although subsequent data have prompted reconsideration. In particular, Brooks *et al.* demonstrated a lower K_M for α-KG in the WT/mutant IDH1 heterodimer than in the R132H/R132H homodimer[25]. In light of the evidence we present against inter-subunit flux, the possibility of intrinsic biochemical superiority of the heterodimer, perhaps involving allosteric modulation of the mutant subunit by the WT[231], warrants further consideration. This possibility is of particular importance given that many recently developed inhibitors of mutant IDH act at the dimerization interface. For IDH1, it has been reported that the R132H mutation destabilizes a regulatory segment (the $\alpha 10$ helix) at the interface where several inhibitors bind[231]. With ongoing trials of IDH inhibitors, one might predict that tumors will evolve therapeutic resistance through mechanisms involving both the mutant and WT IDH alleles, potentially at residues impacting the dimer interface.

4.4 EXPERIMENTAL PROCEDURES

4.4.1 CELL CULTURE

HT1080 fibrosarcoma cells (which harbor an endogenous, monoallelic R132C IDH1 mutation) and HEK 293T cells (which lack endogenous IDH mutations) were cultured routinely in Dulbecco's modified Eagle's medium (DMEM), 10% fetal bovine serum, 25 mM glucose, and 6 mM glutamine. For metabolite tracing experiments, DMEM without glutamine or pyruvate (Gibco) and with 10% dialyzed FCS was supplemented with 4 mM [U-^{13}C]-L-glutamine (Aldrich 605166). DMEM without glucose or pyruvate (Gibco) was supplemented with 10 mM [U-^{13}C]-D-Glucose (Cambridge Isotope Laboratories CLM-1396). All cells were in logarithmic growth phase during the entire period of labeling.

4.4.2 METABOLITE EXTRACTION AND GC-MS ANALYSIS

Following gentle removal of culture medium from proliferating cells, cells were rapidly quenched with 80% methanol pre-chilled to -80 °C. For experiments where absolute quantification of 2-HG was performed, this 80% methanol was spiked with a M+5 internal standard of R(-)-2-HG containing five deuterium atoms (D-2-hydroxyglutaric-2,3,3,4,4-d5 acid; details for synthesis are provided in[222]). Following incubation at -80 °C for at least 30 min, cell extracts in 80% MeOH were collected, sonicated, and centrifuged at 14,000 g for 20 min at 4 °C to remove precipitated protein. Supernatants next were dried under nitrogen gas, redissolved in 20 μl MOX reagent (Thermo Scientific), and heated at 37 °C for 90 min, followed by addition of 25 μl of tert-butyldimethylchlorosilane (t-BDMS; Regis Technologies) and heating at 60 °C for 60 min. Derivatized samples were analyzed by GC-MS using a DB-35MS column (30.25 mm i.d., Agilent J&W Scientific) installed in an Agilent 7890A gas chromatograph (GC) interfaced with an Agilent 5975C mass spectrometer (MS). Mass isotopomer distributions were determined by integrating metabolite ion fragments and corrected for natural abundance using in-house algorithms adapted from[60] and described in further detail in[125]. Absolute 2-HG levels were obtained by quantifying the peak area of the ion at m/z 433, formed through the loss of a t-butyl group (-57 amu) from the molecular ion tri-TBDMS-2-HG, and normalizing to the peak area of the m/z 438 ion (representing the analogous derivative of the d5-2-HG internal standard spiked at known concentration) and the total cellular protein as measured by BCA.

4.4.3 Protein harvest and quantitation, western blot, and immunoprecipitation

Cells were lysed 48 h following transfection with mammalian protein extraction reagent (Pierce) supplemented with protease inhibitor cocktail (Complete Mini, EDTA-free, Roche 11-836-170-001) and phosphatase inhibitor cocktails 2 and 3 (Sigma). Lysates were sonicated with 2 x 30 s pulses and then centrifuged at 14,000 g for 20 min at 4 °C. Supernatants were subsequently collected and assessed for protein concentration with BCA Protein Assay (Pierce). For western blotting, lysates were separated by SDS-PAGE on 10% polyacrylamide gels, transferred to PVDF membranes, and blocked in 5% non-fat milk in PBS containing 0.2% Tween-20. Primary antibodies used were: anti-IDH1 (Santa Cruz sc-49996, 1:200 dilution), anti-IDH2 (Abcam ab55271, 1:500), anti-HSP70 (Cell Signaling 48729, 1:1000), anti-FLAG (Cell Signaling 2368, 1:10,000), and anti-Myc tag (Babco, clone 9e10, 1:1000). Detection was performed with horseradish-peroxidase-conjugated anti-rabbit, anti-mouse, or anti-goat antibodies (GE Healthcare NA934V, GE Healthcare NA931V, and Santa Cruz sc-2020, all 1:10,000 dilution).

For immunoprecipitation experiments, concentrated cell lysates in M-PER lysis buffer containing protease inhibitors as above was diluted to 500 μl total volume with hypotonic lysis buffer (20 mM HEPES, 5 mM KCl, 1 mM $MgCl_2$, pH 7.0) and supplemented with dithiothreitol (DTT) to 5 mM. Anti-FLAG M2 affinity gel resin (Sigma A220, lot SLBG5784V, 20 μl beads) was washed in hypotonic lysis buffer plus protease inhibitor. 40 μl of washed bead suspension was mixed with 500 μl of diluted whole cell lysate, centrifuged for 3 min at 1000 rpm at 4 °C, then washed twice with hypotonic lysis buffer and protease inhibitor. Elution weas performed by resuspending beads gently in 3 μl of 5 mg/ml 3x FLAG peptide (Sigma F4799) and 97 μl hypotonic lysis buffer plus protease inhibitor plus DTT, rotating 30 min at 4 °C, then centrifugation for 1 min at 21,000 g. The supernatant from this final spin was saved as the eluate and separated by SDS-PAGE as above.

4.4.4 Plasmid construction and transfection

The cDNA clone of human IDH1 (BC012846.1) was obtained from American Type Culture Collection. R132H IDH1 point mutation was generated as previously described[25]. FLAG or Myc tags were added to the C termini of the open reading frames by standard PCR techniques, as detailed in[222]. Integrity of constructs was confirmed by direct sequencing prior to transfection into HEK 293T cells in pCMV-Sport6 expression vector using Lipofectamine 2000, according to the manufacturer's instructions.

4.4.5 *In vitro* kinetics of WT and R132H IDH1

N-terminally 6xHis-tagged recombinant IDH1 (WT or R132H) was expressed from pET28a(+) in BL21 *E. coli* by inducing an OD_{600} = 0.7 culture with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside for 6 h at room temperature. IDH1 was batch purified using Ni-NTA agarose beads (Qiagen), and purified protein was dialyzed against 50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 25 mM NaCl, 20% glycerol, 0.15% β-mercaptoethanol. Enzyme concentration was determined by Bradford assay using a bovine serum albumin standard. Enzymatic activity was monitored by a change in absorbance at 340 nm, corresponding to production or consumption of NADPH. Each reaction was carried out at room temperature in a 100 μL volume containing buffer (100 mM Tris-HCl, pH 7.5, 1.3 mM $MnCl_2$) and varying concentrations of substrates. To assay WT IDH1 activity, 10 ng of enzyme was included in each reaction, and either $NADP^+$ was titrated in the presence of 1 mM DL-isocitrate, or DL-isocitrate was titrated in the presence of 1 mM $NADP^+$. (The enzyme is specific to D-isocitrate, so the effective concentration of D-isocitrate is half that of total DL-isocitrate.) To assay R132H IDH1 activity, 10 μg of enzyme was included in each reaction, and either α-KG was titrated in the presence of 500 μM NADPH, or NADPH was titrated in the presence of 10 mM α-KG. For each assay, activity was compared to that of a 'no-enzyme' control reaction. K_m and k_{cat} values were determined by fitting the initial-rate data to the Michaelis-Menten model in GraphPad Prism (GraphPad Software, Inc.).

4.4.6 Parameter estimation

Kinetic parameters for the mathematical model of 2-HG production ($k_1 \cdots k_{14}$) were estimated from the data in Table 4.1. The parameters fall into 10 major classes, on-rate and off-rate (for isocitrate and $NADP^+$) and catalytic rate for the WT reaction, and on-rate and off-rate (for α-KG and NADPH) and catalytic rate for the neomorphic reaction. Within a particular class, all parameters are assumed to have an identical value. We set the on-rates for both reactions to 1×10^8 M^{-1}s^{-1}, a standard value for diffusion-limited interactions[37]. We set the catalytic rate constants for WT and R132H equal to the k_{cat} values that we determined experimentally (Table 4.1). Finally, we calculated off-rates from the on-rates, catalytic rates, and experimentally determined Michaelis-Menten constants according to the formula:

$$K_M = \frac{k_{off} + k_{cat}}{k_{on}}.$$

Estimation of diffusion coefficients. Diffusion coefficients for α-KG and NADPH were computed using the Stokes-Einstein equation

$$D = \frac{k_B T}{6\pi\eta r},$$

where T is the temperature, η is the dynamic viscosity, r is the molecular radius, and k_B is the Boltzmann constant. We set $T = 310$ K and $\eta = 1.0$ kg m^{-1} s^{-1}. To determine r, we computed molecular volume using the formula of Abraham and McGowan,

$$V = \Sigma V_a - 6.56 \Sigma N_b,$$

where V_a is the volume of constituent atoms and N_b is the number of covalent bonds[1]. This method yielded radii of 3.4×10^{-10} m and 5.5×10^{-10} m for α-KG and NADPH, respectively, corresponding to $D_{\alpha-KG} = 6.7 \times 10^{-10}$ m^2/s and $D_{NADPH} = 4.1 \times 10^{-10}$ m^2/s. The calculated values are close to experimentally determined literature values for the diffusion coefficients of related molecules such as citrate and NADH[201,85,182].

4.4.7 Algebraic calculations

All algebraic calculations were done in Mathematica 9.0 (Wolfram Research). Gröbner basis calculations on the steady-state polynomial system were performed following the procedure described previously in[139] and applied to a prior mathematical analysis of bacterial IDH[48]. In brief, we computed a lexicographically ordered Gröbner basis for the 11 polynomials using the built-in Mathematica function GroebnerBasis. The variable ordering was [E · NADPH · α-KG], [E · NADP$^+$ · I], [E · I], [E · NADPH], [E · NADP$^+$], [E], [I], [α-KG], [2-HG], [NADP$^+$], and [NADPH]. Eq. 4.7 follows directly from one of the terms of the Gröbner basis.

4.5 Acknowledgements

We thank members of the Vander Heiden laboratory for helpful discussions. J.P.D. was supported by a NSF Graduate Research Fellowship (GE1144152).

5

Model discrimination for Ca^{2+}-dependent regulation of myosin light chain kinase in smooth muscle contraction

AUTHOR LIST

Joseph P. Dexter, John W. Biddle, and Jeremy Gunawardena

CHAPTER SUMMARY

Excitation-contraction coupling in smooth muscle is mediated by the Ca^{2+}- and calmodulin-dependent regulation of myosin light chain kinase. The precise mechanism of this regulation remains a matter of debate, and several mathematical models have been proposed for the interaction of the three species. These models have previously been analyzed at steady state primarily by numerical simulation of differential equations, for which parameter values must be estimated from data and constrained to satisfy detailed balance at thermodynamic equilibrium. Here we use the linear framework for timescale separation to demonstrate that models of this general kind can be solved analytically for an equilibrium steady state, without having to determine numerical values for parameters. This analysis leads to parameter-independent methods for discriminating between the models, for which we propose experiments that could be feasibly performed with existing technology.

5.1 Introduction

Smooth muscle is non-striated, involuntary muscle that surrounds hollow organs such as the stomach, intestines, bladder, and uterus. As with skeletal and cardiac muscle, the contraction of smooth muscle is a Ca^{2+}-dependent process[224,91,177]. The specific mechanisms of contractile regulation, however, are distinct in smooth muscle, which does not contain troponin. In response to various contractile agonists, Ca^{2+} enters the cytosol and binds to the multifunctional protein calmodulin (CaM), which has four Ca^{2+} binding sites, two on the N-terminus and two on the C-terminus (Fig. 5.1A)[217]. Binding of Ca^{2+}-CaM to myosin light chain kinase (MLCK) induces a conformational change that displaces an autoinhibitory sequence from the MLCK catalytic domain. Activated MLCK then phosphorylates a 20 kD light chain of myosin, which initiates actin/myosin cross-bridge cycling and contraction.

Several theoretical models have been proposed for the interaction of Ca^{2+}, CaM, and MLCK. Binding of Ca^{2+} at either end of CaM is highly cooperative[40,218,16], and most models therefore assume two binding sites for CaM, one N-terminal and one C-terminal, each of which can bind two Ca^{2+} ions in a single step. Without further restrictions, these assumptions give rise to a ternary model in which $2 \times Ca^{2+}$ and MLCK may bind to CaM in any order ("model 1," Fig. 5.1B)[26,57]. Several simplified models have also been suggested on the basis of additional biochemical assumptions. Fajmut *et al.* 2005b proposed a six-state model in which Ca^{2+} binding to the C-terminus, but not the N-terminus, obligately precedes binding of MLCK ("model 2," Fig. 5.1C)[58], consistent with the order-of-magnitude difference in affinity for Ca^{2+} between the two domains[218,185,96]. Finally, an ordered-sequential mechanism, in which Ca^{2+} must be bound at both the C-terminus and N-terminus of CaM prior to the binding of MLCK, has also been proposed ("model 3," Fig. 5.1D)[113]. In addition, variants of these models have been incorporated into much larger reaction networks describing agonist-induced Ca^{2+} mobilization and the physiology of airway smooth muscle[130,221,56].

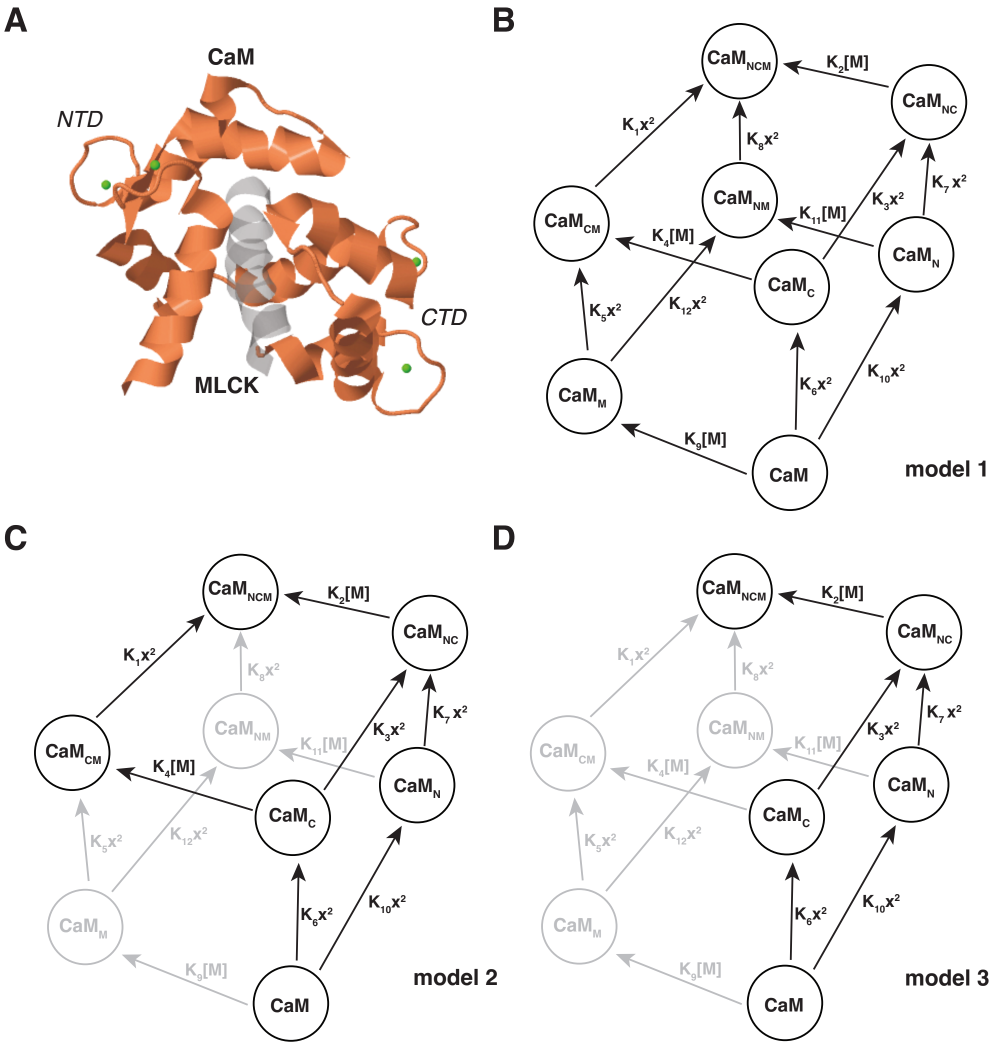


Figure 5.1: Proposed models of the Ca^{2+}- and CaM-dependent activation of MLCK. A Crystal structure at 1.08 Å resolution of an MLCK peptide (gray) bound to CaM (orange). Two Ca^{2+} (green dots) are bound at the N-terminus and two at the C-terminus. Drawn from data reported by Valentine *et al.* 2006[217]. B Equilibrium graph of the eight-state (full) model proposed by Brown *et al.* 1997 and Fajmut *et al.* 2005a. The notation for equilibrium graphs is described in the text. C Equilibrium graph of the six-state model proposed by Fajmut *et al.* 2005b. D Equilibrium graph of the five-state (sequential) model proposed by Kato *et al.* 1984. The graphs in C and D are truncations of the graph in B; the omitted edges and vertices are colored light gray. In all graphs CaM denotes free calmodulin, and M denotes free MLCK. The subscripts N and C indicate that Ca^{2+} is bound to CaM at the N-terminus and C-terminus, respectively, and the subscript M indicates that MLCK is bound. Free $[Ca^{2+}]$ is denoted by x. Two Ca^{2+} ions are assumed to bind CaM simultaneously, following the convention of earlier models.

There is experimental support for the assumptions of both the full and the truncated models. Activation of MLCK is known to require the binding of four Ca^{2+} to CaM, which was the original justification for the ordered-sequential model[41,20]. There is mounting evidence, however, that MLCK can bind apo CaM and CaM in complex with two Ca^{2+}, which suggests that the full model or the six-state model might provide a more biochemically realistic description[26,106,228]. Previous analyses of the models have tended to rely on numerical simulation of the systems of ordinary differential equations that arise from the underlying reactions under the assumption of mass-action kinetics. Such simulations require that numerical values be estimated for all parameters in the model. Furthermore, since the binding and unbinding is assumed to take place within a system that is closed overall, without external sources of matter or energy, the steady state is one of thermodynamic equilibrium, so that paramter values must be constrained to satisfy detailed balance. Fajmut *et al.* 2005a and 2005b used numerical simulation to calculate the expected fraction of bound to total MLCK at steady state for the full and six-state models, with kinetic parameters drawn from *in vitro* biochemical data in the literature. They found that either model could reproduce an experimental binding curve acquired using a FRET-based sensor[57,58,64]. Kato *et al.* 1984 did report a mathematical binding analysis of the simplest (ordered-sequential) model at steady state but did not consider any other models for comparison[113].

We point out here that it is unnecessary to simulate any of these models to determine their steady-state behavior at thermodynamic equilibrium. Steady states can be readily calculated analytically without advance knowledge of the numerical values of any parameters. The advantage of such a mathematical analysis, in contrast to numerical simulation, is that it reveals precisely how the steady state depends on the parameters of the system and on the total amounts of the various components. In turn, this information suggests parameter-independent ways in which the different models can be discriminated.

5.2 RESULTS

5.2.1 EQUILIBRIUM BINDING ANALYSIS OF THREE MODELS OF CA^{2+}- AND CAM-DEPENDENT ACTIVATION OF MLCK

Our mathematical analysis relies on the graph-based linear framework for timescale separation developed in[75,147], which should be consulted for more details; for a review, see[77], and for other applications, see[43,3,55]. Briefly, the framework can be applied to a timescale separation, in which a sub-system is taken to operate sufficiently fast with respect to its environment that it can be assumed to have reached steady state. The steady-state assumption is then used to eliminate the components of the sub-system, in the sense of expressing their steady-state concentrations in terms of the parameters and the concentrations of components in the environment that may be interacting with the sub-system. Here the sub-system is taken to be the various forms of CaM, and it is the steady-state concentrations of these forms that are to be determined in this way.

Biochemical reactions typically give rise to nonlinear dynamics, as is the case for binding reactions. However, this nonlinear dynamics can sometimes be rewritten as a linear dynamics arising from a graph with directed edges and labels on the edges. Such a graph for model 1 resembles that shown in Fig. 5.1B, but its edges and labels have been simplified to reflect thermodynamic equilibrium, as explained below. The vertices of the graph represent the components of the sub-system (the different forms of CaM), while the edges represent the biochemical reactions that inter-convert these components through interaction with components in the environment outside the sub-system. The nonlinearity is incorporated into the edge labels, which may be complex expressions containing both typical parameters, such as rate constants, as well as the free concentrations of the interacting components in the environment. Here these interacting environmental components are Ca^{2+} and MLCK, whose free concentrations are denoted in the graph by x and $[M]$, respectively.

The linear dynamics arising from the graph is common to all applications of the linear framework. In contrast, the nonlinearity in the labels must be dealt with in different ways depending on the specific application. Here a conservation law for the total amount of MLCK, denoted M_{tot}, is used to determine the free concentration of MLCK, while it is assumed as an approximation that the free concentration of Ca^{2+} does not change as a result of binding. In this case, the free concentration of Ca^{2+} is effectively the same as its total concentration. This is a reasonable approximation if the number of free molecules is large compared to the number bound and has been assumed in all previous studies[57,58,113].

Since only the steady state is analyzed here, it is not necessary to write down the linear dynamics that governs the approach to the steady state. Provided that the linear framework graph is strongly connected, so that any two vertices can be joined by a contiguous path of directed edges, the steady state is unique up to a scalar factor. A specific steady state can be calculated from the structure of the graph in terms of the edge labels. The unknown scalar factor is removed by normalizing to the total concentration of material in the graph, which in the present case is the total amount of CaM, denoted CaM_{tot}. The normalization to CaM_{tot} gives rise to the typical algebraic structure for the steady state of each component of the sub-system, as a rational expression in the edge labels.

A second conservation law describes how CaM_{tot} is composed from the various forms of CaM represented by the vertices of the graph. The interplay between the conservation laws for M_{tot} and CaM_{tot} resolves the remaining nonlinearities in the system, as we will see below.

If the steady state is one of thermodynamic equilibrium, as it is for the models considered here, then the steady state is equivalent to what is derived from equilibrium statistical mechanics, with the denominator in the rational expressions being the partition function. (One of the advantages of the linear framework is that it can also be applied away from thermodynamic equilibrium[55].) When considering binding and unbinding reactions at thermodynamic equilibrium, the steady state depends only on the association constants, given by the ratio of the binding on-rate to the unbinding off-rate. Accordingly, the linear framework graphs in this paper have been simplified to show a single edge in the direction of binding, labeled with the ratio of the binding label to the unbinding label.

In more detail, Fig. 6.1B shows the simplified linear framework graph for the model of Fajmut *et al.* 2005a (model 1). CaM is assumed to have two sites for binding of $2 \times$ Ca^{2+} and one site for binding of M, so that there are $2^3 = 8$ vertices in the graph. The association constants for binding of Ca^{2+} and M are denoted $K_1, \cdots, K_{12}$. The graphs for models 2 and 3 are truncations of the graph for model 1, as shown in Fig. 5.1C and D (omitted vertices and edges are colored light gray). Throughout the following analysis, we use square brackets to denote concentration at steady state.

The labeling we have used is particularly convenient for calculating steady states, for which there is a simple rule at thermodynamic equilibrium. For instance, taking the unbound form of CaM as the reference vertex, the steady-state concentration of any vertex can be calculated from $[CaM]$ by choosing any contiguous path of directed edges from CaM to that vertex and multiplying the labels along the path. For instance, taking the path from CaM to CaM$_N$ to CaM$_{NC}$ to CaM$_{NCM}$, the steady-state concentration of the active complex, CaM$_{NCM}$, is given by

$$[CaM_{NCM}] = (K_2 K_7 K_{10} [M] x^4)\,[CaM].$$

There are, of course, many paths from which to choose. It is a consequence of thermodynamic equilibrium that all such paths give the same result, which implies relationships among the equilibrium constants that arise from detailed balance and the cycle condition[75]. For instance, taking instead the path from CaM to CaM$_M$ to CaM$_{CM}$ to CaM$_{NCM}$, we see that detailed balance requires

$$K_2 K_7 K_{10} = K_1 K_5 K_9\,.$$

It is not necessary to impose these relationships for the calculations made below, but they can always be used to reorganize the resulting formulas using association constants found on equivalent paths.

It is now straightforward to write down analytic expressions for two quantities of potential experimental interest, the ratio of the active complex (CaM$_{NCM}$) to M$_{tot}$, which we call A_1, and the ratio of all bound MLCK to M$_{tot}$, which we call F_1. These two quantities are the focus of previous computational analyses[177,57,58]. For the full model, we have the following two conservation laws:

$$\begin{aligned} \mathrm{CaM}_{tot} &= [CaM] + [CaM_M] + [CaM_N] + [CaM_C] \\ &\quad + [CaM_{NM}] + [CaM_{CM}] + [CaM_{NC}] + [CaM_{NCM}] \\ &= [CaM]\,(1 + A(x) + B(x)[M]) \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} \mathrm{M}_{tot} &= [M] + [CaM_M] + [CaM_{NM}] + [CaM_{CM}] + [CaM_{NCM}] \\ &= [M]\,(1 + B(x)[CaM])\,, \end{aligned} \tag{5.2}$$

where

$$\begin{aligned} A(x) &= (K_6 + K_{10})\,x^2 + K_7 K_{10} x^4 \\ B(x) &= K_9 + (K_4 K_6 + K_{10} K_{11})\,x^2 + K_2 K_7 K_{10} x^4. \end{aligned}$$

Thus, after canceling $[M]$, we have

$$A_1 = \frac{[CaM_{NCM}]}{\mathrm{M}_{tot}} = \frac{K_2 K_7 K_{10} x^4 [CaM]}{1 + B(x)[CaM]} \tag{5.3}$$

and

$$F_1 = \frac{\mathrm{M}_{bound}}{\mathrm{M}_{tot}} = \frac{B(x)[CaM]}{1 + B(x)[CaM]}. \tag{5.4}$$

Eqs. 5.3 and 5.4 are expressed in terms of $[CaM]$, which is difficult to measure or manipulate directly in an experiment. We can, however, determine this quantity using the conservation law for total CaM (Eq. 5.1). Eqs. 5.1 and 5.2 are both independently of first order in $[CaM]$ and in $[M]$ (since MLCK binds only once to CaM), despite being highly nonlinear in x. It is therefore straightforward to solve these equations for $[CaM]$ and $[M]$ in terms of x. We find that [CaM] satisfies a quadratic equation,

$$[CaM]^2 + b_1[CaM] + c_1 = 0\,, \tag{5.5}$$

whose coefficients can be expressed in terms of x as

$$\begin{aligned} b_1 &= \frac{1 + A(x) + B(x)\left(\mathrm{M}_{tot} - \mathrm{CaM}_{tot}\right)}{\left(1 + A(x)\right)B(x)} \\ c_1 &= -\frac{\mathrm{CaM}_{tot}}{\left(1 + A(x)\right)B(x)}\,. \end{aligned}$$

A quadratic equation of the form shown in Eq. 5.5 has a single positive real root if, and only if, its constant term is negative. $A(x)$ and $B(x)$ are positive for all positive values of the equilibrium constant and of x, so Eq. 5.5 has one positive real root. Substituting this root into Eq. 5.3 gives, as a function of x,

$$A_1(x) = \frac{2K_2K_7K_{10}\mathrm{CaM}_{tot}x^4}{p_1 + \sqrt{p_2}}\,, \tag{5.6}$$

where

$$\begin{aligned} p_1 &= d_1 + d_2x^2 + d_3x^4 \\ p_2 &= d_4 + d_5x^2 + d_6x^4 + d_7x^6 + d_8x^8\,. \end{aligned}$$

Here $d_1, \cdots, d_8$ are algebraic expressions in the equilibrium constants and the conserved totals.

Throughout the paper, the full expressions for coefficients of this kind are listed in the Methods, with the details of their calculation provided in a Supplementary *Mathematica* notebook.

Similarly, we have

$$F_1(x) = \frac{2B(x)\mathrm{CaM}_{tot}}{p_1 + \sqrt{p_2}}. \tag{5.7}$$

Eqs. 5.6 and 5.7 give A_1 and F_1 as functions of x, in terms of the Ks and the conserved totals (CaM$_{tot}$ and M$_{tot}$). The free concentrations of CaM and MLCK, which are inaccessible to measurement, have been eliminated.

Models 2 and 3 can be analyzed in an identical fashion, so that we have

$$A_2(x) = \frac{2K_2K_7K_{10}\mathrm{CaM}_{tot}x^4}{p_3 + \sqrt{p_4}}$$

and

$$F_2(x) = \frac{2\mathrm{CaM}_{tot}\left(K_4K_6x^2 + K_2K_7K_{10}x^4\right)}{p_3 + \sqrt{p_4}},$$

where p_3 and p_4 have the same algebraic form as p_1 and p_2, respectively. For model 3 there is only a single complex with MLCK bound (CaM$_{NCM}$) and therefore no distinction between A and F. We have

$$A_3(x) = F_3(x) = \frac{2K_2K_7K_{10}\mathrm{CaM}_{tot}x^4}{p_5 + \sqrt{p_6}},$$

where p_5 and p_6 again have the same algebraic form as p_1 and p_2, respectively.

These algebraic expressions for $A_1(x), \cdots, A_3(x)$ and $F_1(x), \cdots, F_3(x)$ can be used to calculate the dependence of MLCK activation and binding on x. Fig. 5.2 shows plots of $A(x)$ and $F(x)$ for the three models, assuming the reference values for $K_1, \cdots, K_{12}$ given in Fajmut *et al.* 2005a

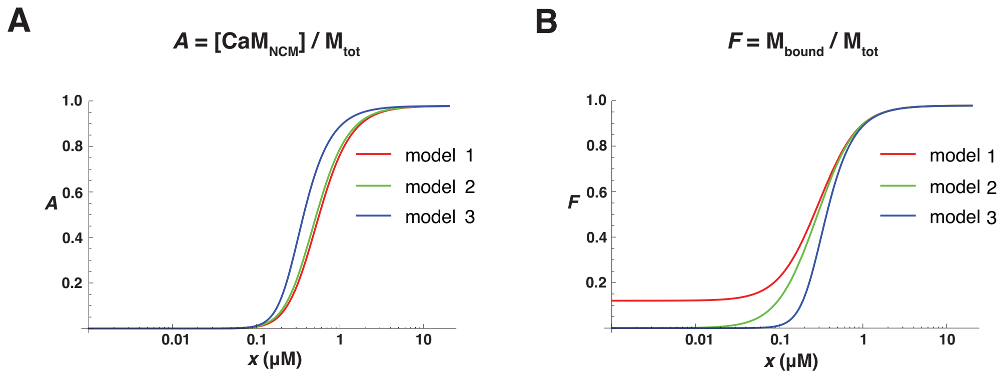


Figure 5.2: Binding curves for the three models. $K_1, \cdots, K_{12}$ were set to the reference values given in Fajmut *et al.* 2005a, and M$_{tot}$ and CaM$_{tot}$ were both set to 2 μM. A Fraction of active MLCK (CaM$_{NCM}$) as a function of x for the three models in Fig. 5.1. B Fraction of bound MLCK as a function of x for the three models in Fig. 5.1.

and setting $\mathrm{M}_{tot} = \mathrm{CaM}_{tot} = 2\ \mu\mathrm{M}$[57].

The curves in Fig. 5.2 are the primary predictions reported in Fajmut *et al.* 2005a and 2005b from numerical simulation of the underlying systems of ordinary differential equations[57,58]. As we have shown here, numerical simulation is not required, however, and the functional form of steady-state quantities like A and F may be calculated with the K parameters and the conserved totals treated symbolically. Parameter values need only be estimated to plot the expressions for A and F calculated above, as in Fig. 5.2.

5.2.2 Parameter-independent criteria for model discrimination

A main goal of our analysis is to identify a minimal set of experimental tests that could distinguish between the three proposed models of MLCK activation, without recourse to numerical fitting of parameter values. For reasons of experimental convenience that will be discussed in greater detail below, our analysis here focuses on the predicted F ratios (i.e., fraction of MLCK bound). We propose a two-step approach that could first differentiate model 1 from models 2 and 3 and then differentiate model 3 from models 1 and 2.

We note that there is a constant term (i.e., a term that is not dependent on x) in the numerator of F_1 but not of F_2 and F_3. Setting $x = 0$ in the relevant expressions, we find that

$$F_1(0) = \frac{2K_9 \mathrm{CaM}_{tot}}{1 + K_9\left(\mathrm{M}_{tot} + \mathrm{CaM}_{tot}\right) + \sqrt{1 + 2K_9\left(\mathrm{CaM}_{tot} + \mathrm{M}_{tot}\right) + K_9^2\left(\mathrm{CaM}_{tot} - \mathrm{M}_{tot}\right)^2}}, \tag{5.8}$$

which is evidently positive, while

$$F_2(0) = F_3(0) = 0.$$

This difference is illustrated in Fig. 5.2B, where models 2 and 3, but not model 1, shows vanishingly low levels of bound MLCK at $x < 10$ nM. Assuming the parameter values and initial conditions used in previous numerical simulations[57], we calculate that, according to model 1, approximately 12% of MLCK would be bound to CaM in 1 nM Ca^{2+}, compared to less than 0.01% according to models 2 or 3. It is also clear from Eq. 5.8 that $F_1(0)$ is an increasing function of CaM$_{tot}$, so that its value can be increased by raising the level of total CaM. For their simulations, Fajmut *et al.* 2005a set M$_{tot}$ $=$ CaM$_{tot}$ $= 2$ μM, but, with M$_{tot}$ $= 2$ μM and CaM$_{tot}$ $= 20$ μM, approximately 59% of MLCK would be bound at $x = 1$ nM according to model 1. Thus, an experimental test of model 1 could involve measurement of MLCK binding in a Ca^{2+}-free buffer, with CaM available in large excess over MLCK.

Furthermore, Eq. 5.8 shows that $F_1(0)$ depends on the single equilibrium constant (K_9), which Fajmut *et al.* 2005a set to 0.078 μM^{-1} after consideration of several experimental measurements[26,130,216]. The experimental values reported, however, vary over several orders of magnitude, from 0.0037 μM^{-1} to 1 μM^{-1}. For any value of K_9 within this range, it is still possible to discriminate model 1 from models 2 and 3 by measuring MLCK binding in low Ca^{2+}; for instance, with $K_9 = 0.0037$

$\mu\mathrm{M}^{-1}$, $\mathrm{M}_{tot} = 2\,\mu\mathrm{M}$, $\mathrm{CaM}_{tot} = 20\,\mu\mathrm{M}$, and $x = 10$ nM, we calculate from model 1 that approximately 7% of MLCK would be bound, compared to negligible levels according to models 2 and 3. Of course, the fraction bound could be increased by having total CaM in even larger excess.

If model 1 were falsified by the above test, it would then be necessary to distinguish between models 2 and 3, which is challenging to do on the basis of the binding of wild-type CaM to MLCK. For a CaM mutant with impaired Ca^{2+} binding at the N-terminus, however, we can give a parameter-independent criterion to differentiate between the corresponding models 2 and 3. Fig. 5.3 shows three additional graphs, corresponding to models 1, 2, and 3, respectively, for a mutant that is unable to bind Ca^{2+} at its N-terminus. The fraction of bound MLCK for model 1 of the mutant, denoted with a superscript asterisk, is given by

$$F_1^*(x) = \frac{2\mathrm{CaM}_{tot}\left(K_9 + K_4K_6x^2\right)}{p_7 + \sqrt{p_8}},$$

where p_7 is a polynomial in x of degree two and p_8 a polynomial in x of degree four, whose coefficients are combinations of the equilibrium constants, M_{tot}, and CaM_{tot}. Similarly, we have

$$F_2^*(x) = \frac{2K_4K_6\mathrm{CaM}_{tot}x^2}{p_9 + \sqrt{p_{10}}},$$

where p_9 and p_{10} have the same algebraic structure as p_7 and p_8, respectively. For model 3, however, there are no longer any reactions involving MLCK binding, so self-evidently

$$F_3^*(x) = 0.$$

For models 1 and 2, MLCK binding is maximized in the high Ca^{2+} limit, for which we have

$$\begin{aligned}\lim_{x\to\infty} F_1^* &= \lim_{x\to\infty} F_2^* \\ &= \frac{2K_4\mathrm{CaM}_{tot}}{1+K_4\left(\mathrm{CaM}_{tot}+\mathrm{M}_{tot}\right)+\sqrt{1+2K_4\left(\mathrm{CaM}_{tot}+\mathrm{M}_{tot}\right)+K_4^2\left(\mathrm{CaM}_{tot}-\mathrm{M}_{tot}\right)^2}}.\end{aligned} \tag{5.9}$$

As such, we calculate, assuming the reference parameter values, that approximately 84% of MLCK would be bound in 1 mM Ca^{2+}, which could be increased to 99% by increasing CaM_{tot} to 20 μM. In the reference parameter set, $K_4 = 16.7\ \mu\mathrm{M}^{-1}$ on the basis of a single experimental measurement[26], and $K_6 = 0.47\ \mu\mathrm{M}^{-1}$, which is on the low end of a range of empirical values considered by Fajmut *et al.* 2005a[57]. Even with both parameters decreased by three orders of magnitude, models 1 and 2 still predict 25% binding, assuming $\mathrm{M}_{tot} = 2\ \mu\mathrm{M}$, $\mathrm{CaM}_{tot} = 20\ \mu\mathrm{M}$, and $x = 1$ mM. Accordingly, the second discrimination test could involve measurement of MLCK binding in a high concentration of Ca^{2+}, with mutant CaM in excess over MLCK.

5.2.3 LIMITING BEHAVIOR OF THE MODELS IN LOW AND HIGH Ca^{2+}

In the previous section, we derived expressions for the fraction of MLCK bound in zero Ca^{2+}(for model 1) and in the high-Ca^{2+}limit (for models 1 and 2 with an N-terminal mutant). Eqs. 5.8 and 5.9 have identical algebraic structure, with the sole difference being the equilibrium constant that appears in the expression. This similarity is not coincidental. In both limits, the reaction network effectively reduces to a bimolecular reaction, between either MLCK and apo CaM (in zero Ca^{2+}) or MLCK and CaM with both Ca^{2+}-binding sites occupied (in high Ca^{2+}). For a general reaction of the form

$$\mathrm{CaM} + \mathrm{M} \overset{K}{\rightleftharpoons} \mathrm{CaM}_M,$$

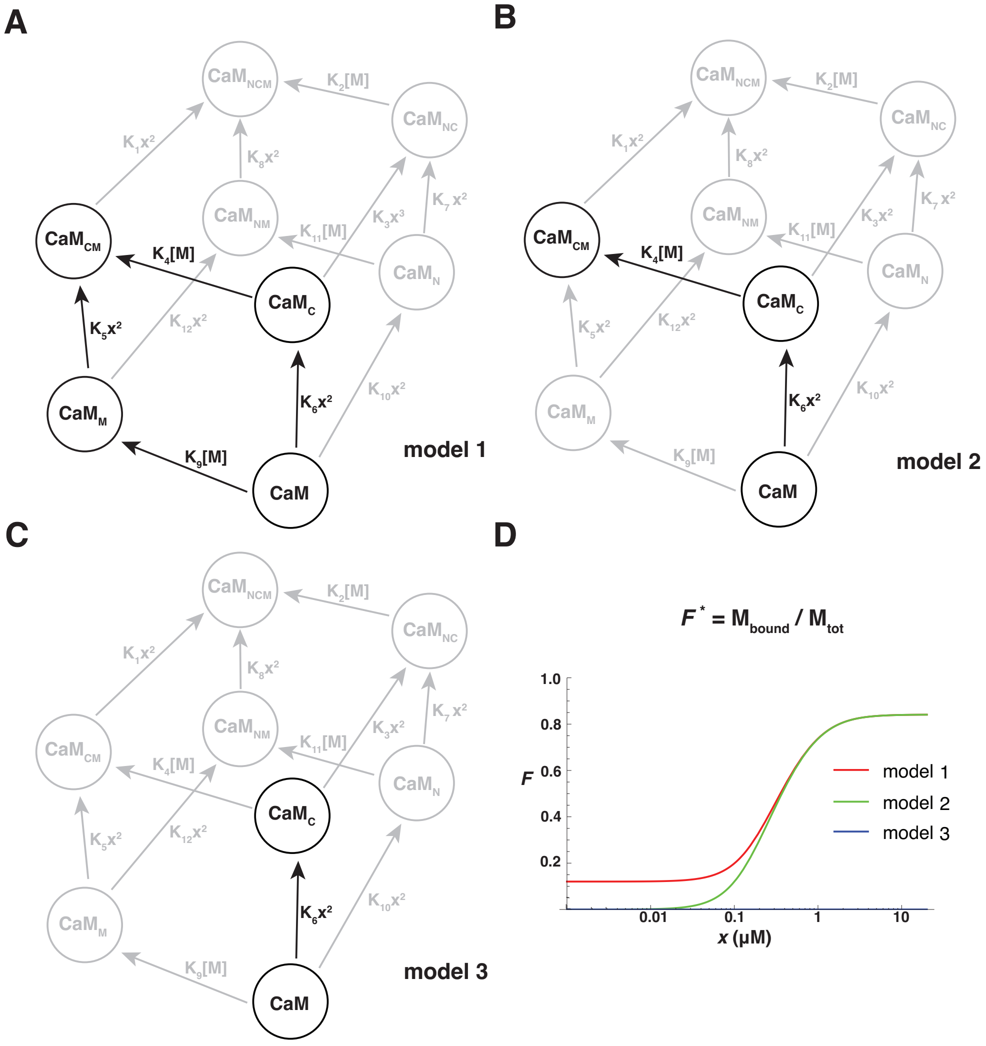


Figure 5.3: Comparison of the three models for CaM mutant with impaired Ca^{2+} binding at the N-terminus. A Equilibrium graph of the eight-state (full) model proposed by Fajmut *et al.* 2005a for mutant. B Equilibrium graph of the six-state model proposed by Fajmut *et al.* 2005b for mutant. C Equilibrium graph of the five-state (sequential) model proposed by Kato *et al.* 1984 for mutant. The notation follows Fig. 5.1. D Fraction of bound MLCK as a function of x for the three models for mutant CaM. The parameter values are the same as for Fig. 5.2.

there is of course only a single complex involving MLCK (CaM_M). Following the usual procedure, we find that

$$F = \frac{[CaM_M]}{\mathrm{M}_{tot}} = \frac{2K\mathrm{CaM}_{tot}}{1 + K\left(\mathrm{CaM}_{tot} + \mathrm{M}_{tot}\right) + \sqrt{1 + 2K\left(\mathrm{CaM}_{tot} + \mathrm{M}_{tot}\right) + K^2\left(\mathrm{CaM}_{tot} - \mathrm{M}_{tot}\right)^2}},$$

which matches the limiting behavior of the more complicated models.

5.3 Discussion

The main predictions of our mathematical analysis are in principle testable using existing experimental technology. Several FRET-based biosensors have been developed for monitoring the binding of Ca^{2+}-CaM to MLCK. These sensors typically consist of two variants of green fluorescent protein tethered by the MLCK binding sequence from CaM, so that binding of Ca^{2+}-CaM disrupts interaction between the fluorescent protein pair and reduces FRET[64,176,34]. Such sensors have enabled the characterization of MLCK localization and binding in a number of cell culture systems, as well as *in vivo* using a transgenic biosensor mouse[34,97,137]. Of particular interest, Geguchadze *et al.* showed that sensor fluorescence tracked phosphorylation of the myosin regulatory light chain in HEK-293T cells, suggesting that their sensor might be useful for quantitative profiling of MLCK activation[64].

Following the analysis presented in the previous section, failure to detect binding of CaM to MLCK in the absence of Ca^{2+} would be sufficient to falsify model 1. Such an experiment could be performed using any of the previously developed biosensors, ideally *in vitro* with purified sensor and CaM (so as to maintain as precise control as possible over the Ca^{2+} concentration in the buffer). As detailed above, assuming the reference parameter values in Fajmut *et al.* 2005a[57], we calculate according to model 1 that approximately 12% of MLCK would be bound to CaM in 1 nM

Ca^{2+} (compared to less than 0.01% for models 2 or 3), a difference which is readily detectable using existing sensors[57,64]. There is already limited experimental evidence against model 1; in lysates of cells stably expressing their biosensor, Geguchadze *et al.* observed no binding at x = 2.5 nM[64].

Discriminating between models 2 and 3 would require binding measurements with a mutant CaM that exhibits impaired Ca^{2+} binding at the N-terminus. The structural and functional properties of CaM have been studied extensively by site-directed mutagenesis, and there are standard strategies for disabling CaM EF-hands without major disruption to the rest of the protein[65,163]. For an appropriate mutant (such as D20A and D56A for human CaM) measurement of non-negligible binding at any reasonably high x would be sufficient to falsify model 3, completing the two-step model discrimination strategy.

Our analysis thus demonstrates that the three models, notwithstanding their numerous similarities, give rise to distinct, experimentally testable predictions about the systems's steady-state behavior. In contrast to previous computational studies of MLCK activation, these predictions are based entirely on algebraic calculations that treat the equilibrium constants symbolically. Inspection of the analytic formulas resulting from these calculations enabled us to identify predictions that are true regardless of the precise numerical values of the parameters, which could not be achieved from numerical simulation alone. In addition to clarifying the relationship between proposed models of CaM/MLCK interaction, variants of our approach should be useful for discriminating between unordered and ordered binding models in diverse physiological contexts.

5.4 METHODS

All conclusions are based on algebraic calculations. A copy of a *Mathematica* notebook (Wolfram Research, Inc., v. 9.0.1.0) containing the main calculations is provided as Supplementary Information. The various coefficients that arise from algebraic analysis of the models are listed below.

Throughout this section we denote CaM$_{tot}$ + M$_{tot}$ as T_1 and CaM$_{tot}$ − M$_{tot}$ as T_2.

For model 1, $p_1 = d_1 + d_2x^2 + d_3x^4$ and $p_2 = d_4 + d_5x^2 + d_6x^4 + d_7x^6 + d_8x^8$, with

$$
\begin{aligned}
d_1 &= 1 + K_9T_1 \\
d_2 &= K_6\left(1 + K_4T_1\right) + K_{10}\left(1 + K_{11}T_1\right) \\
d_3 &= K_7K_{10}\left(1 + K_2T_1\right) \\
d_4 &= 1 + 2K_9T_1 + K_9^2T_2^2 \\
d_5 &= 2\left(K_6\left(1 + K_9T_1 + K_4\left(T_1 + K_9T_2^2\right)\right) + K_{10}\left(1 + K_9T_1 + K_{11}\left(T_1 + K_9T_2^2\right)\right)\right) \\
d_6 &= K_6^2\left(1 + 2K_4T_1 + K_4^2T_2^2\right) + 2K_6K_{10}\left(1 + K_4T_1 + K_{11}\left(T_1 + K_4T_2^2\right)\right) \\
&\quad + K_{10}\left(2K_7\left(1 + K_2T_1 + K_9\left(T_1 + K_2T_2^2\right)\right) + K_{10}\left(1 + 2K_{11}T_1 + K_{11}^2T_2^2\right)\right) \\
d_7 &= 2K_7K_{10}\left(K_6\left(1 + K_4T_1 + K_2\left(T_1 + K_4T_2^2\right)\right) + K_{10}\left(1 + K_2T_1 + K_{11}\left(T_1 + K_2T_2^2\right)\right)\right) \\
d_8 &= K_7^2K_{10}^2\left(1 + 2K_2T_1 + K_2^2T_2^2\right).
\end{aligned}
$$

For model 2, $p_3 = d_9 + d_{10}x^2 + d_{11}x^4$ and $p_2 = d_{12} + d_{13}x^2 + d_{14}x^4 + d_{15}x^6 + d_{16}x^8$, with

$$
\begin{aligned}
d_9 &= 1 \\
d_{10} &= K_6 \left(1 + K_4 T_1\right) + K_{10} \\
d_{11} &= K_7 K_{10} \left(1 + K_2 T_1\right) \\
d_{12} &= 1 \\
d_{13} &= 2 \left(K_6 \left(1 + K_4 T_1\right) + K_{10}\right) \\
d_{14} &= K_6^2 \left(1 + 2K_4 T_1 + K_4^2 T_2^2\right) + 2K_{10} \left(K_6 \left(1 + K_4 T_1\right) + K_7 \left(1 + K_2 T_1\right)\right) + K_{10}^2 \\
d_{15} &= 2K_7 K_{10} \left(K_6 \left(1 + K_4 T_1 + K_2 \left(T_1 + K_4 T_2^2\right)\right) + \left(1 + K_2 T_1\right) K_{10}\right) \\
d_{16} &= K_7^2 K_{10}^2 \left(1 + 2K_2 T_1 + K_2^2 T_2^2\right).
\end{aligned}
$$

For model 3, $p_5 = d_{17} + d_{18} x^2 + d_{19} x^4$ and $p_6 = d_{20} + d_{21} x^2 + d_{22} x^4 + d_{23} x^6 + d_{24} x^8$, with

$$
\begin{aligned}
d_{17} &= 1 \\
d_{18} &= K_6 + K_{10} \\
d_{19} &= K_7 K_{10} \left(1 + K_2 T_1\right) \\
d_{20} &= 1 \\
d_{21} &= 2 \left(K_6 + K_{10}\right) \\
d_{22} &= 2K_{10} \left(K_6 + K_7 \left(1 + K_2 T_1\right)\right) + K_6^2 + K_{10}^2 \\
d_{23} &= 2K_7 K_{10} \left(1 + K_2 T_1\right) \left(K_6 + K_{10}\right) \\
d_{24} &= K_7^2 K_{10}^2 \left(1 + 2K_2 T_1 + K_2^2 T_2^2\right).
\end{aligned}
$$

For model 1 with impaired N-terminal Ca^{2+} binding, $p_7 = d_{25} + d_{26} x^2$ and $p_8 = d_{27} + d_{28} x^2 + d_{29} x^4$, with

$$d_{25} = 1 + K_9 T_1$$

$$d_{26} = K_6 \left(1 + K_4 T_1\right)$$

$$d_{27} = 1 + 2K_9 T_1 + K_9^2 T_2^2$$

$$d_{28} = 2K_6 \left(1 + K_9 T_1 + K_4 \left(T_1 + K_9 T_2^2\right)\right)$$

$$d_{29} = K_6^2 \left(1 + 2K_4 T_1 + K_4^2 T_2^2\right).$$

For model 2 with impaired N-terminal Ca^{2+} binding, $p_9 = d_{30} + d_{31} x^2$ and $p_{10} = d_{32} + d_{33} x^2 + d_{34} x^4$, with

$$d_{30} = 1$$

$$d_{31} = K_6 \left(1 + K_4 T_1\right)$$

$$d_{32} = 1$$

$$d_{33} = 2K_6 \left(1 + K_4 T_1\right)$$

$$d_{34} = K_6^2 \left(1 + 2K_4 T_1 + K_4^2 T_2^2\right).$$

5.5 ACKNOWLEDGEMENTS

We thank Aldebaran Hofer, Yasemin Sancak, and Felix Wong for helpful discussions and comments on the manuscript. J.P.D. was supported by a NSF Graduate Research Fellowship (GE1144152), and J.W.B. and J.G. were supported by NSF grant 0856285.

6

A complex hierarchy of avoidance behaviors in a single-cell eukaryote

AUTHOR LIST

Joseph P. Dexter,* Sudhakaran Prabakaran,* and Jeremy Gunawardena

* J.P.D. and S.P. contributed equally.

CHAPTER SUMMARY

Complex behavior is usually associated with animals having nervous systems, but elementary forms of learning, such as habituation or conditioning, occur in non-neural organisms and even in single cells. In 1906, Herbert Spencer Jennings described, in the sessile ciliate *Stentor roeseli*, a hierarchy of responses to repeated stimulation, which are among the most complex behaviors ever reported for a single cell. Widespread interest in these results has been undermined by claims of non-reproducibility. These claims, however, were based on experiments not with *S. roeseli* but instead with the motile ciliate *Stentor coeruleus*. We acquired and maintained the correct organism in laboratory culture and used micromanipulation and video microscopy to confirm Jennings' observations. *S. roeseli* exhibits avoidance behaviors in a characteristic hierarchy of bending, ciliary alteration, contraction, and detachment, which is distinct from habituation or conditioning. Despite significant individual variation, the data support complex decision making by the organism. Such behavioral complexity may have an evolutionary advantage in protist ecosystems, and the ciliate cortex may have evolved to provide information processing capabilities for implementing such behavior prior to the emergence of multicellularity. Our work resurrects Jennings' biological insights and adds to the list of exceptional features, including regeneration, genome rearrangement, codon reassignment, and cortical inheritance, for which the ciliate clade is renowned. It suggests that learning may be a more widespread phenotype, and may have had greater evolutionary significance, than has emerged from studying metazoans.

6.1 INTRODUCTION

Ciliates form a clade of single-cell eukaryotes with a cosmopolitan distribution in aquatic habitats. They are highly diverse morphologically, varying in size from 10 μm to 4.5 mm, and are characterized by their eponymous cilia used for locomotion and feeding, nuclear dimorphism (they possess both transcriptionally-silent germline micronuclei and an active somatic macronucleus), and conjugation for sex[131]. Their nearest relatives are the apicomplexans and dinoflagellates, with whom they share the sub-cortical vesicles that characterize the alveolates (Fig. 6.1A)[131,52]. Ciliates are thought to have emerged more than two billion years ago, and the earliest known fossils date to the Vendian period[127].

Stentor is a genus of trumpet-shaped heterotrich ciliates, of which several dozen species have been described. Measuring up to several millimeters in length and visible to the naked eye, they are among the largest known unicellular organisms. All species are covered with rows of cilia and have specialized ciliary bundles called "membranellae" at the broad end (Fig. 6.1B), which beat in unison to generate a fluid vortex and draw prey organisms (such as bacteria, algae, and other ciliates) toward the "mouth." The feeding behaviors of *Stentor* vary between species, with some species more sessile and others more motile. Among the best-known sessile species is *Stentor roeseli*, a mid-sized, colorless organism that typically anchors itself to algal detritus through pseudopodia and a holdfast of secreted mucus.

At the beginning of the 20th century, Herbert Spencer Jennings undertook an extensive study of behavior in ciliates and other "lower organisms"[101,102]. Of particular interest to Jennings was the response of *S. roeseli* to repeated exposure to noxious stimuli. Using a fine capillary pipette, he delivered a suspension of carmine dye particles to the mouth of the organism, which would ingest the particles inadvertently while attempting to continue normal feeding behavior. Jennings observed that, in response to this stimulation, *S. roeseli* engaged in series of increasingly elaborate avoidance

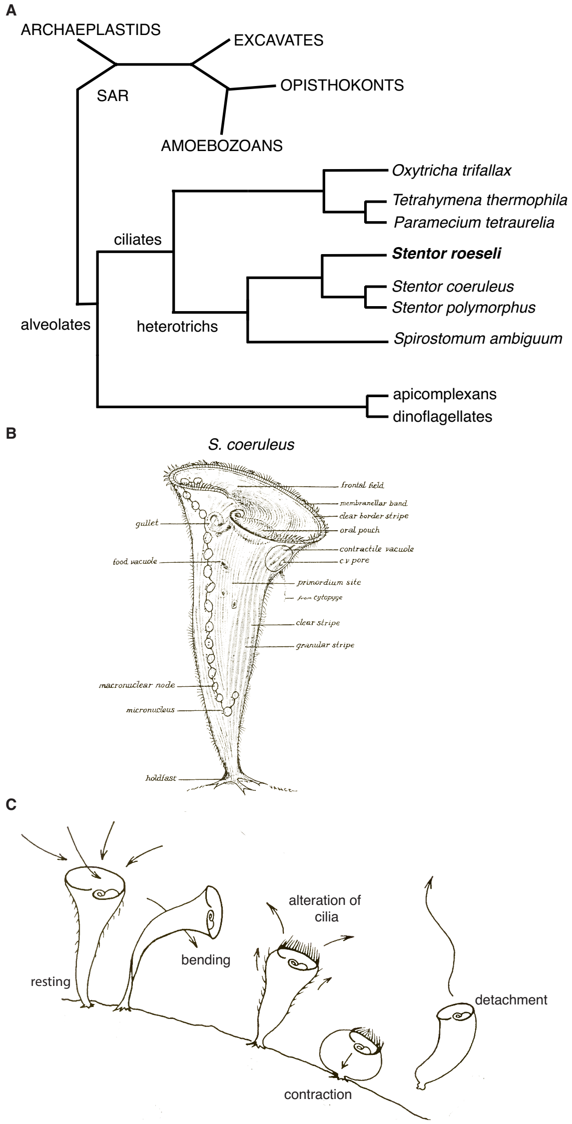


Figure 6.1: Ciliate evolution, morphology, and behavior. A Simplified phylogeny of well-known ciliate species in relation to several major branches of the eukaryotic phylogenetic tree. Adapted from Refs.[131,52,2]. B Anatomical sketch of *S. coeruleus* illustrating the major morphological characteristics of the species. Reprinted from Ref.[215]. C The five *S. roeseli* avoidance behaviors, as sketched by Tatar based on Jennings' descriptions and his own observations. Reprinted from Ref.[215].

behaviors - bending away, changes in the beat of the oral cilia, contraction, and detachment of the holdfast - before ultimately swimming away (Fig. 6.1C). He noted that these behaviors were usually performed in the sequence listed above, with "easy" maneuvers (bending and ciliary reversal) preceding more disruptive and energetically costly ones (contraction, detachment, and swimming).

These results suggest that *S. roeseli* may be capable of hierarchical responses, with the internal state of the organism altered so that it responds differently depending on the history of past stimuli. Such behavior is therefore indicative of learning, memory, and decision making. In light of the organism's active engagement with the environment and the hierarchical organization of the steps involved, the *S. roeseli* avoidance response remains among the most complex behaviors ever described for a unicellular organism. It is important to emphasize, however, that complex behavior in aneural organisms and even single cells is not unprecedented[50]. For instance, habituation to repetitive mechanical, electrical, or chemical stimuli has been demonstrated in other species of *Stentor*, in the ciliate *Spirostomum ambiguum*, and in mammalian PC12 cells[229,230,9,145].

Jennings' observations have enjoyed a colorful history. In their day, they played a important role in debates on the evolutionary origin of behavior and were taken as evidence of functional continuity between unicellular organisms and metazoa with nervous systems[160]. His work has continued to attract interest and fascination to the present day, both within biology and for the general public[179,202]. Yet Jennings' observations have also been dogged by questions of reproducibility, especially among specialists in organismic learning[144]. The primary basis for this skepticism is a 1967 replication attempt by Reynierse and Walsh, which found no evidence for Jennings' hierarchical sequence of avoidance behaviors[172]. Their work, however, suffers from an important limitation that has been overlooked in subsequent discussions of *Stentor* behavior. Reynierse and Walsh were unable to obtain *S. roeseli* and instead used *Stentor coeruleus*, a larger species that is readily obtainable from commercial supply companies. In contrast to the sessile *S. roeseli*, however, *S. coeruleus* strongly prefers to be motile. As such, it is unsurprising that the *S. coeruleus* in Reynierse and Walsh's experiments quickly became free-swimming before completing a sequence of prior avoidance behaviors[172].

Here we report a quantitative analysis of *S. roeseli* avoidance behavior. We obtained and cultured *S. roeseli* and developed a modified version of Jennings' stimulation protocol that permitted video recording of responses in large numbers of organisms. We recorded videos of individual *S. roeseli* engaging in all five behaviors described by Jennings, and we find, across the population of organisms imaged, that contraction obligates precedes detachment. Furthermore, statistical analysis of contraction frequencies suggests that aspects of the behavioral response are non-random, perhaps reflecting complex information processing and internal changes in the organism over the course of stimulation.

6.2 RESULTS AND DISCUSSION

6.2.1 IDENTIFICATION OF *S. ROESELI* AND EXPERIMENTAL SETUP

We obtained *S. roeseli* from a protist supply company and cultured them in pond water as described in the Methods. We confirmed their identity based on observation of a vermiform macronucleus and colorless cortical granules and the absence of symbiotic algae, as specified in the taxonomic classification of heterotrich ciliates (Supplemental Fig. E.S1)[62]. To elicit and record *S. roeseli* avoidance behavior, we constructed a custom microstimulation apparatus consisting of a Signatone micropositioning system with an attached glass microinjection needle placed next to the stage of a Nikon TE-200 inverted microscope equipped with a low magnification, long working distance objective and a CCD camera (Supplemental Fig. E.S2A and Methods). Immediately prior to an experiment, the microinjection needle was loaded with the stimulating agent and connected by tubing to an elevated reservoir of pond water fitted with a two-way stopcock, so that a brief pulse of the agent could be delivered by opening the stopcock. Organisms were transferred to a glass slide without disruption of their holdfasts, and the tip of the needle was positioned in close proximity to the mouth of an individual organism (Supplemental Fig. E.S2B).

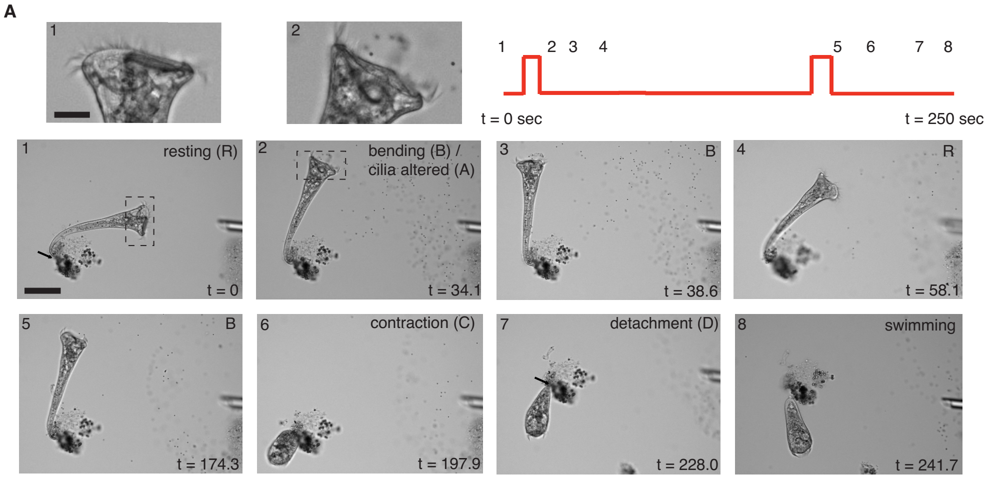
In our initial experiments, we found that stimulation of *S. roeseli* with carmine powder suspended in pond water, as used by Jennings[101,102], rarely elicited any avoidance behavior. Carmine is a natural product derived from the wings of cochineal beetles, and it is possible that its precise composition may have changed in the intervening century. After screening a variety of other particulate suspensions, we found that small polystyrene beads in aqueous suspension with 0.1% NaN_3 reproducibly elicited avoidance behavior (Methods). These beads were used for all subsequent experiments reported here.

6.2.2 MULTI-STEP HIERARCHICAL SEQUENCES OF AVOIDANCE BEHAVIOR

We found that individual *S. roeseli* can engage in all five of Jennings' avoidance behaviors when stimulated with beads. Fig. 6.2A shows a sequence of still images that capture the sequence of behaviors for a sample organism, in the same order described by Jennings. The temporal profile of stimulation, as determined from inspection of the video (Methods), is given in the top right of Fig. 6.2A, and the full video is provided as Supplemental Movie E.S1. These results provide compelling evidence for the fundamental validity of Jennings' verbal descriptions and Tatar's later sketches (Fig. 6.1C).

6.2.3 BEHAVIORAL HETEROGENEITY AND *S. ROESELI* DECISION MAKING

Jennings' original publications contain only qualitative descriptions and drawings of "typical" *S. roeseli* behavior, not quantitative data[101,102]. To investigate the heterogeneity of *Stentor* behavior, we recorded avoidance responses of a large number of organisms on 18 different days. After quality-control review of the videos (Methods), we were left with 57 videos of 70 total organisms. We annotated the avoidance behavior observed for each of these organisms, as summarized in Fig. 6.2B and



B

date	day	video	behavior
11/3/2014	1	1A	RAC1D
11/3/2014	1	1B	RABC1D, RABC1D
11/10/2014	4	4E	RAC2D
11/26/2014	9	9A	RAC12
12/3/2014	10	10C	RC1D
12/10/2014	14	14E	RC9ABD — RBAC2D
12/15/2014	15	15B	RC6BAD
1/22/2015	17	17C	RAC6
1/22/2015	17	17F	RBAC1D
1/30/2015	18	18A	RC3D

C

	(A or B)C	(A or B)D	CD
N	43	45	45
instances	38	29	45
z-score	5.03	1.94	6.71

Figure 6.2: Hierarchical avoidance behavior in *S. roeseli*. A Time-lapse imaging of the response of an individual organism to stimulation with beads. Each of the eight images is labeled with the behavior observed and the time elapsed (in seconds). In the first and penultimate images, the holdfast is indicated with a black arrow. Insets of the first two frames (top) show changes in the the orientation of the membranellae. Scale bars 100 μm (main images) and 50 μm (inset). B Annotated behavior sequences for a small set of representative organisms. The full dataset is provided in Supplemental Tables E.S1 and S2. C Z-scores for three behavioral sequences, compared to expected values for a binomial distribution in which either order is assumed to be equally likely.

Supplemental Tables E.S1 and E.S2. Unsurprisingly, there was substantial heterogeneity in the individual responses - for instance, not all organisms engaged in the full sequence of behaviors, nor were all ultimately induced to detach and swim away.

Despite this heterogeneity, our quantitative analysis of *S. roeseli* behavior supports Jennings' claim of an avoidance hierarchy. Of the 69 organisms that contracted at least once, 45 eventually detached. For those 45 organisms, contraction invariably preceded detachment. Bending or alteration of ciliary beat was observed in 43 organisms that also contracted and preceded the first contraction for 38 of them (88%; Fig. 6.2C).

The design of our microstimulation apparatus and the complexities of recording *Stentor* movement made it infeasible to deliver uniform pulses of beads to all organisms studied. As such, it is possible that the observed heterogeneity reflects a consistent pattern of differential response to variable conditions, rather than true heterogeneity in the organisms' internal state and decision making, which could complicate interpretation of the results (Fig. 6.3A). We reasoned that day-to-day differences in the stimulation protocol (due to disassembly and reassembly of the microstimulation apparatus) would be much greater than any intraday differences. Accordingly, we examined the distribution of contractions prior to detachment across the 18 experimental days and found no significant differences (Fig. 6.3B, $p = 0.116$ by a Kruskal-Wallis H text). In addition, we were occasionally able to stimulate two organisms at the same time (Supplemental Tables E.S1 and E.S2). In these rare cases, we found that the two *S. roeseli*, despite receiving very similar stimuli, tended not to exhibit identical avoidance responses. Supplemental Fig. E.S3 shows a sequence of still images of two such organisms; the full video is provided as Supplemental Movie E.S2.

If the behavioral heterogeneity does indeed reflect differences in the internal state of the organisms, it is instructive to consider the nature of *S. roeseli* "decision making." To investigate whether the decision to contract is a random process, we attempted to fit a Poisson distribution to the predetachment contraction data (Fig. 6.3C). We find that there is a significant difference between the

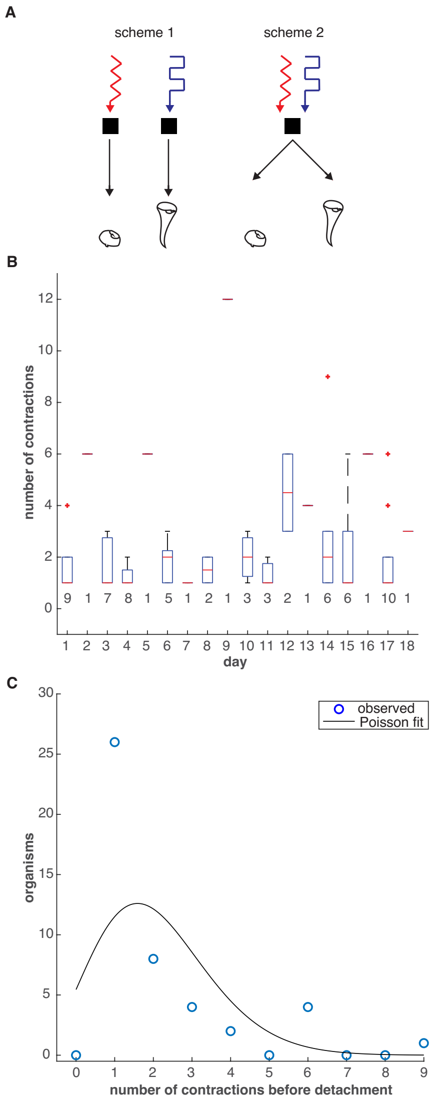


Figure 6.3: Day-to-day heterogeneity in *S. roeseli* behavior. A Cartoons illustrating two possible explanations in the observed heterogeneity. In "scheme 1," organisms respond differently because of fundamental differences in the stimulation; in "scheme 2," organisms make heterogeneous decisions in response to the same kind of stimulus. B Box plot summarizing the number of contractions prior to detachment on each of the 18 experimental days. The red horizontal bar denote the median number of contractions on that day, and the top and bottom of the blue boxes indicate the 75th and 25th percentiles, respectively. Whiskers extend to the furthest non-outlier points, and outliers (defined as >Q3 + 1.5IQR or <Q1 - 1.5IQR) are denoted by red crosses. One extreme outlier (an organism that contracted 17 times) was excluded. $N = 68$ organisms; the N for each day is written under the boxes. C Distribution of contractions prior to detachment is non-random. Plot of the experimentally observed distribution (circles) and of a fitted Poisson distribution with $\lambda = 2.11$ (black line). $N = 45$ organisms.

Poisson best-fit and experimental distributions ($p = 4.49 \times 10^{-6}$ by a χ^2 goodness-of-fit test), suggesting that the decision to contract may be a regulated information processing task for *S. roeseli*.

In the wild, *Stentor* live in complex and dynamic aquatic environments, where prey organisms may be available only intermittently and interspersed with debris[131]. It is plausible to speculate that the *S. roeseli* avoidance response is optimized to achieve a trade-off between easy feeding and self-protection. Much remains to be understood, however, about the relationship between the avoidance response and other *S. roeseli* behaviors. Complex, multi-step mating rituals have been reported for many ciliates, and, although little is known about *S. roeseli* conjugation, *S. coeruleus* does have at least two mating types[28,225,209]. Perhaps putative *S. roeseli* mating rituals share some common elements with the five avoidance behaviors. The study of ciliate behaviors stands to benefit from further observational work, likely coupled with molecular and genomic interrogation[196,197], a topic considered in greater detail in the conclusion of this dissertation.

6.3 METHODS

6.3.1 *S. ROESELI* SOURCE AND MAINTENANCE

Cultures of *S. roeseli* were purchased from Sciento (Manchester, UK), which harvested the organisms from a pond on the property of Whitefield Golf Club (83 Higher Lane, Whitefield, Manchester, UK). The identity of the organisms in the cultures was confirmed by morphological examination (Supplemental Fig. E.S1). We maintained *S. roeseli* in pond water (Carolina Biological Supply Company, Burlington, NC) supplemented with 50 mL/L soil-water supernant (Carolina) in loosely covered glass flasks. Flasks were kept at room temperature in indirect sunlight. Organisms were fed 1 mL of dense cultures of *Chilomonas* and *Chlamydomonas* (Carolina) twice per week, and several wheat seeds (Carolina) were added to the flasks promote bacterial growth. All behavior experiments were performed on organisms purchased no more than two weeks beforehand.

6.3.2 Stimulation of avoidance behavior

The original protocol of Jennings called for stimulation of organisms using an aqueous suspension of carmine particles[102]. Carmine is a natural product extracted from cochineal insects. We were unable to find any information about the source or method of production of Jennings' carmine, and we observed that commercially available carmine powder (Sigma-Aldrich, St. Louis, MO) suspended in pond water elicited avoidance behavior in *S. roeseli* only infrequently. We therefore investigated other particulate suspensions as potential stimuli, including alumina, glass, sand, and polystyrene beads. We found that fluorescent-red, carboxylate-modified polystyrene beads in aqueous suspension with 0.1% NaN_3 (Sigma-Aldrich; mean diameter 2 μm) consistently elicited an avoidance response when microinjected near the mouth of *S. roeseli*. These beads were used for all subsequent behavior experiments.

6.3.3 Needle pulling

Borosilicate glass capillaries with I.D. = 1.10 mm and O.D. = 1.5 mm (Sutter Instrument, Novato, CA) were pulled into microinjection needles using a P-1000 Flaming/Brown micropipette puller (Sutter). The following parameters were used for pulling: Heat 850, Pull 50, Velocity 80, Time 200, Pressure 500. The pulled needle was then broken manually so that the tip diameter was approximately 50% smaller than the mouth of the organism (Supplemental Fig. E.S2B).

6.3.4 Microstimulation apparatus and behavior experiments

We built a custom apparatus to deliver controlled pulses of polystyrene beads directly to the mouth of the organism (Supplemental Fig. E.S2). A Signatone S-931 micropositioner (Gilroy, CA) was placed next to the stage of an inverted microscope on a lab jack. The glass needle was loaded with the suspension of beads and connected to an elevated reservoir of pond water using Tygon tubing

(United States Plastics Corporation, Lima, OH). The needle was then taped to the end of the micropositioner.

Prior to behavior experiments, organisms were removed from culture and placed on a glass slide on the microscope stage without disrupting their attachment to algae. The needle was positioned next to the mouth of the organism manually, and pulses of beads were generated by opening and closing a two-way stopcock connected to the bottom of the reservoir (Bio-Rad Industries, Hercules, CA).

6.3.5 MICROSCOPY

All images were collected on a Nikon TE2000-U inverted microscope (Melville, NY) equipped with a 4x Plan Apo objective lens (N.A. 0.1). An objective with low magnification and long working distance (30 mm) was required because of the substantial movement of the organisms during behavior experiments. Images were acquired with a Hamamatsu ORCA-100 CCD camera (Hamamatsu City, Japan) controlled by MetaMorph 7 software (Molecular Devices, Sunnyvale CA). For timelapse experiments, images were collected at a rate of 7 frames per second, using an exposure time of 5 ms and 1x1 binning, with illumination light shuttered between acquisitions.

6.3.6 ANALYSIS OF VIDEOS

After completion of experimentation, all videos were reviewed manually, and poor-quality videos in which either the organism or bead pulses were not visible were set aside. All of the remaining videos were watched again by at least two of the authors to establish the consensus behavior sequences (Supplemental Tables E.S1 and E.S2) and the approximate pulse times. Only behaviors that occurred after initial stimulation were included in the annotations.

6.4 ACKNOWLEDGEMENTS

We thank Robert McNuff for advice on *Stentor* culture, Joseph Frankel and Denis Lynn for guidance on identification, and Patrick Herring and Marie Herring for help with micromanipulation. All microscopy was performed at the Nikon Imaging Center at Harvard Medical School, and we thank Jennifer Waters and Tally Lambert for assistance with imaging. J.P.D. was supported by a NSF Graduate Research Fellowship (GE1144152).

7

Conclusion

This dissertation considers five related case studies in cellular information processing, covering topics such as robustness in biochemical networks, substrate channeling and metabolic flux in enzyme oligomers, ligand binding, and complex behavior by single-cell organisms, and attempts to demonstrate the usefulness of quantitative analysis for linking form and function. The following conclusion revisits several topics of general interest to the dissertation, with particular attention to review of very recent literature and to follow-up theoretical and experimental work suggested by the research presented in the main chapters.

7.1 LINKING THEORY TO EXPERIMENT

Particular attention is given throughout to the interplay between mathematical modeling and experimental validation. All of the models presented in Chapters 2-5 make clear, experimentally testable predictions. For instance, the critical prediction of our model of the Sln1-Ypd1-Ssk1 phospho-relay is that robust inactivation of the HOG pathway in low-osmolarity conditions requires a large excess of Ypd1. In Chapter 3, we describe multiple experiments that validate this prediction both directly and indirectly. We demonstrate that massive overexpression of Ssk1, which depletes Ypd1 indirectly by increasing levels of intermediate complexes, leads to elevated phosphorylation of Hog1 and impaired growth, and that direct underexpression of Ypd1 also impairs fitness. Subsequent to our work, Stojanovski *et al.* undertook a detailed computational and experimental analysis of Sln1-Ypd1 binding, which also showed that indirect reduction of phospho-Ypd1 leads to low-osmolarity activation of the HOG pathway[206].

For the other models described in this dissertation, one or more key predictions remain to be validated and therefore invite further experimental investigation. In Chapter 4, we present an integrated analysis of 2-hydroxyglutarate (2-HG) production by oncogenic IDH1 mutants. Although the primary focus of the chapter is the use of *in vitro* and *in vivo* kinetic measurements to characterize putative substrate channeling and inter-subunit flux, this analysis motivated the development of a biochemically realistic model of IDH1 heterodimers with potentially broader applications. From this model we derive a polynomial invariant that relates the steady-state concentrations of NADPH, NADP$^+$, α-KG, and isocitrate. As mitochondrial matrix concentrations of NADP$^+$, α-KG, and isocitrate (but not of NADPH) were reported recently in a large-scale characterization of the "MITObolome"[31], we used the invariant to estimate $[\mathrm{NADPH}]_m/[\mathrm{NADPH}]_c$. In combination with the published data, this estimate implies $([\mathrm{NADPH}]_m/[\mathrm{NADP}]_m)/\,([\mathrm{NADPH}]_c/[\mathrm{NADP}]_c)$ < 1, suggesting the intriguing possibility that NADPH pools may be more reduced in the cytosol than in mitochondria. A direct test of this prediction would require measurement of mitochondrial [NADPH], which could become feasible as methods for interrogating metabolic compartmentalization increase in sophistication. Alternatively, it might be possible to test the constraints implied by the invariant *in vitro* using a "systems biochemistry" approach, as described below in greater detail.

Our equilibrium binding analysis of calmodulin (CaM) and myosin light chain kinase (MLCK) suggests a clear experimental strategy for distinguishing between competing models, which is described in detail in Chapter 5. In contrast to some of the other biochemical systems we analyze, research on CaM/MLCK is in a mature phase. The physiology of smooth muscle contraction is in general well-understood, and the quantitative models we consider have been in the literature for over a decade[224,91]. Nevertheless, the usefulness and generalizability of these models has been hampered by repeated attempts to fit them numerically to limited experimental datasets. We show instead that a simple mathematical analysis can inform a two-step model discrimination strategy that does not require estimation of any parameter values. We believe that it is feasible to carry out the strategy using existing technology, and we hope that our analysis will inform future work in this area.

7.2 Systems biochemistry and polynomial invariants

In general, the polynomial invariants that we derive in Chapters 2, 3, and 4 (and in other previous work, such as Ref.[48]) constrain the steady-state concentrations of a small number of species in the appropriate reaction network. For instance, all of the invariants derived in Chapter 2 involve the concentrations of just two species, unmodified and modified substrate. As such, the most direct test of an invariant is to measure those concentrations following incubation with the appropriate enzyme(s) for a length of time sufficient to reach steady state and then check if the algebraic relationship is satisfied. As noted in Chapter 2, checking if the relationship is satisfied raises non-trivial statistical issues, for which various strategies have been proposed and, more recently, applied to real biochemical data on the Wnt/β-catenin signaling pathway[139,81,134]. We refer to this kind of experimental approach as "systems biochemistry," as it involves *in vitro* characterization of systems of multiple components (instead of a single enzyme and single substrate, as is common in traditional biochemistry)[139]. Although systems biochemical approaches have not been widely adopted, bifunctional enzymes may prove to be a convenient test case, for reasons detailed in Chapter 2. Bifunctionality obviates the need to co-incubate multiple enzymes, as would be required for analysis of a standard Goldbeter-Koshland loop[234], and leads to invariants that are always independent of total enzyme and total substrate. This second consideration means that total levels can be set purely for experimental convenience.

7.3 Recent developments in enzyme bifunctionality and robustness

Many experimental demonstrations of robustness in bifunctional enzyme systems date to the 1970s or 1980s[186,122,82]. The recent interest in mathematical modeling of bifunctional enzymes, however, has prompted renewed experimental efforts as well, including detailed characterization of the uridylyltransferase/uridylyl-removing bifunctional enzyme and of the PhoB/PhoR two-component sensor[104,103,63]. As diverse bifunctional enzymes continue to be profiled in greater biochemical detail, the compendium of invariants presented in Chapter 2 should be a useful guide to interpreting their systems-level properties. Gu *et al.*'s recent work on the mammalian bifunctional 5-InsP7 kinase/InsP8 phosphatase is an example of this approach in action[72]. Although they did not try to demonstrate concentration robustness directly, Gu *et al.* characterized the kinetic and mechanistic properties of the enzyme in sufficient detail to conclude, based on existing theoretical results, that the system is robust[72,208].

At the same time, advances continue to be made in theoretical approaches to biological robustness, perhaps motivated in some part by the results described in Chapters 2 and 3. Interest in genome-level analyses of robustness have led to the identification of new structural properties that are necessary for concentration robustness and that hold for arbitrarily complicated networks, and improved software tools for identifying robust motifs[51,119]. On the basis of one such property, Eloundou-Mbebi *et al.* predicted that more than 1,000 human metabolites may be maintained at a robust concentration[51]. Other recent work has focused on integrating biological notions of robustness with related concepts from civil engineering, and on transient, time-dependent forms of concentration robustness[159,53,6].

7.4 ELUCIDATING THE MECHANISMS OF COMPLEX *Stentor* BEHAVIOR

The molecular mechanisms underlying the *Stentor* avoidance behaviors described in Chapter 6 remain unexplored. Recent advances in the molecular biology and genomics of *Stentor*, however, suggest that detailed mechanistic characterization of ciliate behavior may be feasible in the near future. Although the distantly related ciliates *Paramecium* and *Tetrahymena* were developed in the mid-20th century as genetic model organisms, *Stentor* remains intractable for classical genetics , in part because it is challenging to maintain dense cultures and induce frequent mating under laboratory conditions[195]. In an initial effort to circumvent this limitation, Slabodnick *et al.* demonstrated in 2014 that standard RNA interference machinery is conserved in *Stentor coeruleus* and that gene expression can be manipulated by feeding the organism bacteria expressing an appropriate double-stranded RNA, as is routinely done for nematodes and planarians[196]. These technical advances enabled Slabodnick *et al.* to identify the kinase Mob1 as an important regulator of *S. coeruleus* morphogenesis and lay the foundation for further functional genetic analyses of *Stentor*[196]. Additionally, in 2017 the *S. coeruleus* macronuclear genome was sequenced[197], following prior reports of the *P. tetraurelia* macronuclear genome sequence and the macronuclear and micronuclear genomes of *T. thermophile* and *Oxytricha trifallax*[11,49,80,210,32].

These enhanced genomic resources have already enabled large-scale characterization of the *S. coeruleus* kinome, and transcriptomic profiling of injured *Stentor* at different stages of regeneration[170,199,78]. In terms of behavior, a natural first experiment would be to perform RNA sequencing on organisms habituated to touch stimulation[229,230], or, more ambitiously, on groups of *Stentor* that responded differently to our polystyrene bead stimulation (e.g., organisms that contracted once prior to detachment vs. serial contractors).

In addition to the sophisticated information processing and integration needed for hierarchical avoidance, aspects of *Stentor* behavior are impressive biomechanical feats. To break its holdfast and swim away, an organism typically employs a violent "corkscrew" motion, the physics of which remain unexplored. Contraction is known to occur on millisecond timescales[150], perhaps relying on an intricate "microtubule sliding" mechanism [94]. Moreover, as we report in Chapter 6, the number of contractions prior to detachment varies across the population with a non-Poissonian distribution, suggesting that it is in some way a regulated process. Exploring the interplay between signaling and mechanics in *Stentor* behavior is likely to be a fruitful endeavor, one that no doubt will require integration of the classical microsurgical methods pioneered by Tartar and De Terra with molecular and genomic approaches[215,45].

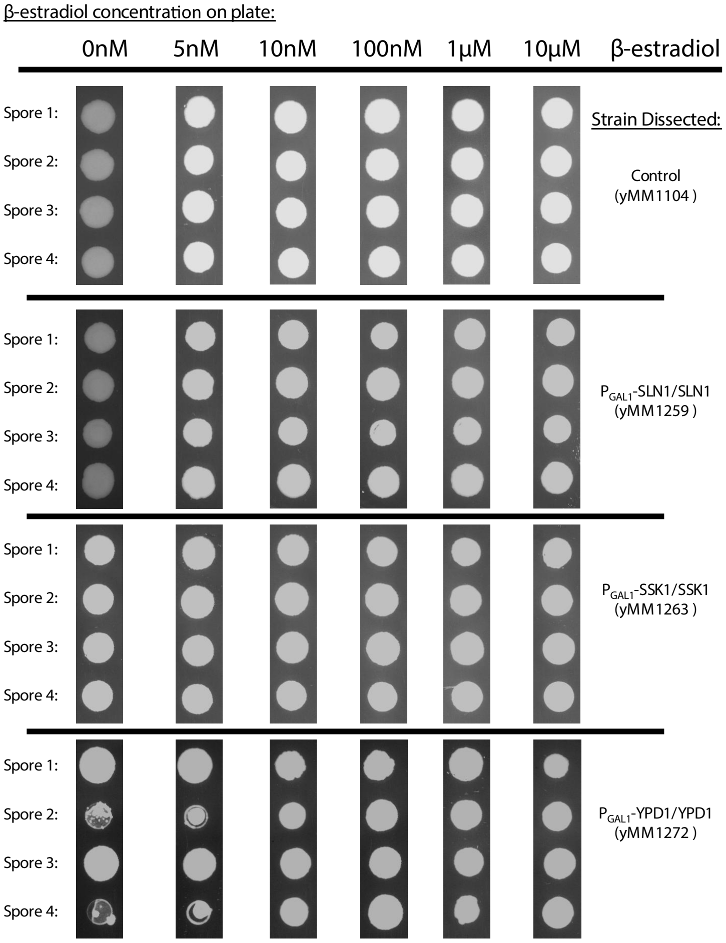
A

Supplemental Material for Chapter 2

All calculations for Chapter 2 are detailed in an accompanying *Mathematica* notebook ("AppendixA_Mathematica_notebook.nb"). Running the notebook requires the additional package "CRN-functions.m."

B

Supplemental Material for Chapter 3



Supplemental Figure B.S1: Underexpression of Ypd1 but not other relay components causes a growth defect. Diploid homozygous GEV strains carrying one inducible allele of a relay component (GENE/PGAL1-GENE) were sporulated onto 10 nM β-estradiol and individual spores were frogged onto plates containing different β-estradiol concentrations. Only underexpression of Ypd1 causes a growth defect.

Strain	Relevant Genotype	Full Genotype	Reference
yMM598	Wild-type GEV Matα	Matα (P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2 HAP1 leu2Δ0::P$_{ACT1}$-GEV-NatMX	McIsaac, et al. 2011
yMM630	Wild-type GEV ura3Δ	Matα (P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX4	McIsaac, et al. 2011
yMM1100	gal10Δ::KanMX	MAT a gal10Δ::KanMX his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	Invitrogen
yMM1101	Wild-type GEV Mata	Mat a (P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2 HAP1 leu2Δ0::P$_{ACT1}$-GEV-NatMX	McIsaac, et al. (DBY12020)
yMM1104	Wild-type GEV/GEV	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2	This study
yMM1259	GEV/GEV P$_{GAL1}$-SLN1/SLN1	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SLN1/KanMX-P$_{GAL1}$-SLN1	This study
yMM1263	GEV/GEV P$_{GAL1}$-SSK1/SSK1	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SSK1/KanMX-P$_{GAL1}$-SSK1	This study
yMM1264	GEV/GEV P$_{GAL1}$-SSK1/SSK1	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 HAP1/HAP1 SSK1/KanMX-P$_{GAL1}$-SSK1	This study
yMM1272	GEV/GEV P$_{GAL1}$-YPD1/YPD1	MAT a/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 YPD1/KanMX-P$_{GAL1}$-YPD1	This study
yMM1277	GEV/GEV P$_{GAL1}$-PBS2/PBS2	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 HAP1/HAP1 PBS2/KanMX-P$_{GAL1}$-PBS2	This study
yMM1286	GEV/GEV P$_{GAL1}$-PBS2/PBS2	MAT a/α HAP1+/HAP1+ (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 PBS2/KanMX-P$_{GAL1}$-PBS2	This study
yMM1287	GEV/GEV P$_{GAL1}$-SSK22/SSK22	MAT a/α HAP1+/HAP1+ (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SSK22/KanMX-P$_{GAL1}$-SSK22	This study
			This study
yMM1296	Wild-type STL1/P$_{STL1}$-yEVenus	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 HAP1/HAP1 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1298	STL1/P$_{STL1}$-yEVenus P$_{GAL1}$-YPD1/YPD1	MATa/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P $_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 YPD1/KanMX-P$_{GAL1}$-YPD1 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1300	STL1/P$_{STL1}$-yEVenus P$_{GAL1}$-SSK1/SSK1	MAT a/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SSK1/KanMX-P$_{GAL1}$-SSK1 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1301	STL1/P$_{STL1}$-yEVenus P$_{GAL1}$-SLN1/SLN1	MAT a/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SLN1/KanMX-P$_{GAL1}$-SLN1 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1304	STL1/P$_{STL1}$-yEVenus P$_{GAL1}$-PBS2/PBS2	MAT a/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 PBS2/KanMX-P$_{GAL1}$-PBS2 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1305	STL1/P$_{STL1}$-yEVenus P$_{GAL1}$-SSK22/SSK22	MAT a/α (P$_{GAL10}$+gal1)Δ:: loxP/ (P$_{GAL10}$+gal1)Δ:: loxP leu2Δ0::P$_{ACT1}$-GEV-NatMX/ leu2Δ0::P$_{ACT1}$-GEV-NatMX gal4Δ:: LEU2/gal4Δ::LEU2 SSK22/KanMX-P$_{GAL1}$-SSK22 STL1/P$_{STL1}$-yEVenus-HphMX	This study
yMM1313	[P$_{GAL1}$-PBS2 scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX [pMM330 P$_{GAL1}$-PBS2 scURA3 2μ]	This study
yMM1314	[P$_{GAL1}$-SSK22 scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX [pMM331 P$_{GAL1}$-SSK22 scURA3 2μ]	This study
yMM1315	[P$_{GAL1}$-SLN1 scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX [pMM332 P$_{GAL1}$-SLN1 scURA3 2μ]	This study
yMM1316	[P$_{GAL1}$-YPD1 scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX4 [pMM333 P$_{GAL1}$-YPD1 scURA3 2μ]	This study
yMM1317	[P$_{GAL1}$-SSK1 scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX4 [pMM334 P$_{GAL1}$-SSK1 scURA3 2μ]	This study
yMM1318	[P$_{GAL1}$ scURA3 2μ]	Matα (P$_{GAL10}$+gal1)Δ:: loxP, gal4Δ::LEU2, HAP1 leu2Δ0:: P$_{ACT1}$-GEV-NatMX ura3Δ:: HphMX4 [pMM329 P$_{GAL1}$ scURA3 2μ]	This study
yMM1333	ssk1Δ	(P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2 HAP1 leu2Δ0::P$_{ACT1}$-GEV-NatMX ura3::HphMX ssk1:: KanMX	This study
yMM1335	ssk1Δ [P$_{GAL1}$ scURA3 2μ]	(P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2 HAP1 leu2Δ0::P$_{ACT1}$-GEV-NatMX ura3::HphMX ssk1:: KanMX [pMM329 P$_{GAL1}$ scURA3 2μ]	This study
yMM1338	ssk1Δ [P$_{GAL1}$-SLN1 scURA3 2μ]	(P$_{GAL10}$+gal1)Δ:: loxP gal4Δ:: LEU2 HAP1 leu2Δ0::P$_{ACT1}$-GEV-NatMX ura3::HphMX ssk1:: KanMX [pMM332 P$_{GAL1}$-SLN1 scURA3 2μ]	This study

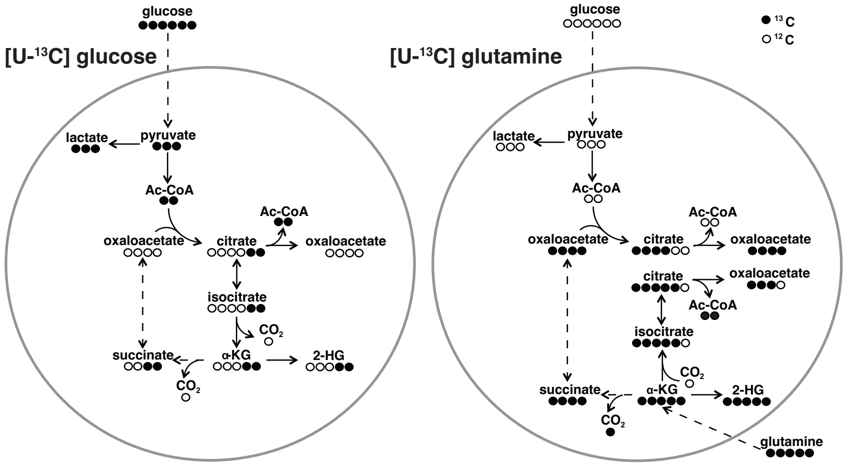
Supplemental Table B.S1: Yeast strains used in this study.

Plasmid ID	Alias	Description	Reference
pMM012	pRS426	scURA3 2µ	Sikorski and Heiter, 1989
pMM131	pFA6-KanMX	KanMX	Goldstein and McCusker, 1999
pMM280	yVenus tag	yEVenus HphMX	This study
pMM329	P_{GAL1} Empty	P_{GAL1} scURA3 2µ	This study
pMM330	P_{GAL1}-PBS2	P_{GAL1}-PBS2 scURA3 2µ	This study
pMM331	P_{GAL1}-SSK22	P_{GAL1}-SSK22 scURA3 2µ	This study
pMM332	P_{GAL1}-SLN1	P_{GAL1}-SLN1 scURA3 2µ	This study
pMM333	P_{GAL1}-YPD1	P_{GAL1}-YPD1 scURA3 2µ	This study
pMM334	P_{GAL1}-SSK1	P_{GAL1}-SSK1 scURA3 2µ	This study

Supplemental Table S2: Plasmids used in this study.

C

Supplemental Material for Chapter 4



Supplemental Figure C.S1: Labeling diagrams for stable isotope tracing experiments. Schematics of expected labeling when cells are grown on glucose uniformly labeled with ^{13}C (U-^{13}C glucose) or on glutamine uniformly labeled with ^{13}C (U-^{13}C glutamine). U-^{13}C glucose is converted to doubly labeled citrate, α-KG, succinate, and 2-HG. U-^{13}C glutamine is converted to fully labeled α-KG, succinate, and 2-HG and to 4- or 5-carbon labeled citrate. Citrate can also be converted to acetyl-CoA and oxaloacetate by ATP citrate lyase. Dashed arrows indicate that one or more intermediate reactions have been omitted from the diagram for clarity.

D

Supplemental Material for Chapter 5

All calculations for Chapter 5 are detailed in an accompanying *Mathematica* notebook ("AppendixD_Mathematica_notebook.nb").

E

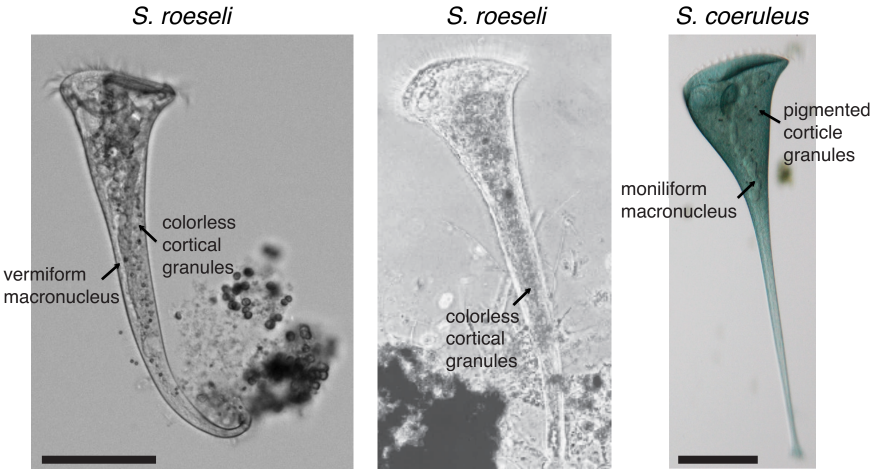
Supplemental Material for Chapter 6

Date	Day	Video	Summary of Behavior
11/3/2014	1	1A	RAC1D
11/3/2014	1	1B	RABC1D, RABC1D
11/3/2014	1	1C	RABC4D
11/3/2014	1	1D	RC1D, RC2D, RC1D
11/3/2014	1	1E	RABC2D
11/3/2014	1	1F	RABC1D
11/5/2014	2	2A	RC6AD
11/7/2014	3	3A	RC1D, RABC1, RABC3
11/7/2014	3	3B	RAC3
11/7/2014	3	3C	RABC1D
11/7/2014	3	3D	RAC2D, RC1D
11/10/2014	4	4A	RC1, RC1, RAC1
11/10/2014	4	4B	RABC1
11/10/2014	4	4C	RAC1
11/10/2014	4	4D	RABC2D
11/10/2014	4	4E	RAC2D
11/10/2014	4	4F	RABC1D
11/12/2014	5	5A	RC6AB
11/14/2014	6	6A	RAC2
11/14/2014	6	6B	RC1D
11/14/2014	6	6C	RC2
11/14/2014	6	6D	RAC1D, RC3
11/15/2014	7	7A	RC1D
11/25/2014	8	8A	RC2
11/25/2014	8	8B	RBAC1
11/26/2014	9	9A	RAC12

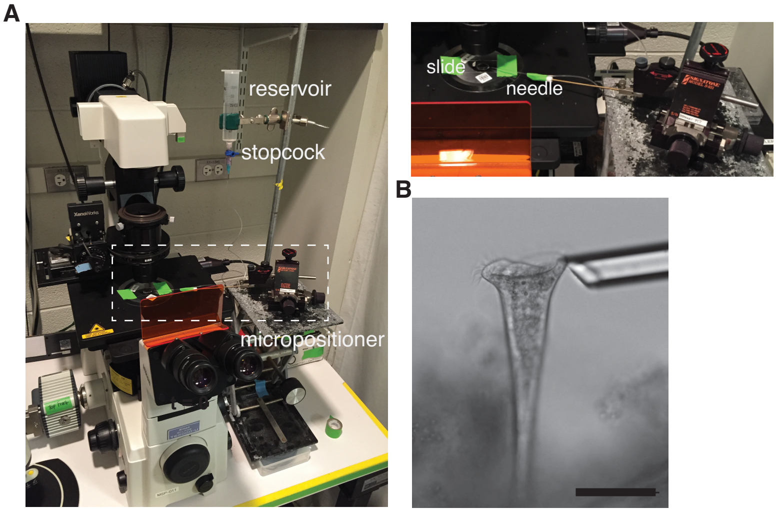
Supplemental Table E.S1: Summary of *S. roeseli* behavior (Part I). For all 70 organisms analyzed, Tables E.S1 and E.S2 list the sequences of avoidance behavior observed (R - resting, B - bending, A - alteration of ciliary beat, C - contraction, D - detachment), along with actual date on which the experiment was performed, the experimental day as defined in Fig. 6.3B, and a code to identify the corresponding raw video, all of which are available from the authors on request. Behaviors are noted once in the order first observed. The number of contractions observed is written immediately after (e.g., "C3" indicates that an organism contracted three times). Only behaviors that were observed subsequent to administration of the first stimulus were included in the annotated sequences. For videos with multiple organisms, the individual sequences are written consecutively, separated by commas. In three unusual cases, an organism detached and subsequently re-anchored while still being imaged. These sequences are noted accordingly in the table, but only behaviors occurring prior to the first detachment were considered for the subsequent statistical analyses. Days 1-9 are summarized in Table E.S1.

Date	Day	Video	Summary of Behavior
12/3/2014	10	10A	RC3
12/3/2014	10	10B	RC2
12/3/2014	10	10C	RC1D
12/5/2014	11	11A	RA
12/5/2014	11	11B	RAC1D
12/5/2014	11	11C	RC2
12/5/2014	11	11D	RC1D
12/8/2014	12	12A	RAC3D
12/8/2014	12	12B	RC6BAD — RC2
12/9/2014	13	13A	RABC4
12/10/2014	14	14A	RAC1, RC1D
12/10/2014	14	14B	RC2D
12/10/2014	14	14C	RAC2D — RC2
12/10/2014	14	14D	RABC3D
12/10/2014	14	14E	RC9ABD — RBAC2D
12/15/2014	15	15A	RABC3D
12/15/2014	15	15B	RC6BAD
12/15/2014	15	15C	RC1D, RC1D, RC1
12/15/2014	15	15D	RC1D
1/21/2015	16	16A	RBAC6D
1/22/2015	17	17A	RC1D, RAC2D
1/22/2015	17	17B	RAC1D
1/22/2015	17	17C	RAC6
1/22/2015	17	17D	RAC1
1/22/2015	17	17E	RABC1D
1/22/2015	17	17F	RBAC1D
1/22/2015	17	17G	RAC1D
1/22/2015	17	17H	RAC1D
1/22/2015	17	17I	RABC4D
1/30/2015	18	18A	RC3D
1/30/2015	18	18B	RAC17

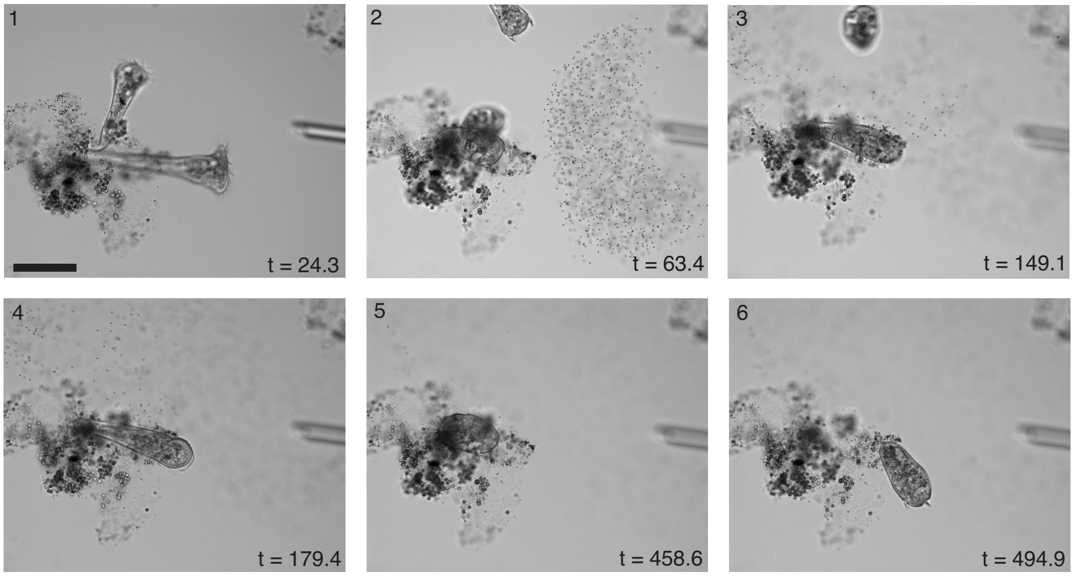
Supplemental Table E.S2: Summary of *S. roeseli* behavior (Part II). Days 10-18 are summarized in Table E.S2.



Supplemental Figure E.S1: *S. roeseli* identification. Brightfield image of a resting and fully extended organism from our culture (left, scale bar 100 μm), along with images of confirmed specimens of *S. roeseli* (middle, from Ref.[62], scale unknown) and *S. coeruleus* (right, from Ref.[195], scale bar 500 μm). Major anatomical features are labeled. *S. roeseli* is distinguished in particular by its vermiform ("worm-like") macronucleus, in contrast to the moniliform ("beads-on-a-string") macronucleus of *S. coerulues*.



Supplemental Figure E.S2: Experimental setup. A Photograph of the custom microstimulation apparatus. B Brightfield image of a resting organism with the glass needle positioned near its oral cavity. Scale bar 100 μm.



Supplemental Figure E.S3: Divergent behavior in response to similar stimuli. Time-lapse imaging of two organisms stimulated simultaneously. One organism detaches much sooner than the other. Scale bar 100 μm.

Two videos of the *S. roeseli* avoidance response are provided as Supplemental Movies E.S1 ("AppendixE_MovieS1.mp4") and E.S2 ("AppendixE_MovieS2.mp4").

F

Simultaneous monitoring of cytosolic and mitochondrial calcium dynamics in single cells

AUTHOR LIST

Joseph P. Dexter,* Yasemin Sancak,* Andrew L. Markhard, Jeremy Gunawardena, and Vamsi K. Mootha

* J.P.D. and Y.S. contributed equally.

Dual-color imaging of cytosolic and mitochondrial calcium

Mitochondrial calcium signaling regulates diverse processes including the coupling of cytosolic and mitochondrial metabolic state[47], cell death[149], and the shaping of cytosolic calcium waves[107,59]. Calcium signaling is a highly dynamic process, with information often encoded in complex oscillatory signals[17]. The recent identification of the molecular components of the mitochondrial calcium uniporter (MCU), which transports calcium from the intermembrane space (IMS) to the matrix, and of the mitochondrial sodium-calcium exchanger (NCLX), which mediates efflux of calcium from the matrix in exchange for sodium, has opened up numerous new opportunities for the study of mitochondrial calcium physiology[110,155].

To enable dynamic characterization of mitochondrial calcium signaling at single-cell resolution, we developed two dual-color reporter systems for simultaneous monitoring of calcium levels in the cytosol and in different parts of the mitochondrion (Fig. F.1). The first system consists of a green genetically encoded calcium indicator (matrix-G-GECO) targeted to the mitochondrial matrix and a red indicator (cyto-R-GECO) expressed in the cytosol[239]. The second systems consists of G-GECO tethered to the cytosolic side of the outer mitochondrial membrane (OMM-G-GECO) and R-GECO targeted to the IMS (IMS-R-GECO).

We used the first system for simultaneous profiling of cytosolic and mitochondrial matrix calcium oscillations. Supplemental Movies F.S1 (cytosol) and F.S2 (matrix) show the response of a population of cyto-R-GECO/matrix-G-GECO HEK-293T cells to 5 μM methacholine, a muscarinic receptor agonist that stimulates IP_3-mediated calcium release[18], and Fig. F.2 shows three typical response dynamics to methacholine stimulation. Additionally, we confirmed that our dual-color system could detect abrogation of mitochondrial calcium uptake following treatment with carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), which dissipates the potential across the inner mitochondrial membrane (IMM). Supplemental Movies F.S3 (untreated) and F.S4 (10 μM CCCP)

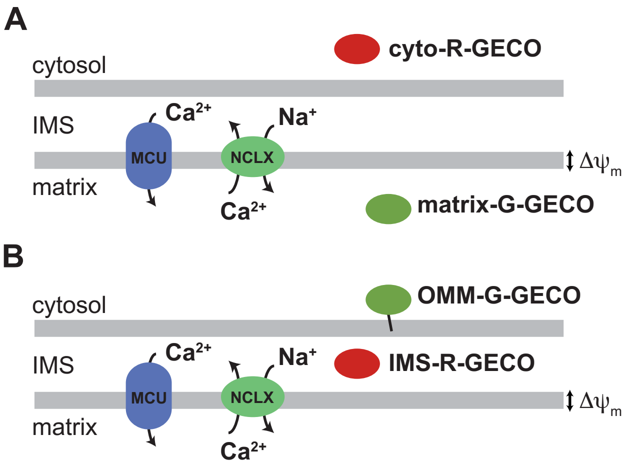


Figure F.1: Schematic of two-color calcium imaging approach. A For monitoring cytosolic and matrix dynamics, a red indicator was expressed in the cytosol (cyto-R-GECO), and a green indicator was targeted to the matrix (matrix-G-GECO). B For monitoring perimitochondrial and IMS dynamics, a red indicator was targeted to the IMS (IMS-R-GECO), and a green indicator was tethered to the outer mitochondrial membrane (OMM-G-GECO).

illustrate representative matrix responses to methacholine captured at 100x magnification.

Few methods for monitoring IMS calcium in live cells have been reported previously. We therefore performed a detailed baseline characterization of our OMM-G-GECO/IMS-R-GECO system. By optical microscopy, both G-GECO and R-GECO were localized to the mitochondrion, as expected (Fig. F.3A). Using a Proteinase K accessibility assay, we confirmed correct localization of R-GECO to the IMS (Fig. F.3B). In addition, we determined the calcium affinity of both reporters and of the inverse pair (i.e., OMM-R-GECO/IMS-G-GECO) in HEK-293T cells (Fig. F.4), which revealed that the *in vivo* affinities are somewhat lower than the *in vitro* values previously reported by Zhao *et al.* ($K_D = 990$ and 1200 nM compared to 618 nM for G-GECO, and $K_D = 580$ and 700 nM compared to 482 nM for R-GECO)[239].

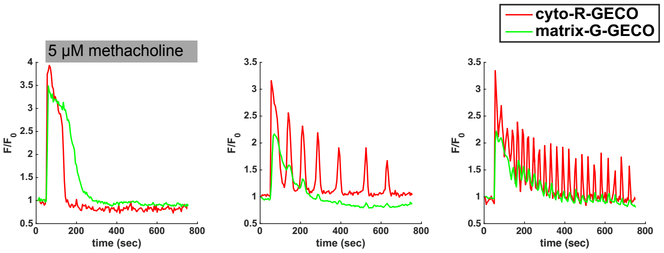


Figure F.2: Representative traces for paired measurement of cytosolic and mitochondrial calcium. Three representative cyto-R-GECO/matrix-G-GECO traces illustrating typical responses of HEK-293T cells to methacholine stimulation (single cytosolic peak, low-frequency oscillation, and high-frequency oscillation).

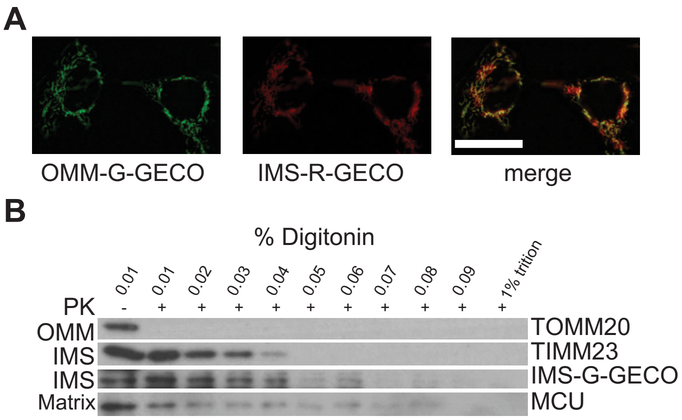


Figure F.3: Localization of OMM-G-GECO/IMS-R-GECO reporters. A HeLa cells expressing OMM-G-GECO (left) and IMS-R-GECO (middle); the overlay image (right) shows colocalization of the reporters. 60x; scale bar 25 µm. B The submitochondrial localization of IMS-GECO reporter was determined using a Proteinase K accessibility assay. The degradation pattern of IMS-GECO is similar to that of IMS-localized TIMM23.

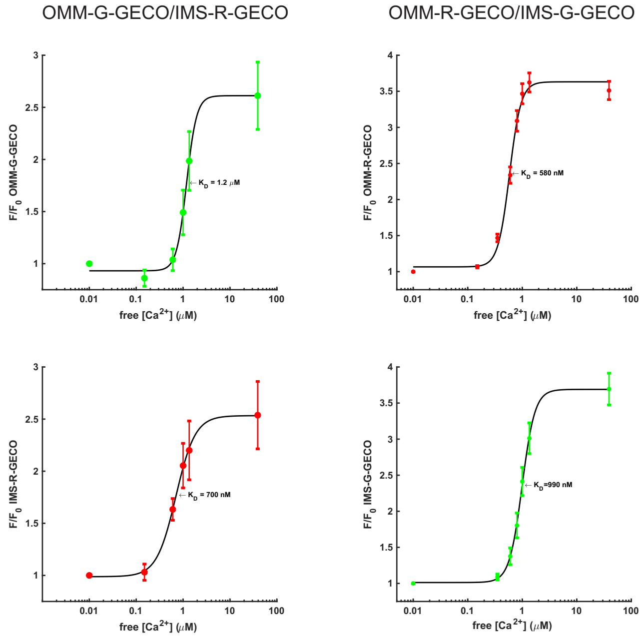


Figure F.4: *In vivo* measurement of reporter calcium affinities. Measurement of calcium-binding affinity (K$_D$) for OMM-G-GECO, IMS-R-GECO, OMM-R-GECO, and IMS-G-GECO in HEK-293T cells permeabilized with digitonin. Points are the mean intensity of $N = 10$ to 50 cells; error bars denote SEM.

NON-EQUIVALENCE OF CALCIUM DYNAMICS ACROSS THE MITOCHONDRIAL OUTER MEMBRANE

The IMS is a small sub-organellar compartment situated between the OMM and IMM. The IMM is composed of the inner boundary membrane, which parallels the OMM, and the cristae membrane, which is folded into extensive invaginations[118]. The distance between the inner boundary membrane has been reported to be approximately 20 nm[161], and the cristae are separated from the peripheral IMS by even narrower junctions[35]. There is an increasing appreciation of the IMS as a functionally distinct compartment that houses numerous processes of importance to mitochondrial and cellular physiology[86], including oxidative folding of proteins[14], export of cytosolic and nuclear Fe-S proteins[120,129], detoxification of reactive oxygen species[86], and regulation of apoptosis[211].

The OMM contains many nonspecific channels (referred to as porins or voltage-dependent anion channels) that are permeable to proteins of mass 5 kD or less[36,19]. As such, a standard assumption has been that the cytosol and IMS are identical in chemical milieu. Multiple direct measurements of IMS composition, however, suggest that this prevailing view is an oversimplification. The pH of the IMS has been reported to be acidic relative to the cytosol in both isolated mitochondria and in intact cells, possibly due to non-uniform distribution of respiratory chain complexes in the IMM and trapping of protons in the cristae[38,166]. Furthermore, targeting of a redox-sensitive YFP variant to the cytosol, IMS, and matrix revealed that the ratio of oxidized to reduced glutathione is higher in the IMS than in the other two compartments[93].

Although the IMS contains multiple calcium-sensing proteins[79], understanding of calcium concentration or dynamics in the compartment remains limited. There is some evidence, in large part indirect, to suggest that the concentration of calcium in the IMS does not simply track that of the cytosol. The permeability to calcium of the OMM-localized voltage-dependent anion channel 1 (VDAC1) varies *in vitro* depending on the confirmation of the channel[212], and loss of VDAC1 has

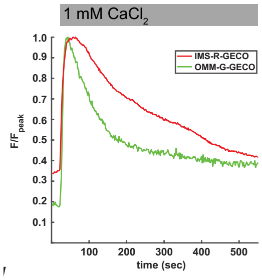


Figure F.5: Representative trace for OMM-G-GECO/IMS-R-GECO. A representative OMM-G-GECO/IMS-R-GECO trace illustrating sustained elevation of IMS calcium above baseline following thapsigargin/CaCl$_2$ treatment of HEK-293T cells.

been reported to lower matrix calcium levels[148]. Similarly, Bcl-2, an outer membrane, anti-apoptotic protein, can potentiate the calcium uptake capacity of the matrix[149]. Finally, direct measurements of IMS calcium suggest amplification of cytosolic calcium spikes in the IMS[174].

In light of these considerations, we undertook a more detailed comparison of cytosolic and IMS calcium dynamics using the OMM-G-GECO/IMS-R-GECO two-color reporter system. For HEK-293T cells treated with 2 μM thapsigargin and 1 mM CaCl$_2$ (to elicit store-operated calcium entry), we observed a striking non-equivalence of the dynamics, with IMS calcium remaining elevated even after decay of the cytosolic transient (Fig. F.5).

To enable quantitative comparison of the dynamics, we parameterized the response curves and examined τ_r (the time from initial stimulation to maximal response) and τ_f (the time from maximal response back to 75% maximal response) (Fig. F.6A). Consistent with our qualitative observations, across the population of cells there was a highly significant difference in both τ_r ($p = 1.36 \times 10^{-11}$ by a Wilcoxon signed-rank test) and τ_f ($p = 3.51 \times 10^{-16}$) between the perimitochondrial cytosol and the IMS (Fig. F.6B and C). This difference was not a trivial consequence of the properties of the two GECO reporters, as we observed a similarly significant difference in cells expressing OMM-R-GECO/IMS-G-GECO instead ($p = 3.63 \times 10^{-23}$ for τ_r and

$p = 3.65 \times 10^{-25}$ for τ_f; $N = 172$ cells).

Nearly all genetically encoded calcium indicators have an intrinsic sensitivity to physiologically relevant changes in intracellular pH; in particular, the pKa of G-GECO is 7.5, and the pKa of R-GECO is 6.59[239]. We therefore examined whether elevation of IMS calcium is accompanied by corresponding changes in pH, which could bias comparisons of perimitochondrial and IMS calcium dynamics. We targeted pHred, a genetically encoded red ratiometric pH sensor with pKa 6.6[213], to the IMS and verified that it responded linearly to changes in pH between 6.5 and 8.0 (Fig. F.7A). Paired measurements of pH (IMS-pHred) and calcium (IMS-G-GECO) demonstrated that calcium transients do not alter IMS pH (Fig. F.7B). This result held for both IP_3-mediated release of calcium from the endoplasmic reticulum stimulated by histamine and for store-operated entry of extracellular calcium.

Methods

Cell culture

HEK-293T and HeLa cells were cultured in DMEM supplemented with 10% Fetal Bovine Serum and Penicillin (100 I.U. / ml), Streptomycin (100 mg/ml), and L-glutamine (0.29 mg/ml). The identities of the cells was confirmed by STR analysis and have been reported before[154]. Cell lines were tested for mycoplasma contamination with MycoAlert Mycoplasma Detection Kit (Lonza, LT07-418) on a monthly basis.

GECO cloning

G-GECO0.9 and R-GECO plasmids were obtained from Addgene (#61247 and # 32444) and were cloned into pLJM1 and pLJM5 lentiviral vectors described before using NheI/EcoRI or NheI/BstsBI sites[183]. For IMS targeting, the first 68 amino acids of Lact8 was used (MYRLLSSVTARAAATAG-

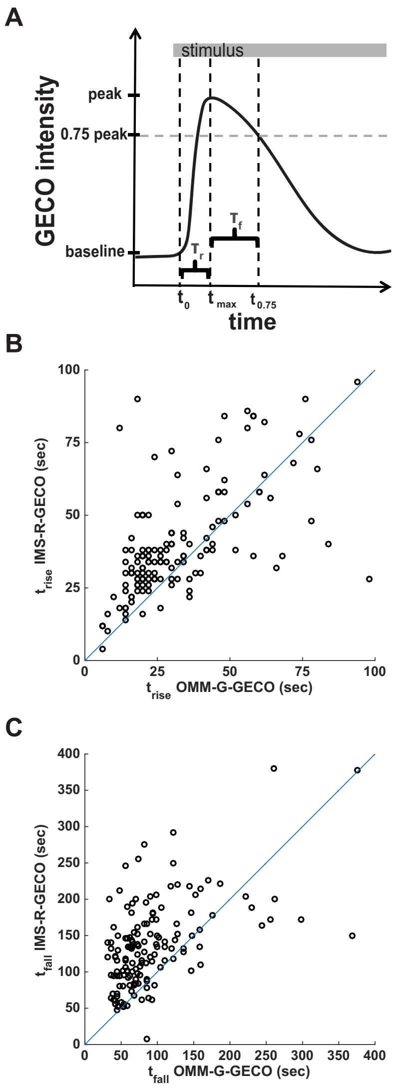


Figure F.6: Calcium dynamics in the IMS are distinct from the perimitochondrial cytosol. A Cartoon response curve showing calculation of key parameters. τ_r is the time from stimulation (t_0) to maximal response (t_{max}), and τ_f is the time to reach 75% of the maximal response from t_{max}. B Comparison of IMS and cytosolic rise kinetics for HEK-293T cells incubated with $CaCl_2$ following thapsigargin treatment ($N = 148$ cells). C Comparison of IMS and cytosolic fall kinetics for the same cells. Each dot denotes one cell. The time resolution for all measurements is 2 sec.

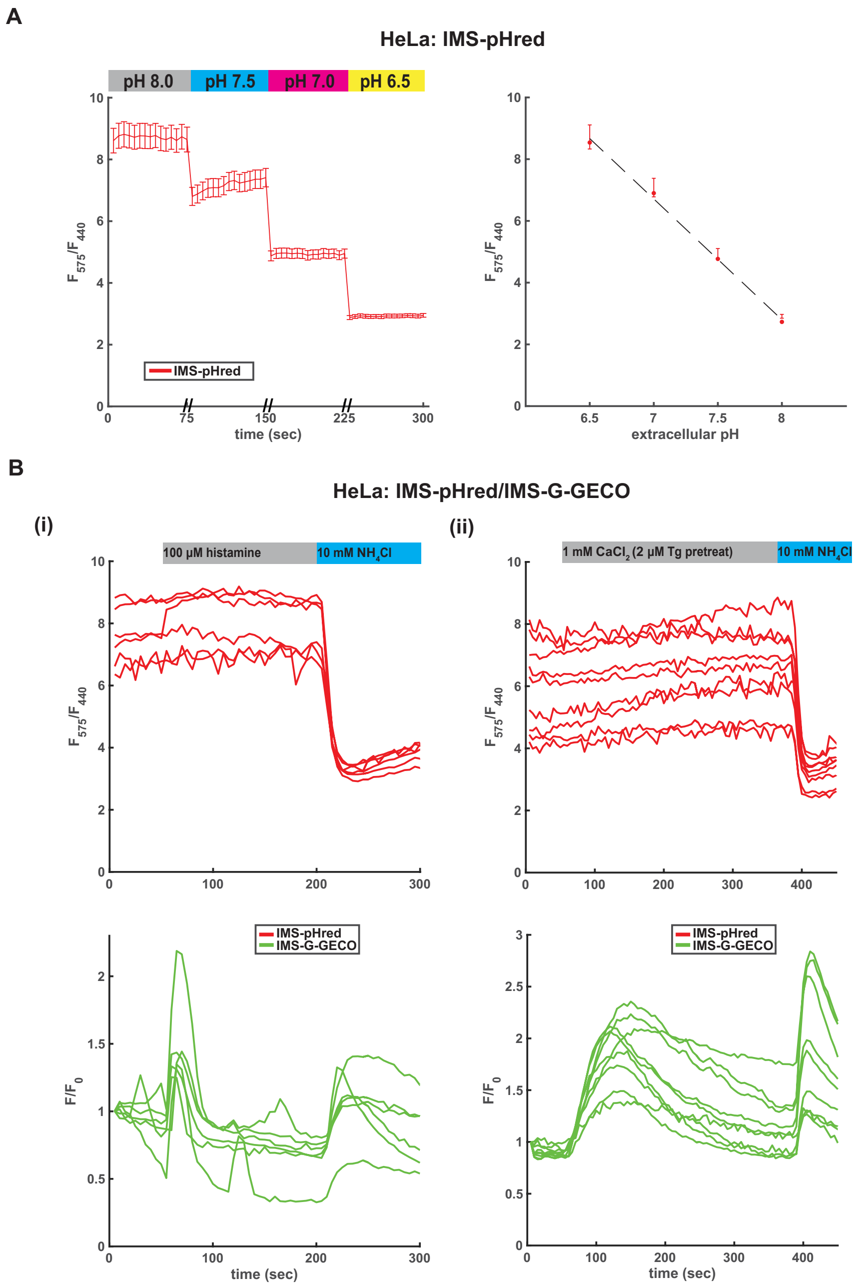


Figure F.7: Calcium transients in the IMS do not alter IMS pH. A IMS-targeted pHred responds to changes in buffer pH in digitonin-permeabilized HeLa cells. Mean of $N = 35$ cells; error bars denote SEM. The response is linear between pH 6.5 and 8 ($R^2 = 0.997$). B Simultaneous recording of IMS-pHred and IMS-G-GECO in intact HeLa cells with (i) histamine stimulation and (ii) CaCl$_2$ incubation following thapsigargin treatment. Final treatment with NH$_4$Cl induced substantial intracellular alkalinization, which was detected by IMS-pHred and also increased the GECO signal. Each line is a single-cell trace. $N = 6$ cells for (i), $N = 9$ cells for (ii).

PAWDGGRRGAHRRPGLPVLGLGWAGGLGLGLGLALGAKLVVGLRGAVPIQS). For OMM targeting, the last 34 amino acids of Bcl-xL was fused to the carboxyl terminus of the GECO sequence (TGESRKGQERFNRWFLTGMTVAGVVLLGSLFSRK).

Proteinase K assay

Proteinase K assay was done as described before[183]. Briefly, HEK-293T cells that stably express IMS-G-GECO were grown to confluency in 15 cm tissue culture dishes. Cells were rinsed once with PBS and were lysed in 1 ml Isolation Buffer IB (10 mM Tris-MOPS, 1 mM EGTA-Tris, 200 mM sucrose, pH 7.4) by passing through a 25g needle 12 times in 1.5 ml microcentrifuge tubes and were centrifuged for 3 min at 800g at 4 C. The supernatant was transferred to another tube, and crude mitochondria were pelleted by centrifugation at 8000g for 3 min at 4 C. The pellet (crude mitochondria) was resuspended in 250 μl of IB. 30 μg of crude mitochondrial prep was incubated with Proteinase K (100 mg/ml of proteinase K) in the presence of digitonin (0-0.9%) or 1% Triton X-100 in 30 μl. The tubes were incubated at room temperature for 20 min, proteinase K was inactivated by addition of PMSF (7 mM), and the samples were boiled in the presence of 1X Laemmli sample buffer. 10 μl of the sample was loaded on an SDS-PAGE for western blotting.

Microscopy

Cells were grown in 35 mm glass bottom dishes with 20 mm No. 1 cover glass and mounted on a Prior Proscan III linear-encoded motorized stage. HBSS without phenol red was used during image acquisition. All images were acquired on a Nikon Ti inverted microscope equipped with a Plan Apo 20x N.A. 0.75 or Plan Apo 100x N.A. 1.4 objective lens and the Perfect Focus Systems to main focus over time. G-GECO fluorescence was excited with a Prior LumenPro light source with an HQ 480/40 nm excitation filter and collected with a 505lp dichroic mirror and ET 535/50 nm emission filter. R-GECO fluorescence was excited with a Prior LumenPro light source with an ET 545/30 nm excitation filter and collected with a 570lp dichroic mirror and 620/60 nm emission filter. Images were acquired with a Hamamatsu ORCA R2 cooled-CCD camera controlled using MetaMorph 7.2 software.

Image analysis

All image analysis was performed using ImageJ 1.49v. To generate the data reported in Fig. F.6, image sequences were first background subtracted using the rolling ball algorithm (radius 50 pixels). To identify individual cells, the OMM-GECO images were superimposed and converted to a binary mask. After despeckling and filling of holes, the mask was segmented using the watershed algorithm to create a template for identification of regions of interest (ROIs). The mean gray-scale value was then measured in each ROI and normalized to baseline concentration. These values (F/F_0) are reported as a proxy for calcium levels in the appropriate intracellular compartment. All ROIs in which the ratio of baseline to maximal intensity was <1.2 (debris or non-responding cells) or in which the ratio of baseline to final intensity was <0.5 or >4 (cells not adhered for the entire experiment) were excluded from further analysis.

References

[1] Abraham, M. & McGowan, J. (1987). The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia*, 23, 243–246.

[2] Adl, S. M., Simpson, A. G., Lane, C. E., Lukes, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Gall, L. L., Lynn, D. H., McManus, H., Mitchell, E. A., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., & Spiegel, F. W. (2012). The revised classification of eukaryotes. *J. Eukaryot. Microbiol.*, 59, 429–493.

[3] Ahsendorf, T., Wong, F., Eils, R., & Gunawardena, J. (2014). A framework for modeling gene regulation which accommodates non-equilibrium mechanisms. *BMC Biol.*, 12, 102.

[4] Alon, U., Surette, M., Barkai, N., & Leibler, S. (1999). Robustness in bacterial chemotaxis. *Nature*, 397, 168–171.

[5] Amary, M., Bacsi, K., Maggiani, F., Damato, S., Halai, D., Berisha, F., Pollock, R., O'Donnell, P., Grigoriadis, A., Diss, T., Eskandarpour, M., Presneau, N., Hogendoorn, P., Futreal, A., Tirabosco, R., & Flanagan, A. (2011). IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J. Pathol.*, 224, 334–343.

[6] Anderson, D., Cappelletti, D., & Kurtz, T. (2017). Finite time distributions of stochastically modeled chemical systems with absolute concentration robustness. *SIAM J. Appl. Dyn. Syst.*, 16, 1309–1339.

[7] Anderson, D., Enciso, G., & Johnston, M. (2014). Stochastic analysis of biochemical reaction networks with absolute concentration robustness. *J. R. Soc. Interface*, 11, 20130943.

[8] Applewhite, P. & Gardner, F. (1973). Tube-escape behavior of paramecia. *Behav. Biol.*, 9, 245–250.

[9] Applewhite, P. & Morowitz, H. (1966). The micrometazoa as model systems for studying the physiology of memory. *Yale J. Biol. Med.*, 39, 90–105.

[10] Astley, H., Parsley, K., Aubry, S., Chastain, C., Burnell, J., Webb, M., & Hibberd, J. (2010). The pyruvate, orthophosphate dikinase regulatory proteins of *Arabidopsis* are both bifunctional and interact with the catalytic and nucleotide-binding domains of pyruvate, orthophosphate dikinase. *Plant J.*, 68, 1070–1080.

[11] Aury, J., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A., Kissmehl, R., Klotz, C., Koll, F., Mouël, A. L., Lepère, G., Malinsky, S., Nowacki, M., Nowak, J., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Bétermier, M., Weissenbach, J., Scarpelli, C., Schächter, V., Sperling, L., Meyer, E., Cohen, J., & Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, 444, 171–178.

[12] Avellaneda Matteo, D., Grunseth, A., Gonzalez, E., Anselmo, S., Kennedy, M., Moman, P., Scott, D., Hoang, A., & Sohl, C. (2017). Molecular mechanisms of isocitrate dehydrogenase 1 (IDH1) mutations identified in tumors: The role of size and hydrophobicity at residue 132 on catalytic efficiency. *J. Biol. Chem.*, 292, 7971–7983.

[13] Babtie, A. & Stumpf, M. (2017). How to deal with parameters for whole-cell modeling. *J. R. Soc. Interface*, 14, 20170237.

[14] Banci, L., Bertini, I., Cefaro, C., Ciofi-Baffoni, S., Gallo, A., Martinelli, M., Sideris, D., Katrakili, N., & Tokatlidis, K. (2009). MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol.*, 16, 198–206.

[15] Batchelor, E. & Goulian, M. (2003). Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proc. Natl. Acad. Sci. USA*, 100, 691–696.

[16] Beccia, M., Sauge-Merle, S., Lemaire, D., Bremond, N., Pardoux, R., Blangy, S., Guilbaud, P., & Berthomieu, C. (2015). Thermodynamics of calcium binding to the calmodulin N-terminal domain to evaluate site-specific affinity constants and cooperativity. *J. Biol. Inorg. Chem.*, 20, 905–919.

[17] Berridge, M. (1997). The AM and FM of calcium signalling. *Nature*, 386, 759–760.

[18] Bird, G. & Putney, J. (2005). Capacitative calcium entry supports calcium oscillations in human embryonic kidney cells. *J. Physiol.*, 562, 697–706.

[19] Blumenthal, A., Kahn, K., Beja, O., Galun, E., Colombini, M., & Breiman, A. (1993). Purification and characterization of the voltage-dependent anion-selective channel protein from wheat mitochondrial membranes. *Plant Physiol.*, 101, 579–587.

[20] Blumenthal, D. & Stull, J. (1980). Activation of skeletal muscle myosin light chain kinase by calcium(2+) and calmodulin. *Biochemistry*, 19, 5608–5614.

[21] Borger, D., Goyal, L., Yau, T., Poon, R., Ancukiewicz, M., Deshpande, V., Christiani, D., Liebman, H., Yang, H., Kim, H., Yen, K., Faris, J., Iafrate, A., Kwak, E., Clark, J., Allen, J., Blaszkowsky, L., Murphy, J., Saha, S., Hong, T., Wo, J., Ferrone, C., Tanabe, K., Bardeesy, N., Straley, K., Agresta, S., Schenkein, D., Ellisen, L., Ryan, D., & Zhu, A. (2014). Circulating oncometabolite 2-hydroxyglutarate is a potential surrogate biomarker in patients with isocitrate dehydrogenase-mutant intrahepatic cholangiocarcinoma. *Clin. Cancer Res.*, 20, 1884–1890.

[22] Borger, D., Tanabe, K., Fan, K., Lopez, H., Fantin, V., Straley, K., Schenkein, D., Hezel, A., Ancukiewicz, M., Liebman, H., Kwak, E., Clark, J., Ryan, D., Deshpande, V., Dias-Santagata, D., Ellisen, L., Zhu, A., & Iafrate, A. (2012). Frequent mutation of isocitrate dehydrogenase (IDH)1 and IDH2 in cholangiocarcinoma identified through broad-based tumor genotyping. *Oncologist*, 17, 72–79.

[23] Borthwick, A., Holms, W., & Nimmo, H. (1984). Isolation of active and inactive forms of isocitrate dehydrogenase from *Escherichia coli* ML308. *Eur. J. Biochem.*, 141, 393–400.

[24] Bralten, L., Kloosterhof, N., Balvers, R., Sacchetti, A., Lapre, L., Lamfers, M., Leenstra, S., de Jonge, H., Kros, J., Jansen, E., Struys, E., Jakobs, C., Salomons, G., Diks, S., Peppelenbosch, M., Kremer, A., Hoogenraad, C., Smitt, P., & French, P. (2011). IDH1 R132H decreases proliferation of glioma cell lines *in vitro* and *in vivo*. *Ann. Neurol.*, 69, 455–463.

[25] Brooks, E., Wu, X., Hanel, A., Nguyen, S., Wang, J., Zhang, J., Harrison, A., & Zhang, W. (2014). Identification and characterization of small-molecule inhibitors of the R132H/R132H mutant isocitrate dehydrogenase 1 homodimer and R132H/wild-type heterodimer. *J. Biomol. Screen.*, 19, 1193–1200.

[26] Brown, S., Martin, S., & Bayley, P. (1997). Kinetic control of the dissociation pathway of calmodulin-peptide complexes. *J. Biol. Chem.*, 272, 3389–3397.

[27] Brucher, B. & Jamal, I. (2014). Cell-cell communication in the tumor microenvironment, carcinogenesis, and anticancer treatment. *Cell Physiol. Biochem.*, 34, 213–243.

[28] Burchill, B. (1967). Conjugation in *Stentor coeruleus*. *J. Protozool.*, 14, 683–687.

[29] Campbell, R., Tour, O., Palmer, A., Steinbach, P., Baird, G., Zacharias, D., & Tsien, R. (2002). A momomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA*, 99, 7877–7882.

[30] Castellana, M., Wilson, M., Xu, Y., Joshi, P., Cristea, I., Rabinowitz, J., Gitai, Z., & Wingreen, N. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.*, 32, 1011–1018.

[31] Chen, W., Freinkman, E., Wang, T., Birsoy, K., & Sabatini, D. (2016). Absolute quantification of matrix metabolites reveals the dynamics of mitochondrial metabolism. *Cell*, 166, 1324–1337.

[32] Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., Perlman, D. H., Doak, T. G., Stuart, A., Amemiya, C. T., Sebra, R. P., & Landweber, L. F. (2014). The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell*, 158, 1187–1198.

[33] Chen, Y., Partow, S., Scalcinati, G., Siewers, V., & Nielsen, J. (2012). Enhancing the copy number of episomal plasmids in *Saccharomyces cerevisiae* for improved protein production. *FEMS Yeast Res*, 12, 598–607.

[34] Chew, T.-L., Wolf, W., Gallagher, P., Matsumura, F., & Chisholm, R. (2002). A fluorescent resonant energy transfer-based biosensor reveals transient and regional myosin light chain kinase activation in lamella and cleavage furrows. *J. Cell Biol.*, 156, 543–553.

[35] Cogliati, S., Enriquez, J., & Scorrano, L. (2016). Mitochondrial cristae: Where beauty meets functionality. *Trends Biochem. Sci.*, 41, 261–273.

[36] Colombini, M. (1979). A candidate for the permeability pathway of the outer mitochondrial membrane. *Nature*, 279, 643–645.

[37] Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics*. London: Portland Press Ltd., 3rd edition.

[38] Cortese, J., Voglino, A., & Hackenbrock, C. (1992). The ionic strength of the intermembrane space of intact mitochondria is not affected by the pH or volume of the intermembrane space. *Biochim. Biophys. Acta*, 1110, 189–197.

[39] Cox, D., Little, J., & O'Shea, D. (2007). *Ideals, Varieties, and Algorithms*. New York: Springer, 3rd edition.

[40] Crouch, T. & Klee, C. (1980). Positive cooperative binding of calcium to bovine brain calmodulin. *Biochemistry*, 19, 3692–3698.

[41] Dabrowska, R., Sherry, J., Aromatorio, D., & Hartshorne, D. (1978). Modulator protein as a component of the myosin light chain kinase from chicken gizzard. *Biochemistry*, 17, 253–258.

[42] Dang, L., White, D., Gross, S., Bennett, B., Bittinger, M., Driggers, E., Fantin, V., Jang, H., Jin, S., Keenan, M., Marks, K., Prins, R., Ward, P., Yen, K., Liau, L., Rabinowitz, J., Cantley, L., Thompson, C., & Su, M. V. H. S. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, 462, 739–744.

[43] Dasgupta, T., Croll, D., Owen, J., Vander Heiden, M., Locasale, J., Alon, U., Cantley, L., & Gunawardena, J. (2014). A fundamental trade-off in covalent switching and its circumvention by enzyme bifunctionality in glucose homeostasis. *J. Biol. Chem.*, 289, 13010–13025.

[44] de la Fuente, M., Young, R., Rubel, J., Rosenblum, M., Tisnado, J., Briggs, S., Arevalo-Perez, J., Cross, J., Campos, C., Straley, K., Zhu, D., Dong, C., Thomas, A., Omuro, A., Nolan, C., Pentsova, E., Kaley, T., Oh, J., Noeske, R., Maher, E., Choi, C., Gutin, P., Holodny, A., Yen, K., DeAngelis, L., Mellinghoff, I., & Thakur, S. (2016). Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma. *Neuro-Oncol.*, 18, 283–290.

[45] De Terra, N. (1985). Cytoskeletal discontinuities in the cell body cortex initiate basal body assembly and oral development in the ciliate *Stentor*. *J. Embryol. Exp. Morph.*, 87, 249–257.

[46] Decker, W., Greuel, G.-M., Pfister, G., & Schönemann, H. (2015). Singular 4-0-2 - a computer algebra system for polynomial computations. http://www.singular.uni-kl.de.

[47] Denton, R. & McCormack, J. (1980). The role of calcium in the regulation of mitochondrial metabolism. *Biochem. Soc. Trans.*, 8, 266–270.

[48] Dexter, J. & Gunawardena, J. (2013). Dimerization and bifunctionality confer robustness to the isocitrate dehydrogenase regulatory system in *Escherichia coli*. *J Biol Chem*, 288, 5770–5778.

[49] Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Jr., R. K. S., Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wilkes, D. E., Wang, Y., Cai, H., Collins, K., Stewart, B. A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C.-C., Gorovsky, M. A., Keeling, P. J., Waller, R. F., Patron, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., del Toro, C., Ryder, H. F., Williamson, S. C., Barbeau, R. A., Hamilton, E. P., & Orias, E. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLOS Biol.*, 4, e286.

[50] Eisenstein, E. (1975). *Aneural Organisms in Neurobiology*. New York: Plenum Press.

[51] Eloundou-Mbebi, J. M., Kuken, A., Omranian, N., Kleessen, S., Neigenfind, J., Basler, G., & Nikoloski, Z. (2016). A network property necessary for concentration robustness. *Nat. Comm.*, 7, 13255.

[52] Embley, T., , & Martin, W. (2006). Eukaryotic evolution, changes and challenges. *Nature*, 440, 623–630.

[53] Enciso, G. (2016). Transient absolute robustness in stochastic biochemical networks. *J. R. Soc. Interface*, 13, 20160475.

[54] Erhart, E. & Hollenberg, C. P. (1981). Curing of *Saccharomyces cerevisiae* 2-μm DNA by transformation. *Curr. Genet.*, 3, 83–89.

[55] Estrada, J., Wong, F., DePace, A., & Gunawardena, J. (2016). Information integration and energy expenditure in gene regulation. *Cell*, 166, 234–244.

[56] Fajmut, A. & Brumen, M. (2008). MLC kinase/phosphatase control of Ca^{2+} signal transduction in airway smooth muscles. *Biophys. J.*, 252, 474–481.

[57] Fajmut, A., Brumen, M., & Schuster, S. (2005a). Theoretical model of the interactions between Ca^{2+}, calmodulin and myosin light chain kinase. *FEBS Lett.*, 579, 4361–4366.

[58] Fajmut, A., Jagodic, M., & Brumen, N. (2005b). Mathematical modeling of the myosin light chain kinase activation. *J. Chem. Inf. Model.*, 45, 1605–1609.

[59] Falcke, M., Hudson, J., Camacho, P., & Lechleiter, J. (1999). Impact of mitochondrial Ca^{2+} cycling on pattern formation and stability. *Biophys. J.*, 77, 37–44.

[60] Fernandez, C., Rosiers, C. D., Previs, S., David, F., & Brunengraber, H. (1996). Correction of ^{13}C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.*, 31, 255–262.

[61] Ferreira, C., van Voorst, F., Martins, A., Neves, L., Oliveira, R., Kielland-Brandt, M., Lucas, C., & Brandt, A. (2005). A member of the sugar transporter family, Stl1p is the glycerol/H+ symporter in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, 16, 2068–2076.

[62] Foissner, W., Berger, H., & Kohmann, F. (1992). *Taxonomische und ökologische Revision der Ciliaten des Saprobiensystems - Band II: Peritrichia, Heterotrichida, Odontostomatida*. Munich: Landesamtes für Wasserwirtschaft.

[63] Gao, R. & Stock, A. (2013). Probing kinase and phosphatase activities of two-component systems *in vivo* with concentration-dependent phosphorylation profiling. *Proc. Natl. Acad. Sci. USA*, 110, 672–677.

[64] Geguchadze, R., Zhi, G., Lau, K., Isotani, E., Persechini, A., Kamm, K., & Stull, J. (2004). Quantitative measurements of Ca^{2+}/calmodulin binding and activation of myosin light chain kinase in cells. *FEBS Lett.*, 557, 121–124.

[65] Geiser, J., van Tuinen, D., Brockerhoff, S., Neff, M., & Davis, T. (1991). Can calmodulin function without binding calcium? *Cell*, 65, 949–959.

[66] Gelman, S. J., Naser, F., Mahieu, N. G., McKenzie, L. D., Dunn, G. P., Chheda, M. G., & Patti, G. J. (2018). Consumption of NADPH for 2-HG synthesis increases pentose phosphate pathway flux and sensitizes cells to oxidative stress. *Cell Rep.*, 22, 512–522.

[67] Gerber, S., Rush, J., Stemman, O., Kirschner, M., & Gygi, S. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA*, 100, 6940–6945.

[68] Ghaemmaghami, S., Huh, W., Bower, K., Howson, R., Belle, A., Dephoure, N., O'Shea, E., & Weissman, J. (2003). Global analysis of protein expression in yeast. *Nature*, 425, 737–741.

[69] Gietz, R. & Schiestl, R. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, 2, 31–34.

[70] Grayson, D. & Stillman, M. (1997). Macaulay2, a software system for research in algebraic geometry. http://www.math.uiuc.edu/Macaulay2.

[71] Gross, S., Cairns, R., Minden, M., Driggers, E., Bittinger, M., Jang, H., Sasaki, M., Jin, S., Schenkein, D., Su, S., Dang, L., Fantin, V., & Mak, T. (2010). Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.*, 207, 339–344.

[72] Gu, C., Nguyen, H.-N., Hofer, A., Jessen, H., Dai, X., Wang, H., & Shears, S. (2017). The significance of the bifunctional kinase/phosphatase activities of diphosphoinositol pentakisphosphate kinases (PPIP5Ks) for coupling inositol pyrophosphate cell signaling to cellular phosphate homeostatis. *J. Biol. Chem.*, 292, 4544–4555.

[73] Gunawardena, J. (2005). Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl. Acad. Sci. USA*, 102, 14617–14622.

[74] Gunawardena, J. (2009). Models in systems biology: the parameter problem and the meanings of robustness. In H. Lodhi & S. Muggleton (Eds.), *Elements of Computational Systems Biology* (pp. 53–76). New York: John Wiley and Sons.

[75] Gunawardena, J. (2012a). A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS ONE*, 5, e36321.

[76] Gunawardena, J. (2012b). Some lessons about models from Michaelis and Menten. *Mol. Biol. Cell*, 23, 517–519.

[77] Gunawardena, J. (2014). Time-scale separation - Michaelis and Menten's old idea, still bearing fruit. *FEBS J.*, 281, 473–488.

[78] Gustafson, H. O., Jamy, M., & Ettema, T. J. G. (2017). RNA sequencing of *Stentor* cell fragments reveals key processes underlying cellular regeneration. *bioRxiv*, https://doi.org/10.1101/232363.

[79] Hajnczky, G., Booth, D., Csordas, G., Debattisti, V., Golenar, T., Naghdi, S., Niknejad, N., Paillard, M., Seifert, E., & Weaver, D. (2014). Reliance of ER-mitochondrial calcium signaling on mitochondrial EF-hand Ca^{2+} binding proteins: Miros, MICUs, LETM1 and solute carriers. *Curr. Opin. Cell Biol.*, 29, 133–141.

[80] Hamilton, E. P., Kapusta, A., Huvos, P. E., Bidwell, S. L., Zafar, N., Tang, H., Hadjithomas, M., Krishnakumar, V., Badger, J. H., Caler, E. V., Russ, C., Zeng, Q., Fan, L., Levin, J. Z., Shea, T., Young, S. K., Hegarty, R., Daza, R., Gujja, S., Wortman, J. R., Birren, B. W., Nusbaum, C., Thomas, J., Carey, C. M., Pritham, E. J., Feschotte, C., Noto, T., Mochizuki, K., Papazyan, R., Taverna, S. D., Dear, P. H., Cassidy-Hanley, D. M., Xiong, J., Miao, W., Orias, E., & Coyne, R. S. (2016). Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *eLife*, 28, e19090.

[81] Harrington, H., Ho, K., Thorne, T., & Stumpf, M. (2012). Parameter-free model discrimination criterion based on steady-state coplanarity. *Proc. Natl. Acad. Sci. USA*, 109, 15746–15751.

[82] Hart, Y. & Alon, U. (2013). The utility of paradoxical components in biological circuits. *Mol. Cell*, 49, 213–221.

[83] Hart, Y., Madar, D., Yuan, J., Bren, A., Mayo, A., Rabinowitz, J., & Alon, U. (2011a). Robust control of nitrogen assimilation by a bifunctional enzyme in *E. coli*. *Mol. Cell*, 41, 117–127.

[84] Hart, Y., Mayo, A., Milo, R., & Alon, U. (2011b). Robust control of PEP formation in the carbon fixation pathway of C_4 plants by a bi-functional enzyme. *BMC Sys. Biol.*, 5, 171.

[85] Hasinoff, B. & Chishti, S. B. (1983). Viscosity dependence of the kinetics of the diffusion-controlled reaction of carbon monoxide with the separated α and β chains of hemoglobin. *Biochemistry*, 22, 58–61.

[86] Hermann, J. & Riemer, J. (2010). The intermembrane space of mitochondria. *Antioxid. Redox Signal.*, 13, 1341–1358.

[87] Hersen, P., McClean, M., Mahadevan, L., & Ramanathan, S. (2008). Signal processing by the HOG MAP kinase pathway. *Proc. Natl. Acad. Sci. USA*, 105, 7165–7170.

[88] Hohmann, S. (2002). Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.*, 66, 300–372.

[89] Hohmann, S. (2009). Control of high osmolarity signalling in the yeast *Saccharomyes cerevisaie*. *FEBS Lett.*, (pp. 4025–4029).

[90] Hohmann, S., Krantz, M., & Nordlander, B. (2007). Yeast osmoregulation. *Meth. Enzymol.*, 428, 29–45.

[91] Hong, F., Haldeman, B., Jackson, D., Carter, M., Baker, J., & Cremo, C. (2011). Biochemistry of smooth muscle myosin light chain kinase. *Arch. Biochem. Biophys.*, 510, 135–146.

[92] Horie, T., Tatebayashi, K., Yamada, R., & Saito, H. (2008). Phosphorylated Ssk1 prevents unphosphorylated Ssk1 from activating the Ssk2 mitogen-activated protein kinase kinase kinase in the yeast high-osmolarity glycerol osmoregulatory pathway. *Mol. Cell. Biol.*, 28, 5172–5183.

[93] Hu, J., Dong, L., & Caryn, E. (2008). The redox environment in the mitochondrial intermembrane space is maintained separately from the cytosol and matrix. *J. Biol. Chem.*, 283, 29126–29134.

[94] Huang, B. & Pitelka, D. (1973). The contractile process in the ciliate, *Stentor coeruleus*. *J. Cell Biol.*, 57, 704–728.

[95] Hurley, J., Thorsness, P., Ramalingam, V., Helmers, N., Koshland, D., & Stroud, R. (1989). Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. USA*, 86, 8635–8639.

[96] Ikura, M., Hiraoki, T., Hikichi, K., Mikuni, T., Yazawa, M., & Yagi, K. (1983). Nuclear magnetic resonance studies on calmodulin: calcium-induced conformational change. *Biochemistry*, 22, 2573–2579.

[97] Isotani, E., Zhi, G., Lau, K., Huang, J., Mizuno, Y., Persechini, A., Geguchadze, R., Kamm, K., & Stull, J. (2004). Real-time evaluation of myosin light chain kinase activation in smooth muscle tissues from a transgenic calmodulin-biosensor mouse. *Proc. Natl. Acad. Sci. USA*, 101, 6279–6284.

[98] Janiak-Spens, F., Cook, P., & West, A. (2005). Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. *Biochemistry*, 44, 377–386.

[99] Janiak-Spens, F., Sparling, D., & West, A. (2000). Novel role for an HPt domain in stabilizing the phosphorylated state of a response regulator domain. *J Bacteriol.*, 182, 6673–6678.

[100] Janiak-Spens, F., Sparling, J., Gurfinkel, M., & West, A. (1999). Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. *J Bacteriol.*, 181, 411–417.

[101] Jennings, H. (1902). Studies on the reactions to stimuli in unicellular organisms IX - on the behavior of fixed infusoria (Stentor and Vorticella) with special reference to the modifiability of protozoan reactions. *Am. J. Physiol.*, 8, 23–60.

[102] Jennings, H. (1906). *Behavior of the Lower Organisms*. New York: Columbia University Press.

[103] Jiang, P., Ventura, A., & Ninfa, A. (2012a). Characterization of the reconstituted UTase/UR-PII-NRII-NRI bicyclic signal transduction system that controls the transcription of nitrogen-regulated (Ntr) genes in *Escherichia coli*. *Biochemistry*, 51, 9045–9057.

[104] Jiang, P., Zhang, Y., Atkinson, M., & Ninfa, A. (2012b). The robustness of the *Escherichia coli* signal-transducing UTase/UR-PII covalent modification cycle to variation in the PII concentration requires very strong inhibition of the UTase activity of UTase/UR by glutamine. *Biochemistry*, 51, 9032–9044.

[105] Jin, G., Reitman, Z., Duncan, C., Spasojevic, I., Gooden, D., Rasheed, B., Yang, R., Lopez, G., He, Y., McLendon, R., Bigner, D., & Yan, H. (2013). Disruption of wild-type IDH1 suppresses D-2-hydroxyglutarate production in IDH1-mutated gliomas. *Cancer Res.*, 73, 496–501.

[106] Johnson, J., Snyder, C., Walsh, M., & Flynn, M. (1996). Effects of myosin light chain kinase and peptides on Ca^{2+} exchange with the N- and C-terminal Ca^{2+} binding sites of calmodulin. *J. Biol. Chem.*, 271, 761–767.

[107] Jouaville, L., Ichas, F., Holmuhamedov, E., Camacho, P., & Lechleiter, J. (1995). Synchronization of calcium waves by mitochondrial substrates in *xenopus laevis* oocytes. *Nature*, 377, 438–441.

[108] Kafri, R., Levy, M., & Pilpel, Y. (2006). The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl. Acad. Sci. USA*, 103, 11653–11658.

[109] Kahn, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J., & Kschischo, M. (2010). Grofit: Fitting biological growth curves with R. *J. Stat. Software*, 33, 1–21.

[110] Kamer, K. & Mootha, V. (2015). The molecular era of the mitochondrial calcium uniporter. *Nat. Rev. Mol. Cell Biol.*, 16, 545–553.

[111] Karim, A. S., Curran, K. A., & Alper, H. S. (2013). Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast. Res.*, 13, 107–116.

[112] Karp, R., Perez-Millan, M., Dasgupta, T., Dickenstein, A., & Gunawardena, J. (2012). Complex-linear invariants of biochemical reaction networks. *J. Theor. Biol.*, 311, 130–138.

[113] Kato, S., Osa, T., & Ogasawara, T. (1984). Kinetic model for isometric contraction in smooth muscle on the basis of myosin phosphorylation hypothesis. *Biophys. J.*, 46, 35–44.

[114] Kitano, H. (2007). Towards a theory of biological robustness. *Mol. Syst. Biol.*, 3, 137.

[115] Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., & Hohmann, S. (2005). Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.*, 23, 975–982.

[116] Krakauer, D. & Plotkin, J. (2002). Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. USA*, 99, 1405–1409.

[117] Krantz, M., Ahmadpour, D., Ottoson, L.-G., Warringer, J., Waltermann, C., Nordlander, B., Klipp, E., Blomberg, A., Hohmann, S., & Kitano, H. (2009). Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway. *Mol. Sys. Biol.*, 5, 1–7.

[118] Kuhlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. *BMC Biol.*, 13, 89.

[119] Kuwahara, H., Umarov, R., Almasri, I., & Gao, X. (2017). ACRE: Absolute concentration robustness exploration in module-based combinatorial networks. *Synth. Biol.*, 2, ysx001.

[120] Lange, H., Lisowsky, T., Gerber, J., Muhlenhoff, U., Kispal, G., & Lill, R. (2001). An essential function of the mitochondrial sulfhydryl oxidase Erv1p/ALR in the maturation of cytosolic Fe/S proteins. *EMBO Rep.*, 2, 715–720.

[121] LaPorte, D. & Koshland, D. (1982). A protein with kinase and phosphatase involved in regulation of tricarboxylic acid cycle. *Nature*, 300, 458–460.

[122] LaPorte, D., Thorsness, P., & Koshland, D. (1985). Compensatory phosphorylation of isocitrate dehydrogenase. A mechanism for adaptation to the intracellular environment. *J. Biol. Chem.*, 260, 10563–10568.

[123] Larhlimi, A., Blachon, S., Selbig, J., & Nikoloski, Z. (2011). Robustness of metabolic networks: A review of existing definitions. *BioSystems*, 106, 1–8.

[124] Leonardi, R., Subramanian, C., Jackowski, S., & Rock, C. (2012). Cancer-associated isocitrate dehydrogenase mutations inactivate NADPH-dependent reductive carboxylation. *J. Biol. Chem.*, 287, 14615–14620.

[125] Lewis, C., Parker, S., Fiske, B., McCloskey, D., Gui, D., Green, C., Vokes, N., Feist, A., Vander Heiden, M., & Metallo, C. (2014). Tracing compartmentalized NADPH metabolism in the cytosol and mitochondria of mammalian cells. *Mol. Cell*, 55, 253–263.

[126] Li, C., Wang, E., & Wang, J. (2012). Landscape topography determines global stability and robustness of a metabolic network. *ACS Synth. Biol.*, 1, 229–239.

[127] Li, C.-W., Chen, J.-Y., Lipps, J., Gao, F., Chi, H.-M., & Wu, H.-J. (2007). Ciliated protozoans from the precambrian Doushantuo formation, Wengan, South China. In P. Vickers-Rich & P. Komarower (Eds.), *The Rise and Fall of the Ediacaran Biota* (pp. 151–156).: The Geological Society of London.

[128] Li, S., Chou, A., Chen, W., Chen, R., Deng, Y., Phillips, H., Selfridge, J., Zurayk, M., Lou, J., Everson, R., Wu, K., Faull, K., Cloughesy, T., Liau, L., & Lai, A. (2013). Overexpression of isocitrate dehydrogenase mutant proteins renders glioma cells more sensitive to radiation. *Neuro-Oncol.*, 15, 57–68.

[129] Lill, R. (2009). Function and biogenesis of iron-sulphur proteins. *Nature*, 460, 831–838.

[130] Lukas, T. (2004). A signal transduction pathway model prototype I: From agonist to cellular endpoint. *Biophys. J.*, 87, 1406–1416.

[131] Lynn, D. H. (2008). *The Ciliated Protozoa. Characterization, Classification and Guide to the Literature*. New York: Springer.

[132] Ma, H., Kunes, S., Schatz, P., & Botstein, D. (1987). Plasmid construction by homologous recombination in yeast. *Gene*, 58, 201–216.

[133] Macia, J., Regot, S., Peeters, T., Conde, N., Sole, R., & Posas, F. (2009). Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Science Signaling*, 2, 1–9.

[134] MacLean, A., Rosen, Z., Byrne, H., & Harrington, H. (2015). Parameter-free methods distinguish Wnt pathway models and guide design of experiments. *Proc. Natl. Acad. Sci. USA*, 112, 2652–2657.

[135] Maeda, T., Takekawa, M., & Saito, H. (1995). Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. *Science*, 269, 554–558.

[136] Maeda, T., Wurgler-Murphy, S., & Saito, H. (1994). A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature*, 369, 242–245.

[137] Maier, L., Ziolo, M., Bossuyt, J., Persechini, A., Mestril, R., & Bers, D. (2006). Dynamic changes in free Ca-calmodulin levels in adult cardiac myocytes. *J. Mol. Cell. Cardiol.*, 41, 451–458.

[138] Mann, M., Ong, S.-E., Gronberg, M., Steen, H., Jensen, O., & Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.*, 20, 261–268.

[139] Manrai, A. & Gunawardena, J. (2008). The geometry of multisite phosphorylation. *Biophys. J.*, 95, 5533–5543.

[140] Mardis, E., Ding, L., Dooling, D., Larson, D., McLellan, M., Chen, K., Koboldt, D., Fulton, R., Delehaunty, K., McGrath, S., Fulton, L., Locke, D., Magrini, V., Abbott, R., Vickery, T., Reed, J., Robinson, J., Wylie, T., Smith, S., Carmichael, L., Eldred, J., Harris, C., Walker, J., Peck, J., Du, F., Dukes, A., Sanderson, G., Brummett, A., Clark, E., McMichael, J., Meyer, R., Schindler, J., Pohl, C., Wallis, J., Shi, X., Lin, L., Schmidt, H., Tang, Y., Haipek, C., Wiechert, M., Ivy, J., Kalicki, J., Elliott, G., Ries, R., Payton, J., Westervelt, P., Tomasson, M., Watson, M., Baty, J., Heath, S., Shannon, W., Nagarajan, R., Link, D., Walter, M., Graubert, T., DiPersio, J., Wilson, R., & Ley, T. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.*, 361, 1058–1066.

[141] Masel, J. & Siegal, M. (2009). Robustness: mechanisms and consequences. *Trends Genet.*, 25, 395–403.

[142] Mavrianos, J., Desai, C., & Chauhan, N. (2014). Two-component histidine phosphotransfer protein Ypd1 is not essential for viability in *Candida albicans*. *Eukaryot. Cell*, 13, 452–460.

[143] Mayya, V., Rezual, K., Wu, L., Fong, M., & Han, D. (2006). Absolute quantification of multisite phosphorylation by selective reaction monitoring mass spectrometry. *Mol. Cell. Proteomics*, 5, 1146–1157.

[144] McConnell, J. (1966). Comparative physiology: learning in invertebrates. *Ann. Rev. Physiol.*, 28, 107–136.

[145] McFadden, P. & Koshland, D. (1990). Habituation in the single cell: diminished secretion of norepinephrine with repetitive depolarization of PC12 cells. *Proc. Natl. Acad. Sci. USA*, 87, 2031–2035.

[146] McIsaac, R., Silverman, S., McClean, M., Gibney, P., Macinskas, J., Hickman, M., Petti, A., & Botstein, D. (2011). Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, 22, 4447–4459.

[147] Mirzaev, I. & Gunawardena, J. (2013). Laplacian dynamics on general graphs. *Bull. Math. Biol.*, 75, 2118–2149.

[148] Monaco, G., Decrock, E., Arbel, N., van Vliet, A., Rovere, R. L., Smedt, H. D., Parys, J., Agostinis, P., Leybaert, L., Shoshan-Barmatz, V., & Bultynck, G. (2015). The BH4 of anti-apoptotic Bcl-XL, but not that of the related Bcl-2, limits the voltage-dependent anion channel 1 (VDAC1) – mediated transfer of pro-apoptotic Ca^{2+} signals to mitochondria. *J. Biol. Chem.*, 290, 9150–9161.

[149] Murphy, A., Bredesen, D., Cortopassi, G., Wang, E., & Fiskum, G. (1996). Bcl-2 potentiates the maximal calcium uptake capacity of neural cell mitochondria. *Proc. Natl. Acad. Sci. USA*, 93, 9893–9898.

[150] Newman, E. (1972). Contraction in *Stentor coeruleus*: A cinematic analysis. *Science*, 177, 447–449.

[151] Nimmo, G., Borthwick, A., Holms, W., & Nimmo, H. (1984). Partial purification and properties of isocitrate dehydrogenase kinase/phosphatase from *Escherichia coli*. *Eur. J. Biochem.*, 141, 401–408.

[152] Ohta, S., Ishida, Y., & Usami, S. (2006). High-level expression of cold-tolerant pyruvate, orthophosphate dikinase from a genomic clone with site-directed mutations in transgenic maize. *Mol. Breeding*, 18, 29–38.

[153] Oldenburg, K., Vo, K., Michaelis, S., & Paddon, C. (1997). Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res.*, 25, 451–452.

[154] Oxenoid, K., Dong, Y., Cao, C., Cui, T., Y.Sancak, Markhard, A., Grabarek, Z., Kong, L., Liu, Z., Ouyang, B., Cong, Y., Mootha, V., & Chou, J. (2016). Architecture of the mitochondrial calcium uniporter. *Nature*, 533, 269–273.

[155] Palty, R., Silverman, W., Hershfinkel, M., Caporale, T., Seni, S., Parnis, J., Nolte, C., Fishman, D., Shosham-Barmatz, V., Hermann, S., Khananshvili, D., & Sekler, I. (2010). NCLX is an essential component of mitochondrial Na^{+}/Ca^{2+}. *Proc. Natl. Acad. Sci. USA*, 107, 436–441.

[156] Paluch, E. (2015). After the greeting: realizing the potential of physical models in cell biology. *Trends Cell. Biol.*, 25, 711–713.

[157] Pansuriya, T., van Eijk, R., d'Adamo, P., van Ruler, M., Kuijjer, M., Oosting, J., Cleton-Jansen, A., van Oosterwijk, J., Verbeke, S., Meijer, D., van Wezel, T., Nord, K., Sangiorgi, L., Toker, B., Liegl-Atzwanger, B., San-Julian, M., Sciot, R., Limaye, N., Kindblom, L., Daugaard, S., Godfraind, C., Boon, L., Vikkula, M., Kurek, K., Szuhai, K., French, P., & Bovée, J. (2011). Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. *Nat. Genet.*, 43, 1256–1261.

[158] Parsons, D., Jones, S., Zhang, X., Lin, J., Leary, R., Angenendt, P., Mankoo, P., Carter, H., Siu, I., Gallia, G., Olivi, A., McLendon, R., Rasheed, B., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D., Tekleab, H., Diaz, L., Hartigan, J., Smith, D., Strausberg, R., Marie, S., Shinjo, S., Yan, H., Riggins, G., Bigner, D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V., & Kinzler, K. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321, 1807–1812.

[159] Paul, D. & Radde, N. (2016). Robustness and filtering properties of ubiquitous signaling network motifs. *IFAC-PapersOnLine*, 49, 120–127.

[160] Pauly, P. (1981). The Loeb-Jennings debate and the science of animal behavior. *J. Hist. Behav. Sci.*, 17, 504–515.

[161] Perkins, G., Renken, C., Martone, M., Young, S., & Ellisman, M. (1997). Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. *J. Struct. Biol.*, 119, 260–272.

[162] Phair, R. (2014). Mechanistic modeling confronts the complexity of molecular cell biology. *Mol. Biol. Cell*, 25, 3494–3496.

[163] Piazza, M., Taiakina, V., Dieckmann, T., & Guillemette, J. (2017). Structural consequences of calmodulin EF hand mutations. *Biochemistry*, 56, 944–956.

[164] Pichler, M., Bodner, C., Fischer, C., Deutsch, A., Hiden, K., Beham-Schmid, C., Linkesch, W., Guelly, C., Sill, H., & Wölfler, A. (2011). Evaluation of mutations in the isocitrate dehydrogenase genes in therapy-related and secondary acute myeloid leukaemia identifies a patient with clonal evolution to IDH2 R172K homozygosity due to uniparental disomy. *Br. J. Haematol.*, 152, 669–672.

[165] Pietrak, B., Zhao, H., Qi, H., Quinn, C., Gao, E., Boyer, J., Concha, N., Brown, K., Duraiswami, C., Wooster, R., Sweitzer, S., & Schwartz, B. (2011). A tale of two subunits: How the neomorphic R132H IDH1 mutation enhances production of αHG. *Biochemistry*, 50, 4804–4812.

[166] Porcelli, A., Ghelli, A., Zanna, C., Pinton, P., Rizzuto, R., & Rugolo, M. (2005). pH difference across the outer mitochondrial membrane measured with a green fluorescent protein mutant. *Biochem. Biophys. Res. Comm.*, 326, 799–804.

[167] Posas, F. & Saito, H. (1998). Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. *EMBO J.*, 17, 1385–1394.

[168] Prabakaran, S., Everley, R., Landrieu, I., Wieruszeski, J.-M., Lippens, G., Steen, H., & Gunawardena, J. (2011). Comparative analysis of Erk phosphorylation suggests a mixed strategy for measuring phospho-form distributions. *Mol. Sys. Biol.*, 7, 5533–5543.

[169] Prabakaran, S., Lippens, G., Steen, H., & Gunawardena, J. (2012). Post-translational modification: Nature's escape from from genetic imprisonment and the basis for cellular information processing. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 4, 565–83.

[170] Reiff, S. & Marshall, W. (2017). A large kinome in a large cell: *Stentor coeruleus* possesses highly expanded kinase families and novel domain architectures. *bioRxiv*, https://dx.doi.org/10.1101/168187.

[171] Rendina, A., Pietrak, B., Smallwood, A., Zhao, H., Qi, H., Quinn, C., Adams, N., Concha, N., Duraiswami, C., Thrall, S., Sweitzer, S., & Schwartz, B. (2013). Mutant IDH1 enhances the production of 2-hydroxyglutarate due to its kinetic mechanism. *Biochemistry*, 52, 4563–4577.

[172] Reynierse, J. & Walsh, G. (1967). Behavior modification in the protozoan *Stentor* reexamined. *Pyschol. Rec.*, 17, 161–165.

[173] Rizk, A., Gatt, G., Fages, F., & Soliman, S. (2009). A general computational method for robustness analysis with applications to synthetic gene networks. *Bioinformatics*, 25, i169–i178.

[174] Rizzuto, R., Pinton, P., Carrington, W., Fay, F., Fogarty, K., Lifshitz, L., Tuft, R., & Pozzan, T. (1998). Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca^{2+} responses. *Science*, 280, 1763–1766.

[175] Robinson, G. L., Philip, B., Guthrie, M. R., Cox, J. E., Robinson, J. P., VanBrocklin, M. W., & Holmen, S. L. (2016). *In vitro* visualization and characterization of wild type and mutant IDH homo- and heterodimers using bimolecular fluorescence complementation. *Cancer Res. Front.*, 2, 311–329.

[176] Romoser, V., Hinkle, P., & Persechini, A. (1997). Detection in living cells of Ca^{2+}-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *J. Biol. Chem.*, 272, 13270–13274.

[177] Roux, E., Mbikou, P., & Fajmut, A. (2012). Role of protein kinase network in excitation-contraction coupling in smooth muscle cell. In G. D. S. Xavier (Ed.), *Protein Kinases* (pp. 287–320). Rijeka, Croatia: InTechOpen.

[178] Russo, F. & Silhavy, T. (1993). The essential tension: opposed reactions in bacterial two-component regulatory systems. *Trends Microbiol.*, 1, 306–310.

[179] Sacks, O. (2014). The mental life of plants and worms, among others. *The New York Review of Books*, 61.

[180] Saito, H. (2001). Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chem Rev*, 101, 2497–2509.

[181] Salamanca-Cardona, L., Shah, H., Poot, A., Correa, F., Di Gialleonardo, V., Lui, H., Miloushev, V., Granlund, K., Tee, S., Cross, J., Thompson, C., & Keshari, K. (2017). *In vivo* imaging of glutamine metabolism to the oncometabolite 2-hydroxyglutarate in IDH1/2 mutant tumors. *Cell Metab.*, 26, 830–841.

[182] Samec, Z., Trojanek, A., & Samcova, E. (1994). Evaluation of ion transport parameters in a nafion membrane from ion-exchange measurements. *J. Phys. Chem.*, 98, 6352–6358.

[183] Sancak, Y., Markhard, A., Kitami, T., Kovacs-Bogdan, E., Kamer, K., Udeshi, N., Carr, S., Chaudhuri, D., Clapham, D., Li, A., Calvo, S., Goldberger, O., & Mootha, V. (2013). EMRE is an essential component of the mitochondrial calcium uniporter complex. *Science*, 342, 1379–1382.

[184] Schneider, C., Rasband, W., & Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, 9, 671–675.

[185] Seamon, K. (1980). Calcium- and magnesium-dependent conformational states of calmodulin as determined by nuclear magnetic resonance. *Biochemistry*, 19, 207–215.

[186] Senior, P. (1975). Regulation of nitrogen metabolism in *Escherichia coli* and *Klebsiella aerogenes*: studies with the continuous-culture technique. *J. Bacteriol.*, 123, 407–418.

[187] Sheff, M. & Thorn, K. (2004). Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast*, 21, 661–670.

[188] Shinar, G., Alon, U., & Feinberg, M. (2009a). Sensitivity and robustness in chemical reaction networks. *SIAM J. Appl. Math.*, 69, 977–998.

[189] Shinar, G. & Feinberg, M. (2010). Structural sources of robustness in biochemical reaction networks. *Science*, 327, 1389–1391.

[190] Shinar, G. & Feinberg, M. (2011). Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work. *Math. Biosci.*, 231, 39–48.

[191] Shinar, G., Milo, R., Martinez, M., & Alon, U. (2007). Input-output robustness in simple bacterial signaling systems. *Proc. Natl. Acad. Sci. USA*, 104, 19931–19935.

[192] Shinar, G., Rabinowitz, J., & Alon, U. (2009b). Robustness in glyoxylate bypass regulation. *PLoS Comput. Biol.*, 5, e1000297.

[193] Sigal, A., Milo, R., Geva-Zatorsky, A. C. N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., & Alon, U. (2006). Variability and memory of protein levels in human cells. *Nature*, 444, 643–646.

[194] Sikorski, R. & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, 122, 19–27.

[195] Slabodnick, M. & Marshall, W. (2014). *Stentor coeruleus*. *Curr. Biol.*, 24, R783–R784.

[196] Slabodnick, M., Ruby, J., Dunn, J., Feldman, J., DeRisi, J., & Marshall, W. (2014). The kinase regulator Mob1 acts as a patterning protein for *Stentor* morphogenesis. *PLOS Biol.*, 12, e1001861.

[197] Slabodnick, M., Ruby, J., Reiff, S., Swart, E., Gosai, S., Prabakaran, S., Witkowska, E., Larue, G., Fisher, S., Freeman, R., Gunawardena, J., Chu, W., Stover, N., Gregory, B., Nowacki, M., DeRisi, J., Roy, S., Marshall, W., & Sood, P. (2017). The macronuclear genome of *Stentor coeruleus* reveals tiny introns in a giant cell. *Curr. Biol.*, 27, 569–575.

[198] Song, H., Lee, J., Lee, M., Moon, J., Min, K., Yang, J., & Suh, S. (1999). Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from *Saccharomyces cerevisiae*. *J Mol Biol*, 293, 753–761.

[199] Sood, P., McGillivary, R., & Marshall, W. (2017). The transcriptional program of regeneration in the giant single cell, *Stentor coeruleus*. *bioRxiv*, https://dx.doi.org/10.1101/240788.

[200] Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S., Cyert, M., Hughes, T., Boone, C., & Andrews, B. (2006). Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell*, 21, 319–330.

[201] Southard, M., Dias, L., Himmelstein, K., & Stell, V. (1991). Experimental determinations of diffusion coefficients in dilute aqueous solution using the method of hydrodynamic stability. *Pharm. Res.*, 8, 1489–1494.

[202] Staddon, J. (2016). *Adaptive Learning and Behavior*. Cambridge: Cambridge University Pres.

[203] Steen, H., Jebanathirajah, J., Springer, M., & Kirschner, M. (2005). Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc. Natl. Acad. Sci. USA*, 102, 3948–3953.

[204] Stein, E., DiNardo, C., Pollyea, D., Fathi, A., Roboz, G., Altman, J., Stone, R., DeAngelo, D., Levine, R., Flinn, I., Kantarjian, H., Collins, R., Patel, M., Frankel, A., Stein, A., Sekeres, M., Swords, R., Medeiros, B., Willekens, C., Vyas, P., Tosolini, A., Xu, Q., Knight, R., Yen, K., Agresta, S., de Botton, S., & Tallman, M. (2017). Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood*, 130, 722–731.

[205] Stelling, J., Sauer, U., Szallasi, Z., Doyle, F., & Doyle, J. (2004). Robustness of cellular functions. *Cell*, 118, 675–685.

[206] Stojanovski, K., Ferrar, T., Benisty, H., Uschner, F., Delgago, J., Jimenez, J., Sole, C., de Nadal, E., Klipp, E., Posas, F., Serrano, L., & Kiel, C. (2017). Interaction dynamics determine signaling and output pathway responses. *Cell Rep.*, 19, 136–149.

[207] Strabue, R. (2014). Reciprocal regulation as a source of ultrasensitivity in two-component systems with a bifunctional sensor kinase. *PLOS Comput. Biol.*, 10, e1003614.

[208] Straube, R. (2013). Sensitivity and robustness in covalent modification cycles with a bifunctional converter enzyme. *Biophys. J.*, 105, 1925–1933.

[209] Sugiura, M., Shiotani, H., Suzaki, T., & Harumoto, T. (2010). Behavioural changes induced by the conjugation-inducing pheromones, gamone 1 and 2, int eh ciliate *Blepharisma japonicum*. *Eur. J. Protistol.*, 46, 143–149.

[210] Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J. S., Goldman, A. D., Nowacki, M., Schotanus, K., Jung, S., Fulton, R. S., Ly, A., McGrath, S., Haub, K., Wiggins, J. L., Storton, D., Matese, J. C., Parsons, L., Chang, W.-J., Bowen, M. S., Stover, N. A., Jones, T. A., Eddy, S. R., Herrick, G. A., Doak, T. G., Wilson, R. K., Mardis, E. R., & Landweber, L. F. (2013). The *Oxytricha trifallax* macronuclear genome: A complex eukaryotic genome with 16,000 tiny chromosomes. *PLOS Biol.*, 11, e1001473.

[211] Tait, S. & Green, D. (2010). Mitochondria and cell death: Outer membrane permeabilization and beyond. *Nat. Rev. Mol. Cell Biol.*, 11, 621–632.

[212] Tan, W. & Colombini, M. (2007). VDAC closure increases calcium ion flux. *Biochim. Biophys. Acta – Biomembr.*, 1768, 2510–2515.

[213] Tantama, M., Hung, Y., & Yellen, G. (2011). Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. *J. Am. Chem. Soc.*, 133, 10034–10037.

[214] Tao, W., Malone, C. L., Ault, A. D., Deschenes, R. J., & Fassler, J. S. (2002). A cytoplasmic coiled-coil domain is required for histidine kinase activity of the yeast osmosensor SLN1. *Mol. Microbiol.*, 43, 459–473.

[215] Tartar, V. (1961). *The Biology of Stentor*. Oxford: Pergammon Press.

[216] Tsvetkov, P., Protasevich, I., Gilli, R., Lafitte, D., Lobachov, V., Haiech, J., Briand, C., & Makarov, A. (1999). Apocalmodulin binds to the mysoin light chain kinase calmodulin target site. *J. Biol. Chem.*, 274, 18161–18164.

[217] Valentine, K., Ng, H., Schneeweis, J., Kranz, J., Frederick, K., Alber, T., & Wand, A. (2006). PDB ID: 2O5G. 10.2210/pdb2O5G/pdb.

[218] VanScyoc, W., Sorensen, B., Rusinova, E., Laws, W., Alexander Ross, J., & Shea, M. (2002). Calcium binding to calmodulin mutants monitored by domain-specific intrinsic phenylalanine and tyrosine fluorescence. *Biophys. J.*, 83, 2767–2780.

[219] Veit, B. E. & Fangman, W. L. (1988). Copy number and partition of the *Saccharomyces cerevisiae* 2 micron plasmid controlled by transcription regulators. *Mol. Cell. Biol.*, 8, 4949–4957.

[220] Ventura, A., Jiang, P., Van Wassenhove, L., Del Vecchio, D., Merajver, S., & Ninfa, A. (2010). Signaling properties of a covalent modification cycle are altered by a downstream target. *Proc. Natl. Acad. Sci. USA*, 107, 10032–10037.

[221] Wang, I., Politi, A., Tania, N., Bai, Y., Sanderson, M., & Sneyd, J. (2008). A mathematical model of airway and pulmonary arteriole smooth muscle. *Biophys. J.*, 94, 2053–2064.

[222] Ward, P., Lu, C., Cross, J., Abdel-Wahab, O., Levine, R., Schwartz, G., & Thompson, C. (2013). The potential for isocitrate dehydrogenase mutations to produce 2-hydroxyglutarate depends on allele specificity and subcellular compartmentalization. *J. Biol. Chem.*, 288, 3804–3815.

[223] Ward, P., Patel, J., Wise, D., Abdel-Wahab, O., Bennett, B., Coller, H., Cross, J., Fantin, V., Hedvat, C., Perl, A., Rabinowitz, J., Carroll, M., Su, S., Sharp, K., Levine, R., & Thompson,

C. (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*, 17, 225–234.

[224] Webb, R. (2003). Smooth muscle contraction and relaxation. *Adv. Physiol. Educ.*, 27, 201–206.

[225] Webb, T. & Francis, D. (1969). Mating types in *Stentor coeruleus*. *J. Protozool.*, 16, 758–763.

[226] Whitacre, J. (2012). Biological robustness: paradigms, mechanisms, and systems principles. *Front. Genet.*, 3, 67.

[227] Wiley, H. S. (2014). Open questions: The disrupted circuitry of the cancer cell. *BMC Biol.*, 12, 88.

[228] Wilson, D., Sutherland, C., & Walsh, M. (2002). Activation of smooth muscle myosin light chain kinase by calmodulin. Role of LYS(30) and GLY(40). *J. Biol. Chem.*, 277, 2186–2192.

[229] Wood, D. (1970). Parametric studies of the response decrement produced by mechanical stimuli in the protozoan, *Stentor coeruleus*. *J. Neurobiol.*, 1, 345–360.

[230] Wood, D. (1988). Habituation in *Stentor*: Produced by mechanoreceptor channel modification. *J. Neurosci.*, 8, 2254–2258.

[231] Xie, X., Baird, D., Bowen, K., Capka, V., Chen, J., Chenail, G., Cho, Y., Dooley, J., Farsidjani, A., Fortin, P., Kohls, D., Kulathila, R., Lin, F., McKay, D., Rodrigues, L., Sage, D., Touré, B., van der Plas, S., Wright, K., Xu, M., Yin, H., Levell, J., & Pagliarini, R. (2017). Allosteric mutant IDH1 inhibitors reveal mechanisms for IDH1 mutant and isoform selectivity. *Structure*, 25, 506–513.

[232] Xu, Q., Porter, S., & West, A. (2003). The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. *Structure*, 11, 1569–1581.

[233] Xu, X., Zhao, J., Xu, Z., Peng, B., Huang, Q., Arnold, E., & Ding, J. (2004). Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. *J. Biol. Chem.*, 279, 33946–33957.

[234] Xu, Y. & Gunawardena, J. (2012). Realistic enzymology for post-translation modification: Zero-order ultrasensitivity revisited. *J. Theor. Biol.*, 311, 136–152.

[235] Yan, H., Parsons, D., Jin, G., McLendon, R., Rasheed, B., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., GJ, G. R., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K., Velculescu, V., Vogelstein, B., & Bigner, D. (2009). IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.*, 360, 765–773.

[236] Yang, B., Zhong, C., Peng, Y., Lai, Z., & Ding, J. (2010). Molecular mechanisms of "off-on switch" of activities of human IDH1 by tumor-associated mutation R132H. *Cell Res.*, 20, 1188–1200.

[237] Yen, K., Travins, J., Wang, F., David, M. D., Artin, E., Straley, K., Padyana, A., Gross, S., DeLaBarre, B., Tobin, E., Chen, Y., Nagaraja, R., Choe, S., Jin, L., Konteatis, Z., Cianchetta, G., Saunders, J. O., Salituro, F. G., Quivoron, C., Opolon, P., Bawa, O., Saada, V., Paci, A., Broutin, S., Bernard, O. A., de Botton, S., Marteyn, B. S., Pilichowska, M., Xu, Y., Fang, C., Jiang, F., Wei, W., Jin, S., Silverman, L., Liu, W., Yang, H., Dang, L., Dorsch, M., Penard-Lacronique, V., Biller, S. A., & Su, S.-S. M. (2017). AG-221, a first-in-class therapy targeting acute myeloid leukemia harboring oncogenic IDH2 mutations. *Cancer Discov.*, 7, 478–493.

[238] Zhao, S., Lin, Y., Xu, W., Jiang, W., Zha, Z., Wang, P., Yu, W., Li, Z., Gong, L., Peng, Y., Ding, J., Lei, Q., Guan, K., & Xiong, Y. (2009). Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1α. *Science*, 324, 261–265.

[239] Zhao, Y., Araki, S., Wu, J., Teramoto, T., Chang, Y.-F., Nakano, M., Abdelfattah, A., Fujiwara, M., Ishihara, T., Nagai, T., & Campbell, R. (2011). An expanded palette of genetically encoded Ca^{2+} indicators. *Science*, 333, 1888–1891.

[240] Zheng, J. & Jia, Z. (2010). Structure of the bifunctional isocitrate dehydrogenase kinase/phosphatase. *Nature*, 465, 961–965.