Mathematical modeling of macronutrient signaling in *S. cerevisiae*

Amogh Jalihal

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Genetics, Bioinformatics, and Computational Biology

John J. Tyson, Co-chair
T. M. Murali, Co-chair
Silke Hauf
Jing Chen
Pavel Kraikivski

May 11, 2019
Blacksburg, Virginia

Keywords: Yeast, Signaling, Mathematical Modeling, Boolean Models, RNAseq

Mathematical modeling of macronutrient signaling in *S. cerevisiae*

Amogh Jalihal

(ABSTRACT)

In eukaryotes, distinct nutrient signals are integrated in order to produce robust cellular responses to fluctuations in the environment. This process of signal integration is attributed to the crosstalk between nutrient specific signaling pathways, as well as the large degree of overlap between their regulatory targets. In the budding yeast *Saccharomyces cerevisiae*, these distinct pathways have been well characterized. However, the significant overlap between these pathways confounds the interpretation of the overall regulatory logic in terms of nutrient-dependent cell state determination. Here, we propose a literature-curated molecular mechanism of the integrated nutrient signaling pathway in budding yeast, focussing on carbon and nitrogen signaling. We build a computational model of this pathway to reconcile the available experimental data with our proposed molecular mechanism. We evaluate the robustness of the model fit to data with respect to the variations in the values of kinetic parameters used to calibrate the model. Finally, we use the model to make novel, experimentally testable predictions of transcription factor activities in mutant strains undergoing complex nutrient shifts. We also propose a novel framework, called BoolODE for utilizing published Boolean models to generate synthetic datasets used to benchmark the performance of algorithms performing gene regulatory network inference from single cell RNA sequencing data.

Mathematical modeling of macronutrient signaling in *S. cerevisiae*

Amogh Jalihal

(GENERAL AUDIENCE ABSTRACT)

An important problem in biology is how organisms sense and adapt to ever changing environments. A good example of an environmental cue that affects animal behavior is the availability of food; scarcity of food forces animals to search for food-rich habitats, or go into hibernation. At the level of single cells, a range of behaviors are observed depending on the amount of food, or nutrients present in the environment. Moreover, different types of nutrients are important for different biological functions in single cells, and each different nutrient type will have to be available in the right quantities to support cellular growth. At the subcellular level, intricate molecular machineries exist which sense the amounts of each nutrient type, and interpret this information in order to make a decision on how best to respond. This interpretation and integration of nutrient information is a complex, poorly understood process even in a simple unicellular organism like the budding yeast. In order to understand this process, termed nutrient signaling, we propose a mathematical model of how yeasts respond to nutrient availability in the environment. Our model advances the state of knowledge by presenting the first comprehensive mathematical model of the nutrient signaling machinery, accounting for a variety of experimental observations from the last three decades of yeast nutrient signaling. We use our model to make predictions on how yeasts might behave when supplied with different combinations of nutrients, which can be verified by experiments. Finally, the cellular machinery that helps yeasts respond to nutrient availablity in the environment is very similar to the machinery in cancer cells that causes them to grow rapidly. Our proposed model can serve as a stepping stone towards the construction of a model of cancer's responses to its nutritional environment.

To my parents.

Acknowledgments

I have many people to thank for my amazing journey through my PhD these last five years, but there are even more people responsible for shaping me into the person I was before setting foot in Blacksburg. I would like to acknowledge these pillars of support first. I would like to start by thanking my teachers in high school, Ms. Vedapathi for introducing to me the world of biology, and Dr.s K.S. Balaji and Balaji Sampath for the amazing energy with which they taught physics in those evening classes. The time spent with cohort of RSI'10, and the amazing conversations with Prof. V. Balakrishnan from the physics department at IIT form my earliest convictions to enter the world of the pure sciences. I have gained immensely from the discussions and support offered by my friends at SASTRA, and for this I would like to thank Vibishan Balasubramanian, Juhi Baskar, C.S. Anirudh, M. Sankaranarayan, Saradha Venkatachalapathy, and Anand Murali. Two professors have played an important role in encouraging me through college: I would like to thank Prof. Vigneshwar Ramakrishnan for introducing me to bioinformatics and for his unwavering support ever since, and Prof. Anbumathi Palanisamy for introducing me to the world of systems biology and for her support as I applied to graduate programs.

For most students, the first exposure to 'real' research decides their interest in continuing to pursue a research career. I will forever remain immensely grateful to Prof. Akatsuki Kimura from the NIG, and the Cell Architecture Lab, for having provided me with an amazing research experience, and constant encouragement throughout my graduate research career. From my second research experience, at Berkeley, I would like to thank Prof. Abby Dernburg, and Dr.s Ofer Rog, Christina Glazier, Keith Cheveralls, and Yumi Kim for their support and encouragement during a lonely period in my life. Last, but not least, I would like to thank Prof. Silke Hauf for an unforgettable semester in my life, where I felt like I could finally be an independent researcher. Thank you Silke, for teaching me so much.

My life in Blacksburg these last five years has been extremely rewarding intellectually, and every person I have met has influenced the way I see the world today. I will start by thanking Mark Petersen, for introducing me to the quaint world of typewriters, and to the rich world of poetry and prose that he lives in; my roommates, Ashwath Kumar and Radhakrishnan Vinayagam, for those wonderful times in Foxridge; my friends from Kala, Prathyush Kashyap, Satyajit Upasani, Rutuja Ksheersagar, and others, who taught me so much about Hindustani and Carnatic music; the entire GBCB family, both past and current, including Daniel Chen, Long Tian, Wei He, Amanda Fisher, Hazem Sharaf, Jiyoung Lee, Nuria Tubau-Juni, Bronson Weston and many others for the many happy times spent on Thursday afternoons during GBCB seminars, or in one of many potlucks on the weekends. Thanks to Alex Song for being a great friend, and for many lovely conversations. Thanks also to Sarah

Kadelka for being a motivational running companion in the times when I most needed the fresh air. For being a reliable source of knowledge and motivation, I would like to thank the past and current members of the Computational Systems Biology Group , including Mitchel Wagner, Aditya Bharadwaj, Dhruva Sahasrabuddhe, Tanya Tyagi, and Nure Tasnina. Outside of Virginia Tech, I would like to thank Tara Narayanan and Sumanth Srinivasan for keeping me grounded, and reminding me that there is life outside of the academy.

The people closest to me have helped me at every step of my way through the problems I faced during my research, giving me honest and encouraging feedback. Special thanks to Jeffrey Law, and Aditya Pratapa for always having my back, for always being around to help me with my code or paper. I have learned a lot from Jeff, who despite being very busy with his work and family, always made time for me and my Emacs obsession. I have greatly enjoyed working with Aditya on two projects, and I owe a great deal of my own development outside my shell as a biologist to my interactions with him. I will greatly miss our coffee breaks. Finally, there are those constants in my life who have been around no matter what I was going through. I would like to thank Brittany Boribong for all her support and companionship, and my parents, and my brother Ameya, for always being there for me.

I owe a great deal to my committee Prof. T.M. Murali, Prof. John Tyson, Prof. Silke Hauf, and Prof. Jing Chen who have given me infinite freedom, and have been infinitely patient, as I stumbled my way around my projects, and to Dr. Pavel Kraikivski for being a fantastic mentor, and for always being around when I had questions.

Contents

List of Figures

List of Tables

List of Abbreviations

p Parameter vector

$\mathcal{H}$ The Hessian matrix

Σ The covariance matrix

C Cost function

DYNGEN Framework to generate synthetic scRNAseq datasets for evaluation of pseudotime inference algorithms.

GNW GeneNetWeaver, a framework to generate synthetic bulk RNAseq datasets for GRN inference algorithms. Originally used in the DREAM4 challenge.

xyzΔ Deletion strain of gene *XYZ*

Bcy1 Bypass of CYclic-AMP requirement. Inhibitory subunit of PKA

cAMP cyclic Adenosine Monophosphate

CSRE Carbon Stress Response Element

Cyr1 CYclic AMP Requirement 1 (Adenylyl Cyclase).

GAAC General Amino Acid Control

GAP GTPase activating factor

GEF Guanine exchange factor

Ira1/2 Inhibitory Regulator of the RAS-cAMP pathway

MCMC Markov Chain Monte Carlo Algorithm

NCR Nitrogen Catabolite Repression

PCA Principal Component Analysis

PDE1/2 Phosphodiesterase.

PDS Post-diauxic shift

PKA Protein Kinase A. Kinase complex regulating glucose response.

RIBI The Ribosome Biogenesis regulon

RTG Retrograde Signaling

Snf1 Sucrose Non-Fermenting 1. Kinase complex regulating carbon adaptation responses.

SSA Stochastic Simulation Algorithm

TORC1 Target of Rapamycin Complex 1. Kinase complex regulating nitrogen responses..

Tpk1/2/3 Takashi's Protein Kinase. Catalytic subunit of PKA

Chapter 1

Introduction

1.1 Background and significance of nutrient signaling

A long standing question in biology has been the relationship between cell growth and metabolism [1, 2, 3, 4, 5, 6, 7]. Empirical models constructed for bacteria show that cell growth is driven by scaling up protein translation [8]. However, in order to sustain high translation rates, the cell needs to maintain an adequate supply of raw materials (amino acids, also referred to as nutrients) and energy (ATP). The biochemical mechanisms that sense and signal the levels of available raw materials, also known as nutrient signaling, are the cornerstone of growth control. In prokaryotes, the nutrient signaling mechanisms coupling nutrient availability to ribosome biogenesis and metabolism are well understood, and are largely mediated by the second messenger ppGpp [9, 10, 11]. However, nutrient signaling is far more complex in higher eukaryotes [12].

The investigation of the mechanism of growth control in yeast has its origins in the practical utility of budding yeasts *S. cerevisiae* for fermentation[1]. Studies of the fermentation pathway in yeast eventually lead to the elucidation of the biochemical pathways composing carbon metabolism. In a departure from bacterial studies studying metabolism, the discovery of yeast vacuoles motivated early studies that quantified the storage levels of amino acid distributed across the cytoplasm and the vacuole [14, 15]. These observations provided impetus to the study of transporters, as well as the regulation of amino acid distribution. Once a clear picture of nitrogen and carbon metabolism emerged in the 1980s and 1990s [16, 17, 18, 19], the focus of research shifted to the mechanisms of regulation of enzyme expression. The last decade of the 20^{th} century saw the beginnings of the description of three key regulatory kinases (Protein Kinase A [20, 21], Snf1 [19], and TORC1 [22]) that defined the distinct nutrient signaling pathways [23].

In essence, the nutrient signaling pathways gauge the sufficiency of environmentally derived nutrient signals and integrate these signals in order to produce a robust adaptation response. The consequences of this signal integration for cellular physiology include the remodeling of global metabolism, modulation of cellular growth rate, as well as the mounting of stress responses and accumulation of storage nutrient reserves. The last two decades of yeast research have uncovered the other roles of the nutrient signaling pathways. Current views place nu-

[1] For an early review of yeasts in fermentation processes, see Harden, 1911[13].

trient signaling as the regulator for decision making of various stress responses [24, 25, 26], the G1-S transition in the cell cycle [27], and autophagy [28, 29]. Finally, in all of the above contexts, it has been recognized that crosstalk between various nutrient signaling pathways is important to accomplish robust decision making in the face of fluctuating nutrient availability [30, 31, 32, 33].

The current picture of nutrient signaling and its relation to growth control is complicated by the large number of interacting molecular players, and the diverse roles they play in various cellular contexts. Multiple reviews have been published outlining the current state of our knowledge of these pathways in yeasts and higher eukaryotes [23, 34, 35, 36, 37]. However, it is unclear if these qualitative descriptions are consistent with, or are even sufficient to explain, the vast body of experimental observations probing yeast's responses to the nutrient environment. In particular, reasoning about the roles of specific molecular interactions is confounded by multi-pronged environmental and genetic perturbations, such as nutrient shift phenotypes in genetic mutants [38], strain-specific metabolic and regulatory differences [39], as well as contradictory and controversial claims regarding molecular interactions.

In this dissertation, we directly interrogate the nutrient signaling system in *S. cerevisiae*, restricting ourselves to the signaling components and the interactions between them. We construct a literature curated interaction network representing the signaling pathway, which we convert into a mathematical model. This model allows us to represent genetic and environmental perturbation, which makes it easy to compare our proposed mechanism against a variety of different types of experimental data. Importantly, we are able to use our proposed model to make novel predictions of *S. cerevisiae*'s response to complex nutrient shifts in a range of genetic mutants. We hope that this first model of integrated nutrient signaling will serve as a step towards understanding the mechanism of growth control in *S. cerevisiae* and higher eukaryotes.

1.2 Mathematical Modeling

Cell physiology, or the collection of cellular responses which emerge from biochemical interactions at the subcellular level, is often described as being "complex". The term "complexity", while difficult to define precisely, carries with it the notion that cell physiology is not simply a sum of a cell's components. The property of cells that precludes an intuitive understanding of cellular phenotypes is the (large) number of interacting molecules composing them, and the time-varying, often non-linear interactions between these molecules [40]. Dynamical modeling is a mathematical framework that has proven to be highly successful in providing insights into the emergence of cellular physiology from biochemical reactions.[2] In this dissertation, we will largely be concerned with dynamical models expressed as systems of

[2]From a physical perspective, a complete mathematical description of a cell, with all its time varying molecular constituents, as well as fluxes of matter and energy, will fall in the domain of non-equilibrium statistical mechanics. Work by Keizer and others demonstrated that even in non-equilibrium conditions, the

differential equations. Such a formulation admits sophisticated analyses of the dynamical process under study, such as stability analysis and bifurcation theory [44].

The construction of a dynamical model of a cellular process is done in stages. First, the experimental evidence for the interaction and dynamics of the key molecular players is curated from the literature. Typically, the existing literature evidence will be summarized as mechanistic hypotheses, represented as interaction diagrams [45]. Once high confidence mechanistic hypothesiss are constructed, they are converted to systems of Ordinary Differential Equation (ODE), where each molecular interaction is represented as a biochemical rate law. Further, the reaction rate laws have to be parameterized by kinetic rate constants which govern the timescales of the molecular processes. These rate constants, which are model parameters, are estimated by 'fitting' the model to quantitative data curated from the literature. This calibrated model can then be analyzed analytically or numerically to explore the implications of the initial mechanistic hypothesis on the observed model behavior. Over the past decades, many dynamical models have been published, each providing great insight into relationship between the underlying molecular mechanism and the resulting cellular phenotype. Some successful mechanistic models of physiological processes include the yeast cell cycle [46], circadian oscillations [47], bacterial motility [48], NF-κB signaling [49], and insulin signaling [50].

1.3 Organization and Contributions

In this work, we construct a dynamical model of the macronutrient signaling pathway in the budding yeast *S. cerevisiae*, integrating the pathways that respond to carbon and nitrogen sufficiency. The objective of this model is to dissect the contributions of nutrient-specific signaling pathways in the

1. integration of diverse nutrient signals derived from the environment, and, subsequently
2. the joint regulation of downstream adaptation responses that define the cellular response to a nutritional shift.

To this end, we constructed a literature curated 'wiring diagram' representing the nutrient signaling system in *S. cerevisiae*. We examined the primary literature in the area of carbon and nitrogen signaling in recent decades in order to collect quantitative as well as qualitative assay measurements in investigations of the molecular interactions relevant to short time scale nutrient responses. This resulted in a collection of quantitative time course measurements and quantitative perturbation data (end point measurements) that were used in model calibration, as well as a collection of qualitative phenotype measurements that were used to validate the model (Chapter 2).

steady states of the phenomenological rate laws remain valid; cf [41, 42, 43]

Next, we converted our proposed 'wiring diagram' into a system of coupled nonlinear Ordinary Differential Equations. We used the Standard Component Framework to define the form of the biochemical rate equations based on the interaction evidence. We calibrated the model with the quantitative time series and perturbation data curated from the literature (Chapter 3).

Having constructed and calibrated the model, we studied the relationship between the behavior of the nutrient signaling model and the model parameterization by carrying out a model robustness analysis. This analysis revealed the parts of the model that are constrained by the data. One of the outcomes of this analysis is a collection of 18,000 alternate sets of parameter values spanning a large region of parameter space that all provide an acceptable quantitative fit to the curated data (Chapter 4).

We studied the predictions made by the nutrient signaling model. We compared model predictions using the collection of parameter sets to a set of qualitative nutrient shift phenotypes curated from the literature. Finally, we used the model to make novel, experimentally testable predictions of the global transcriptional response of yeast undergoing complex nutrient shifts in 18 genetic mutants (Chapter 5).

In an exploration of automated construction of ODE models, one of the applications that we came across was the use of models of gene regulatory networks to create benchmarking datasets for algorithms that attempt to infer network structure from experimental time courses. We identified drawbacks in current approaches, and proposed a novel framework, BOOLODE, to harness manually constructed Boolean models to generate synthetic datasets. These datasets were used in a benchmarking study to evaluate the performance of algorithms that infer gene regulatory networks from single cell RNAseq data. BOOLODE is discussed in Chapter 6.

Chapter 2

Nutrient Signaling in Yeast

Jalihal, A.P., Kraikivski, P., Murali, T.M., Tyson, J.J., 2020. Modeling and analysis of macronutrient signaling in budding yeast. *Under review.* Also available on *bioRxiv* (https://doi.org/10.1101/2020.02.15.950881)

Depending on the amounts derived from the environment, nutrients can be categorized as macro- or micro-nutrients. Macronutrients are composed of compounds rich in carbon, nitrogen, phosphate, and sulphur, while micronutrients are composed of minerals and vitamins which are typically scarce in the environment. In this work, we will be focusing on the signaling pathways associated with macronutrients, specifically carbon and nitrogen signaling. As discussed later in this chapter, these signaling pathways play significant roles in the adaptation of cellular metabolism, regulation of growth, and the decision to mount stress responses.

The macronutrient signaling system in yeast is composed of three distinct signaling pathways. Broadly, the TORC1-Sch9 pathway senses nitrogen sufficiency, the cAMP-PKA pathway conveys the glucose status, and the Snf1 pathway regulates carbon adaptation responses. The first part of this chapter will describe the details of these pathways as they relate to nutrient signaling. In the second part of this chapter, we will outline the relevant mechanistic interactions curated from the literature used to create our proposed interaction network of nutrient signaling.

Carbon sensing and signaling The cAMP/Ras/PKA pathway is activated by upstream glucose sensing mechanisms. Subsequently, PKA rapidly carries out two main functions, the inhibition of respiration pathway, and the activation of fermentation pathway. This drastic response to glucose availability is referred to as "glucose rexpression" [51]. The addition of glucose to carbon-starved yeast cells activates Adenylate Cyclase (Cyr1) indirectly via Ras2 and directly through the action of the glucose sensing G-protein Gpr1/Gpa2 [52]. This stimulation of Cyr1 activity results in a rapid increase in the intracellular concentration of cyclic Adenosine Mono-Phosphate (cAMP) [20, 53]. The yeast Protein Kinase A (PKA) exists as a heterotetramer, as a complex of two regulatory subunits, Bcy1, and two catalytic subunits, which can be one of the three TPK genes Tpk1-3. cAMP subsequently binds to the regulatory subunit Bcy1, freeing the catalytic subunits. The active Tpk subunits phosphorylate and activate the phosphodiesterases Pde1/2 (present in our mathematical model but not shown in Figure 2.1), which degrade cAMP, implementing a negative feedback loop [54]. Furthermore, PKA has been suggested to implement another negative feedback loop by

inhibiting Ras2 activation, upstream of cAMP production. Putative PKA target sites have been identified on Ira1,2 the Ras2 GTPase activating proteins (GAPs) that inactivate Ras2. PKA-mediated hyperphosphorylation of Cdc25, the Ras2 guanine nucleotide exchange factor (GEF) has been proposed as another mechanism by which Ras2 activity is inhibited [55].

BOX 1: Convention used in cartoons of molecular mechanism

The network diagrams (molecular wiring diagrams) shown in this chapter follow a fixed convention. Molecular species are represented by filled black shapes, while the biochemical reactions they participate in are represented by lines with arrow heads. A pair of curved, reversible black arrows represents the interconversion between the active and inactive forms of a molecular species by post translational modification. A gray, dashed line represents a regulatory interaction. Such a line can end in an arrow, representing the activation of the process on which it is incident, or it can end in a blunt 'T,' representing an inhibition of the process on which it is incident. Straight, solid black lines represent production or degradation of a species. Solid gray lines ending in gray circles represent complex formation. The presence of a gray circle on a molecular species represents a subunit or a chemical moeity. Phosphate groups are represented by red filled circles. Degradation events are represented using a red cross. In order to make the cartoon legible, multiple regulatory interactions originating from a given molecular species are represented by forking lines, where the fork is represented by a black filled circle. Finally, black filled rectangles represent transcriptional states that lead to distinct cellular phenotypes. Note that some species may be left out in the diagram in order to retain readability.

Activated PKA regulates cell growth in two distinct ways. First, PKA inhibits respiration and upregulates the fermentation pathway. Specifically, PKA implements the glucose repression pathway by inhibiting carbon adaptation responses [56] and general stress responses [57]. This is an adaptation to the natural environment of the yeast, where glucose is nonlimiting, and rapid energy production provides a fitness advantage. Second, PKA upregulates the ribosome biogenesis (RIBI) pathway. PKA regulates expression of the *RIBI* regulon by inhibiting transcriptional repressors like Dot6/Tod6 [58, 59]. Finally, PKA activates trehalase which triggers the consumption of storage carbohydrates like trehalose or glycogen that are accumulated during starvation. This activation results in a rapid availability of ATP for anabolic processes which further aids the transition to fermentative growth on the re-availability of glucose in the environment. The interactions described above constitute the "Fermentative Growth" block appearing on the let in Figure 2.1.

Carbon adaptation responses Early investigation into the mechanism of yeast adaptation to growth on sucrose (as compared to growth on glucose, a fermentable sugar) led to the identification of the Sucrose Non Fermenting (SNF) mutants. Of these, the Snf1 kinase was shown to be essential for the activation of invertase in order to degrade sucrose to its monosaccharide constituents. [18]. Snf1 was later shown to be the yeast homolog of mam-

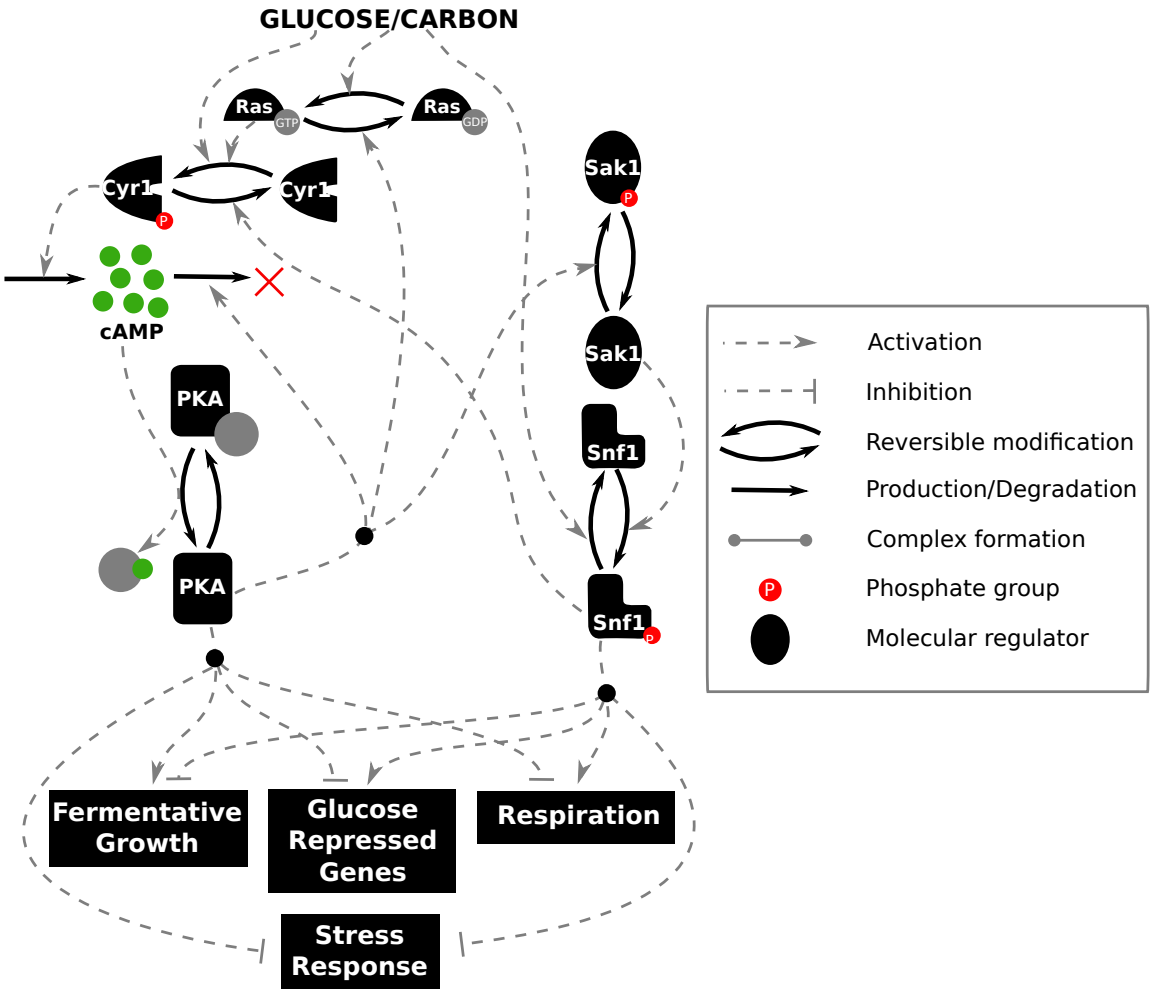


Figure 2.1: The carbon signaling pathway in yeast is composed of the glucose signaling branch, i.e., the cAMP-Ras-PKA pathway depicted on the left, and the carbon adaptation branch, i.e. the Snf1 pathway depicted on the right. The exact mechanism of inhibition of Snf1 in the presence of glucose is not well understood, but is known to originate at Gpr1 (not shown), and is effected by the Reg1/Glc7 phosphatase. Here, we represent a direct inhibition of Snf1 in the presence of glucose. Note that the Phosphodiesterases (PDE1/2) which degrade cAMP to AMP are not shown on this diagram. The boxes at the bottom of the figure indicate the biological processes regulated by PKA and Snf1. The transcription factors which implement the gene regulatory program leading to these cellular phenotypes are discussed in the text.

malian AMP Kinase [19]. Interestingly, unlike in higher eukaryotes AMP binding does not seem to be crucial for Snf1 activity, even though structural studies have shown that ADP and AMP can bind to Snf1 *in vitro* [60].

In the presence of glucose, the Reg1-Glc7 phosphatase complex dephosphorylates and subse-

quently inactivates Snf1. The relationship between glucose uptake/sensing and the activity of Reg1/Glc7 remains unclear [61]. In the absence of glucose, the Sak1/Tos3/Elm1 kinases phosphorylate and activate Snf1 [62, 63].

Activated Snf1 upregulates Carbon Stress Response (CSRE) genes which effect carbon source dependent control of gluconeogenesis. This is achieved via the inhibition of the Mig1 transcriptional repressor which implements the glucose repression pathway [64]. Apart from mediating the post-diauxic shift (PDS) via the stimulation of Gis1 activity, Snf1 has been implicated in the inhibition of general stress response regulators such as Msn2/4 [26] and the nutrient dependent regulation of nitrogen adaptation response regulators like Gcn4 [65].(We do not examine the mechanism of Msn2/4 regulation here. A mathematical investigation of Msn2/4 nucleocytoplasmic oscillations was carried out by Garmendia-Torres et al. 2007 [66].) Overall, the Snf1 pathway plays the crucial role of adapting cellular metabolism to non-glucose environments. We believe that these molecular players and interactions that they participate in are sufficient to describe the logic of carbon signaling; The right hand side of Figure 2.1 illustrates these molecular interactions.

Nitrogen sensing and signaling TORC1 senses the quality and quantity of nitrogen sources in the environment, the so-called amino acid sufficiency, in order to regulate growth pathways. TORC1 is localized at the lysosome-like vacuolar membrane in yeast, and is activated by a network of GTPases which directly or indirectly sense the concentrations of various nitrogen sources inside and outside the cell [67]. The presence of rich nitrogen sources results in the relocalization of TORC1 to the vacuolar membrane, followed by the stimulation of TORC1 kinase activity, leading to the phosphorylation and activation of Sch9 (the yeast homolog of mammalian Protein Kinase B) [68]. Sch9 regulates a variety of growth processes and inhibits stress responses. On the other hand, the TORC1 dependent activity of the Tap42/Sit4 Phosphatase branch ensures the upregulation of nitrogen stress responses and inhibition of growth responses in response to nitrogen starvation [69] (Figure 2.2).

Nitrogen adaptation responses The nitrogen adaptation response is intimately linked to two metabolic processes: amino acid transport and amino acid biosynthesis. The molecular mechanism governing the nitrogen adaptation response, termed the General Amino Acid Control (GAAC) response has been elucidated in great detail. In the presence of rich amino acid sources like glutamine, amino acid biosynthetic pathways and high specificity transporters are downregulated in a process called Nitrogen Catabolite Repression (NCR). The TORC1 pathway regulates the factors involved in the GAAC and NCR responses [16, 17, 70]. Gln3 is the primary transcriptional regulator that has been implicated in the NCR response. In response to declining nitrogen status, a graded Gln3 response has been suggested to modulate amino acid consumption and increase the expression of non-specific amino acid permeases like Gap1 [71] [72]. Gcn4 is another important transcription factor that is activated by the Gcn2 kinase in response to amino acid starvation, which results in the derepression of amino acid biosynthesis pathways. Finally, the retrograde pathway regulated by the transcription factors Rtg1/3 are stimulated when ammonia derived from the environment can be used

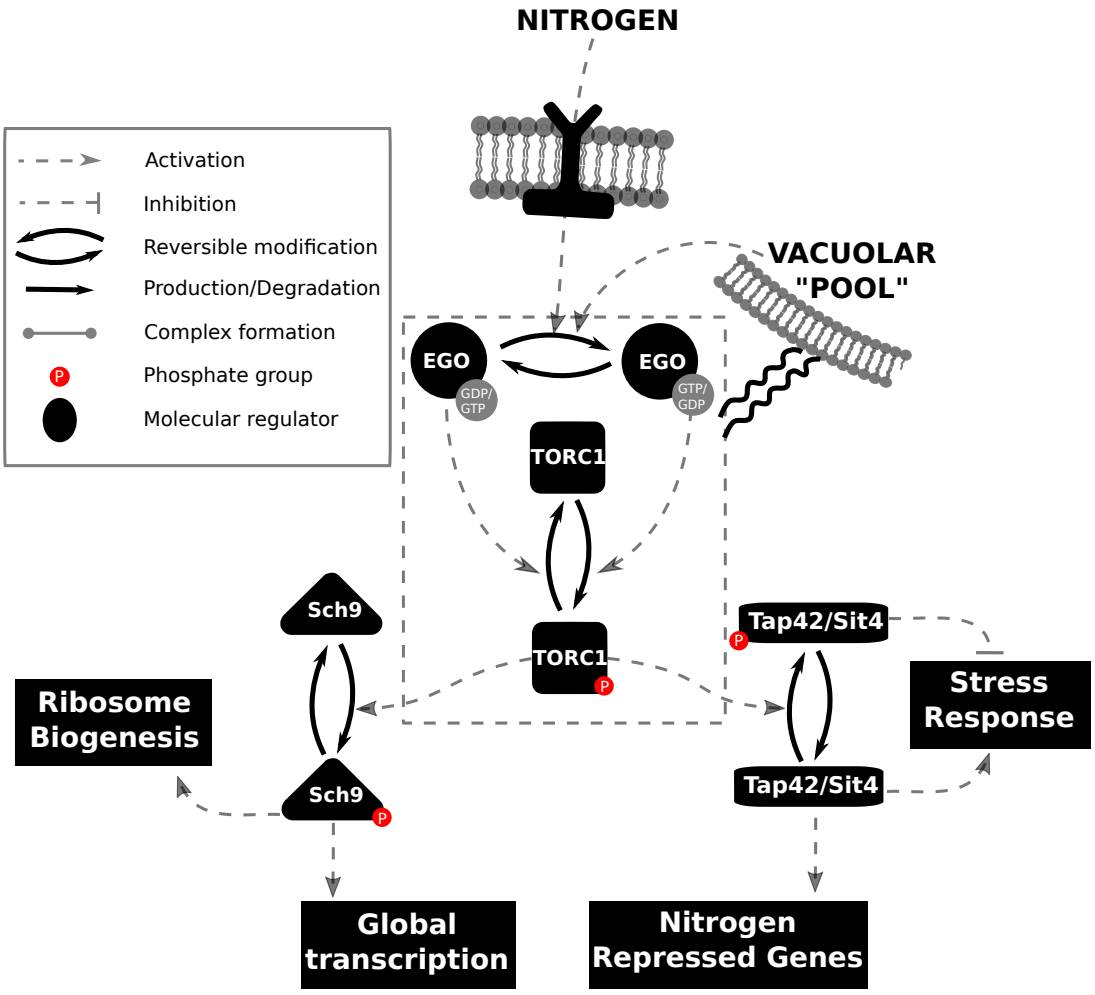


Figure 2.2: The nitrogen signaling system is represented by the TORC1 pathway. The presence of rich nitrogen sources leads to the activation of TORC1 kinase, subsequently leading to the activation of the Sch9 branch (left) and the inhibition of the Tap42-Sit4 branch (right). Nitrogen starvation leads to the inhibition of TORC1 activity, in turn activating nitrogen adaptation and stress responses and inhibiting growth responses.

in transamination of α ketoglutarate to produce glutamine and glutamate in response to decline in amino acid availability. All the transcription factors described above are regulated directly or indirectly by TORC1, usually by localization to the cytoplasm. The summary of these interactions is presented in Figure 2.2.

2.1 Pathway crosstalk

Extensive crosstalk has been documented between the pathways described above. This pathway crosstalk has been implicated as being essential for integrating disparate nutrient signals and coordinating nutrient availability with metabolic demands, in turn constraining global transcription and translation.

Interactions between Sch9 and PKA have been reported multiple times in the literature, but the significance of these interactions remain uncertain [73, 74, 75]. Furthermore, Sch9 and PKA share a large number of common targets. In fact, Sch9 was first identified to be a PKA homolog in a screen for mutants that could rescue the PKA^{-} phenotype [76]. The stress regulatory kinase Rim15 which regulates a variety of stress and nutrient adaptation factors is jointly regulated by both Sch9 and PKA, and this overlap of regulation has been shown to cause a diverse set of transcriptional responses downstream [77, 78, 79]. We do not consider the regulation of Rim15 in our mathematical model.

Further, this crosstalk is also significant for the regulation of ribosome biogenesis. The Sch9 and PKA pathways jointly regulate the expression of rRNAs and the transcription of Ribosomal Protein (RP) genes via the inhibition of transcriptional repressors like Dot6/Tod6 [30, 80] [1]. Apart from the regulation of growth responses, the two pathways regulate global cellular states in response to nutritional shifts [31, 32, 83, 84].

Unlike in mammals where AMPK has been shown to be upstream of mTOR, there is evidence that the relationship between these regulators is different in yeast. Crosstalk between the Snf1 and TORC1 pathways has been documented, with major consequences for global transcriptional regulation [85, 86]. However, the exact mechanism mediating the crosstalk remains controversial (cf. [87] and [88]). Irrespective of the exact mechanism, Snf1 and TORC1 also share common transcriptional targets, including the nitrogen adaptation factor Gln3 [89]. The summary of the crosstalk between the carbon and nitrogen signaling pathways, along with the full carbon and nitrogen signaling pathways, is presented in Figure 2.3.

2.2 Perspective

Over the last half a century, the biochemistry of yeast metabolism has been studied in great detail. In recent decades, the regulatory mechanisms governing transitions in global cellular metabolism in response to nutrient shifts have been elucidated. These regulatory

[1] Sfp1 is another key transcriptional regulator that affects *RP* expression in response to nutrient shifts. Results from Jorgensen et al. cemented the idea that Sch9 and Sfp1 were the global regulators of ribosomal gene exression [81]. Unlike Sch9, Sfp1's primary role as a transcriptional regulator is activated when it complexes with TORC1 and localizes to the nucleus [82]. For the sake of simplicity, we will consider only the Sch9 branch of TORC1 signaling to the downstream regulation of ribosome biogenesis here.

mechanisms are composed of multiple signaling pathways which integrate nutrient signals from the extracellular environment, the vacuole, and the cytoplasm, and coordinate the expression of hundreds of metabolic enzymes via the localization of transcription factors. The structure of the signaling pathways exhibit feedback loops, and cross-talk interactions with components of other signaling pathways including the cell cycle and autophagy pathways [28, 90]. The nutrient signaling pathways are typically studied in specific media or under specific drug treatments. The discovery of cross-talk interactions between nutrient signaling pathways indicate roles of these pathways outside of the specific contexts in which they have been described. As we have discussed in this chapter, the functional significance of the cross-talk interactions in signal integration remains unclear [30, 38, 91, 92]. We wish to investigate the molecular interactions that govern the nutrient signal integration mechanisms in yeast with the goal of reconciling disparate pathway-specific experimental data with the dynamics demonstrated by the full integrated model. In particular, we wish to investigate the combined regulation of the targets of the nutrient signaling pathways, namely the regulatory transcription factors, and the effects of complex nutrient shifts and genetic perturbations of the signaling molecules on this regulation. This chapter has discussed the major molecular interactions in yeast that capture the complexity of the nutrient response pathway. The regulatory interactions described here are summarized in the wiring diagram illustrated in Figure 2.3.

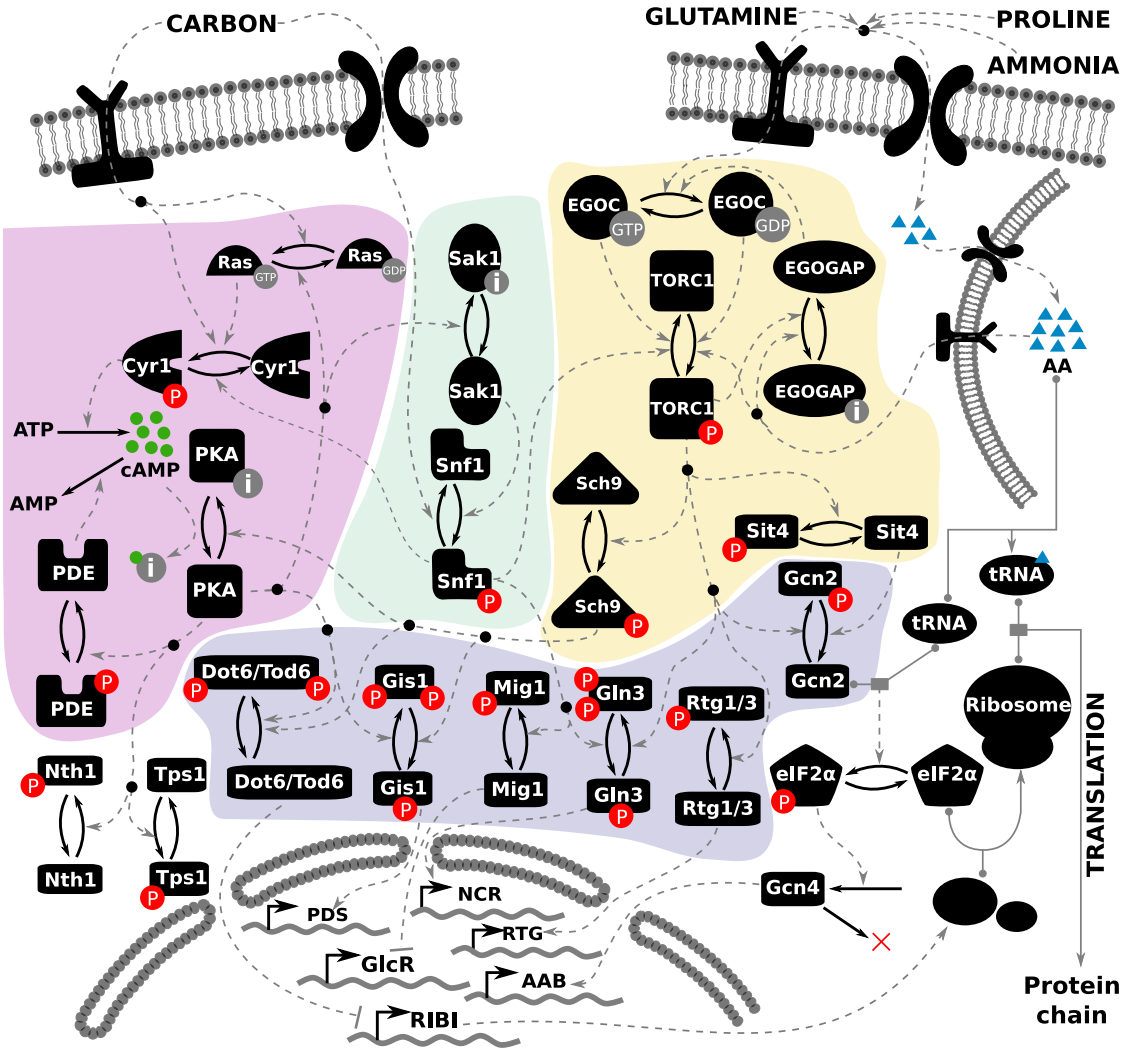


Figure 2.3: Literature-curated molecular regulatory network of nutrient signaling in yeast cells. The cAMP-PKA pathway is depicted on the left (highlighted in purple), the Snf1 pathway in the middle (highlighted in green) and the TORC1 pathway on the right (highlighted in yellow). The readouts of the model are the activities of the transcription factors towards the bottom of the diagram (highlighted in blue), whose target genes at the bottom of the figure comprise various regulons: GlcR, glucose-repressed genes; PDS, post-diauxic shift element; NCR, nitrogen catabolite repression genes; RTG, retrograde genes; AAB, amino acid biosynthesis; and RIBI, ribosome biogenesis. Each black icon represents a molecular species. Post-translational modifications are represented by a pair of solid arrows pointing in opposite directions. Phosphorylation is indicated by 'P' in a red circle. The inactive form of a species is indicated by the letter 'i' in a gray circle. Guanidylation is indicated by 'GTP' in a grey circle. Regulatory signals are represented by dashed gray lines. Complex formation is indicated by a solid gray line, with the binding partners indicated by gray circles. The inputs to the model are shown at the top of the figure, represented by one carbon input, and three nitrogen inputs: glutamine, ammonia and proline. The intracellular amino acid pool stored in the vacuole is represented by the membrane-bound structure on the top right.

Chapter 3

Model Construction and Calibration

Jalihal, A.P., Kraikivski, P., Murali, T.M., Tyson, J.J., 2020. Modeling and analysis of macronutrient signaling in budding yeast. *Under review.* Also available on *bioRxiv* (https://doi.org/10.1101/2020.02.15.950881)

Dynamical models have proven to be very useful tools in reasoning and interrogating complex biological systems. Since these models are easy to simulate numerically, multiple hypotheses can be quickly tested by carrying out *in silico* experiments. This provides a method to distinguish competing molecular mechanisms, interrogate contradictory experimental results, and ultimately generate novel biological hypotheses that can be validated.

In this chapter, we will briefly review previously published models that have explored various aspects of nutrient signaling in eukaryotes. We will then describe the Standard Component Modeling framework (Section 3.2) which we use to define our proposed model of nutrient signaling in yeast (Section 3.3).

3.1 Previously published models

This section presents a survey of the previously published models that have investigated nutrient-specific signaling pathways. These models tend to focus on a small number of interactions and on specific nutrient condition dependent phenotypes. The published methods are organized by the pathway that they model. The cAMP-Ras-PKA pathway sense glucose availability. The Snf1 pathway regulates carbon adaptation responses in the absence of glucose. Finally, the TORC1 pathway senses amino acid sufficiency and regulates nitrogen adaptation responses. A detailed description of these pathways is presented in Chapter 2.

The cAMP-PKA pathway. Using a nonlinear ODE model, Garmendia-Torres *et al.*, 2007 investigated the PKA-mediated regulation of Msn2/4 nucleocytoplasmic oscillations, capturing the essential feedback loops in the PKA activation mechanism [66]. Gonze *et al.*, 2008 proposed a stochastic version of this model [93]. Cazzaniga *et al.*, 2008 investigated the guanine nucleotide exchange reactions in the PKA activation mechanism but failed to capture the short time-scale behavior of cAMP dynamics [94, 95]. Stewart-Ornstein et al., 2017 presented a simplified mechanism to investigate, both computationally and experimentally, the upstream mechanism of cAMP activation [96].

Williamson *et al.*, 2009 created a full model of the PKA pathway, incorporating substantial mechanistic details with respect to alternative PKA activation mechanisms, with a focus on understanding features of cAMP regulation [97]. In order to account for experimental observations not captured by these models, Gonzales *et al.*, 2013 used a simplified mechanism to investigate strain-specific cAMP oscillatory dynamics [39]. All the models described in this paragraph have focused on the mechanism of cAMP regulation, or the regulation of Msn2/4 nucleocytoplasmic oscillations purely in response to glucose abundance or starvation.

The Snf1 pathway. Garcia-Salcedo *et al.*, 2014 systematically explored a series of hypothetical interactions in the Snf1 pathway in order to identify a high confidence network topology that recapitulated experimentally observed phenotypes [98]. More recently Welkenhuysen *et al.*, 2017 [99] have investigated the structure of the Snf1 pathway in the context of glucose starvation.

The TORC1 pathway. Many mathematical models of mTOR signaling have been proposed. Of particular relevance to our work, Vinod et al. 2009 investigated the role of mTOR in metabolic regulations in response to insulin signaling [50]. Sonntag et al. 2012 carried out a systematic investigation of regulatory mechanisms underlying the interactions between AMPK and mTORC2 [100]. Pezze et al. 2016 expanded this model to study its response to amino acids [101] . Thobe et al. 2017 studied the regulation of mTORC2 using logical modeling [102].

Integrated pathway models. Sengupta et al., 2007 studied the steady-state properties of an integrated model of the cAMP-PKA pathway and the MAPK signaling system in regulating the pseudohyphal transition of yeast cells to response to nitrogen starvation [103]. Recently Welkenhuysen et al., 2019 created a Boolean model of the crosstalk between various signaling pathways [104], to show how the network confers robustness to diverse nutrient environments.

3.2 The Standard Component Modeling approach

Dynamical modeling frameworks in biological systems are based on the biochemical details of the system under consideration. Oftentimes, the actual biochemical mechanism might remain elusive. Furthermore, the rate constants of enzyme catalyzed reactions are difficult to measure experimentally, and this poses a major hurdle for model parameterization. Here, we use a general framework that circumvents the above mentioned problems by using generic functions to describe biochemical reactions that occur at three qualitative timescales, slow (eg. gene expression, protein synthesis and degradation), intermediate (eg. post-translational modification) and fast (eg. complex formation or dissociation). This approach is referred to as the standard component modeling strategy [105].

The slow timescale reactions, referred to as *Class I* reactions are represented by a Mass Action Law.

$$\frac{d[\mathrm{A}]}{dt} = k_p - k_d[\mathrm{A}] \tag{3.1}$$

Here, k_p represents the production rate and k_d represents the degradation rate of A.

The intermediate timescale reactions, denoted the *Class II* equations are represented by a generic sigmoidal Soft-Heaviside function.

$$\frac{d[\mathrm{P}]}{dt} = \gamma([\mathrm{P}_T]\mathcal{H}(\sigma,\omega) - [\mathrm{P}]) \tag{3.2}$$

where P_T represents the total amount of species P, and $\mathcal{H}$ is the Soft-Heaviside function, given by

$$\mathcal{H}(\sigma,\omega) = \frac{1}{1+e^{-\sigma\omega}}, \quad \omega = \omega_0 + \sum_{i=1}^{N} w_i[X_i] \tag{3.3}$$

is a generic sigmoidal function used to represent a switch-like biochemical mechanism. Here the ω term is composed of a linear function of the activities of the regulators X_i of P with appropriate signed coefficients ω_i.

In this work, the total amount of all variables (P_T) is assumed to be 1. Thus, variables representing signaling components take values between 0 and 1, and denote the fraction of the given regulator in a given post-translational modification state.

Finally the reactions occurring on fast time scales such as protein complex formations are represented by the *Class III* equations governed by the stoichiometric relationships between the constituents of the protein complexes. Thus, the stoichiometric association of components X and Y to produce a complex C would be represented as

$$C = \min(X, Y) \tag{3.4}$$

3.3 Model Construction

In brief, we manually surveyed the literature related to the nutrient signaling system in yeast, and codified the consensus mechanism that emerged into the regulatory mechanism represented in Figure 2.3. The literature support for the interactions used to construct the molecular regulator network are summarized in Table 3.1.

Table 3.1: Literature support for molecular interactions

Regulator	Target	Notes	Citation
Snf1	Cyr1	phosphorylation	[91]
PKA	Snf1	via Sak1 inhibition	[91]
PKA	Snf1		[92]
PKA	Ras2	via Cdc25 inhibition	[106]
Ras2	Cyr1	phosphorylation	
Glucose	Cyr1	via Gap2, glycolysis	[107]
Sch9	PKA	via bcy1 stimulation?	[74, 75]
Glutamine	Gtr1/2 (EGO)		
Snf1	TORC1	via PASk and Pbp1	[85]
Snf1	TORC1		[87]
TORC1	Sch9	phosphorylation, activation	[68]
PKA	Gis1	phosphorylation, inhibition	[56]
Snf1	Mig1	phosphorylation, inhibition	[99]
TORC1	Gcn2	phosphorylation, inhibition	[65]
Snf1	Gcn2	phosphorylation	[65]
PKA	Pde1/2	phosphorylation, activation	[54]
TORC1	Rtg1/3	phosphorylation	[108]
TORC1	Gln3	phosphorylation	[109]
Glucose	Snf1	via Reg1/2-Glc7	[61]
Glucose/pH	Sch9		[88]
Sch9	eIF2alpha	phosphorylation	[68]
PKA	Gis1	via Rim15	[56]
Sch9	Gis1	via Rim15	[32]
Snf1	Gln3	Direct phosphorylation	[89]

3.4 Kinetic expressions

We expressed the interactions curated in Section 3.3 as biochemical rate laws using the Standard Component Modeling framework described in Section 3.2. The system of equations is subdivided into three subsections for the purpose of presentation. The components involved in sensing and signal transduction are presented first, followed by the master regulatory proteins, followed by the downstream responses which consist of the model readouts. The equations are presented here in this order.

3.4.1 Nutrient signal sensing and transduction

$$\frac{d[\text{Glutamine}]}{dt} = (k_{acc-glu}[\text{Glutamine}_{ext}] + k_{acc-pro}\text{Proline} + k_{acc-nh4}\text{NH}_4[\text{Gln1}]\text{Carbon}) - k_{degr}[\text{Glutamine}]$$

$$\frac{d[\text{Cyr1}]}{dt} = \gamma_{cyr}([\text{Cyr1}_\text{T}]\mathcal{H}(\sigma_{cyr}, \omega_{cyr-glu}\text{Carbon}[\text{Ras}] - \omega_{cyr} - \omega_{cyr-snf}[\text{Snf1}]) - [\text{Cyr1}])$$

$$\frac{d[\text{Ras}]}{dt} = \gamma_{ras}([\text{Ras}_\text{T}]\mathcal{H}(\sigma_{ras}, -\omega_{ras-pka}[\text{PKA}] + \omega_{ras-glu}\text{Carbon} + \omega_{ras)} - [\text{Ras}])$$

$$\frac{d[\text{EGO}]}{dt} = \gamma_{ego}([\text{EGO}_\text{T}]\mathcal{H}(\sigma_{ego}, \omega_{ego-gap}[\text{EGOGAP}]([\text{Glutamine}]_{ext} + 0.5\text{NH}_4 + 0.01\text{Proline}) - \omega_{ego}(1 - [\text{Glutamine}]) - \omega_{ego-basal}) - [\text{EGO}])$$

$$\frac{d[\text{EGOGAP}]}{dt} = \gamma_{gap}([\text{EGOGAP}_\text{T}]\mathcal{H}(\sigma_{gap}, \omega_{gap-N}(1 - [\text{Glutamine}]) - \omega_{gap-torc}[\text{TORC1}]) - [\text{EGOGAP}])$$

$$\frac{d[\text{cAMP}]}{dt} = k_{camp-cyr}[\text{Cyr1}]ATP - k_{camp-pde}[\text{PDE}][\text{cAMP}] - k_{camp-deg}[\text{cAMP}]$$

$$\frac{d[\text{PDE}]}{dt} = \gamma_{pde}([\text{PDE}_\text{T}]\mathcal{H}(\sigma_{pde}, \omega_{pde-pka}[\text{PKA}] - \omega_{pde)} - [\text{PDE}])$$

$$\frac{d[\text{Sak}]}{dt} = [\text{Sak}_\text{T}]\mathcal{H}(\sigma_{sak}, \omega_{sak} - \omega_{sak-pka}[\text{PKA}]) - [\text{Sak}]$$

3.4.2 Master regulators

$$\frac{d[\text{TORC1}]}{dt} = \gamma_{tor}([\text{TORC1}_\text{T}]\mathcal{H}(\sigma_{tor}, \omega_{torc-glut}[\text{Glutamine}] + \omega_{torc-ego}[\text{EGO}] - \omega_{torc-egoin}(1 - [\text{EGO}]) - \omega_{torc} - \omega_{torc-snf}[\text{Snf1}]) - [\text{TORC1}])$$

$$\frac{d[\text{Snf1}]}{dt} = \gamma_{snf}([\text{Snf1}_\text{T}]\mathcal{H}(\sigma_{snf}, -\omega_{snf-glc}\text{Carbon} + \omega_{snf-sak}[\text{Sak}] - \omega_{snf)} - [\text{Snf1}])$$

$$\frac{d[\text{PKA}]}{dt} = \gamma_{pka}([\text{PKA}_\text{T}]\mathcal{H}(\sigma_{pka}, \omega_{pka-camp}[\text{cAMP}] - \omega_{pka} - \omega_{pka-sch9}[\text{Sch9}]) - [\text{PKA}])$$

$$\frac{d[\text{Sch9}]}{dt} = \gamma_{sch9}([\text{Sch9}_\text{T}]\mathcal{H}(\sigma_{sch9}, \omega_{sch9-torc}[\text{TORC1}] - \omega_{sch9)} - [\text{Sch9}])$$

3.4.3 Downstream responses

$$\frac{d[\text{Gcn2}]}{dt} = \gamma_{gcn2}([\text{Gcn2}_\text{T}]\mathcal{H}(\sigma_{gcn2}, \omega_{gcn} - \omega_{gcn-torc}[\text{Sch9}]) - [\text{Gcn2}])$$

$$\frac{d[\text{Gcn4}]}{dt} = \gamma_{gcn4}([\text{Gcn4}_\text{T}]\mathcal{H}(\sigma_{gcn4}, \omega_{gcn4-gcn2-trna}\min([\text{Gcn2}], 74.5([\text{tRNA}_\text{total}] - \min([\text{tRNA}_\text{total}], [\text{Glutamine}]))) - \omega_{gcn4)} - [\text{Gcn4}])$$

$$\frac{d[\text{eIF}]}{dt} = \gamma_{eif}([\text{eIF}_\text{T}]\mathcal{H}(\sigma_{eif}, \omega_{eif} - \omega_{eif-gcn2}[\text{Gcn2}]) - [\text{eIF}])$$

$$\frac{d[\text{Gln3}]}{dt} = \gamma_{gln3}([\text{Gln3}_\text{T}]\mathcal{H}(\sigma_{gln}, -\omega_{gln3} + \omega_{gln-snf}[\text{Snf1}] + \omega_{gln-sit}(1 - [\text{TORC1}])) - [\text{Gln3}])$$

$$\frac{d[\text{Gln1}]}{dt} = \gamma_{gln1}([\text{Gln1}_\text{T}]\mathcal{H}(\sigma_{gln1}, \omega_{gln1-gln3}[\text{Gln3}] - \omega_{gln1)} - [\text{Gln1}])$$

$$\frac{d[\text{Rtg13}]}{dt} = \gamma_{rtg13}([\text{Rtg13}_\text{T}]\mathcal{H}(\sigma_{rtg}, -\omega_{rtg-torc}[\text{TORC1}] + \omega_{rtg)} - [\text{Rtg13}])$$

$$\frac{d[\text{Gis1}]}{dt} = \gamma_{gis1}([\text{Gis1}_\text{T}]\mathcal{H}(\sigma_{gis1}, -\omega_{gis-pka}[\text{PKA}] - \omega_{gis-sch}[\text{Sch9}] + \omega_{gis)} - [\text{Gis1}])$$

$$\frac{d[\text{Mig1}]}{dt} = \gamma_{mig}([\text{Mig1}_\text{T}]\mathcal{H}(\sigma_{mig1}, \omega_{mig-pka}[\text{PKA}] - \omega_{mig-snf}[\text{Snf1}] + \omega_{mig)} - [\text{Mig1}])$$

$$\frac{d[\text{Dot6}]}{dt} = \gamma_{dot6}([\text{Dot6}_\text{T}]\mathcal{H}(\sigma_{dot}, -\omega_{dot-sch-pka}[\text{Sch9}][\text{PKA}] + \omega_{dot)} - [\text{Dot6}])$$

$$\frac{d[\text{Tps1}]}{dt} = \gamma_{tps}([\text{Tps1}_\text{T}]\mathcal{H}(\sigma_{tps}, \omega_{tps-pka}([\text{PKA}_\text{T}] - [\text{PKA}]) - \omega_{tps)} - [\text{Tps1}])$$

$$\frac{d[\text{Trehalase}]}{dt} = \gamma_{tre}([\text{Trehalase}_\text{T}]\mathcal{H}(\sigma_{trehalase}, \omega_{tre-pka}[\text{PKA}] - \omega_{tre)} - [\text{Trehalase}])$$

$$\frac{d[\text{Protein}]}{dt} = k_{pr}[\text{ATP}]\min(\min([\text{Rib}], [\text{eIF}]), \min([\text{tRNA}_\text{total}], [\text{Glutamine}]))[\text{Protein}]$$

$$\frac{d[\text{Rib}]}{dt} = k_{\text{transcription}}(1 - [\text{Dot6}]) - k_{mRNA-degr}[\text{Rib}]$$

Heuristics and assumptions

While the equations related to the regulatory interactions reflect the mechanism curated from the literature, such a representation is outside the scope of the model for some components. The heuristics used to represent such model components are summarized here.

1. **Storage molecules** In yeast, amino acids are stored both in the cytosol and the lysosome-like vacuole[15, 110, 111, 112]. Since regulation of the these distinct, dynamic nutrient pools is complex, and since TORC1 has been shown to detect amino acid levels by detecting glutamine concentrations [113], a single variable representing

glutamine levels are used to represent the intracellular nitrogen sufficiency, depicted as AA in Figure 2.3. Currently, this is modeled as a mass action law proportional to the extracellular concentrations of Glutamine ($\text{Glutamine}_{\text{ext}}$), ammonia ($NH_4$), and proline. The consumption rate of gutamine currently is not proportional to growth rate in this model.

$$\begin{aligned}\frac{d[\text{Glutamine}]}{dt} =&(k_{acc\_glu}[\text{Glutamine}_{ext}] + k_{acc\_nh4}[\text{NH4}][\text{Rtg13}][\text{Carbon}] \\ &+ k_{acc\_pro}[\text{Proline}]) - k_{degr}[\text{Glutamine}]\end{aligned}$$

For maximal growth rate, the cell requires a non limiting amount of ATP. Instead of modeling the dynamic regulation of the cellular energy charge (EC), we assume that the EC is proportional to the carbon sufficiency of the environment while acknowledging that the observed behavior is highly non-trivial [114]. The cellular EC is represented by the parameter ATP, which is fixed to the value of the carbon input for a given simulation. Finally, during glucose starvation, trehalose is synthesized as the carbon storage molecule. During relief from starvation, the stored trehalose is rapidly degraded to glucose by trehalase[115].

2. **Nitrogen sensing** The precise mechanism of nitrogen sensing in yeast by TORC1 remains controversial. The TOR complex has been shown to interact with multiple G-proteins at the vacuolar membrane which sense the levels of distinct intracellular amino acids [23, 67] . The EGO complex comprising of the G-proteins Gtr1/2 with the membrane anchors Ego1/2/3 (represented as EGO), along with their GTPase activating proteins Lst4/7 (represented as EGOGAP) ultimately activate the TORC1 complex. This process occurs at the vacuolar membrane, represented in the dashed box in Figure 2.3 (EGOGAP is not represented in the figure). The activation of the EGO by the various classes of nitrogen sources is represented as follows

$$\begin{aligned}\frac{d[\text{EGO}]}{dt} =& \gamma_{ego}([\text{EGO}_T]\mathcal{H}(\sigma_{ego}, \omega_{ego\_gap}[\text{EGOGAP}]([\text{Glutamine}_{ext}] + 0.5[\text{NH}_4] + 0.01[\text{Proline}]) \\ &- \omega_{ego}(1 - \text{Glutamine}) - \omega_{ego\_basal}) - \text{EGO})\end{aligned}$$

3. **PKA activation** Multiple published models have explored the complexity of Protein Kinase A activation via the Ras-cAMP mechanism with a focus on cAMP dynamics. In this model, we attempt to include the minimum number of interactions required to explain the cAMP dynamics, and shift our attention to the interactions between the PKA pathway and the rest of the nutrient signaling system. Glucose has been shown to stimulate Adenylate Cyclase (Cyr1) activity both directly [21, 116] and via Ras2[117]. Cyr1 catalyzes the formation of cAMP from ATP, which binds to the regulatory subunit Bcy1, thus activating the catalytic subunits Tpk1/2/3 of PKA, here represented as a single species PKA. PKA activates the phosphodiesterases Pde1/2 which have been shown to affect cAMP dynamics to different degrees, while it is represented as a single variable in our model.

4. **The Snf1 pathway** While the exact mechanism remains unclear, it is known that glucose-dependent phosphatases Reg1-Glc7 inhibit Snf1 activity. Snf1 activity is stimulated by the Sak1/Elm3/Tos1 kinases [98]. In our model, we simplify the upstream mechanism of glucose dependent Snf1 repression, and focus on the Sak1 mediated interactions with the rest of the pathway.

$$\frac{d[\text{Snf1}]}{dt} = \gamma_{snf}([\text{Snf1}_T]\mathcal{H}(\sigma_{snf}, -\omega_{snf\_glc}[\text{Carbon}] + \omega_{snf\_sak}[\text{Sak}]) - [\text{Snf1}])$$

5. **Crosstalk between pathways** By examining the available time series data, we have retained controversial interactions depending on the ability to qualitatively demonstrate experimentally observed dynamics. PKA and Snf1 have been shown to mutually inhibit each other; Snf1 has been shown to inhibit Cyr1 activity, effectively inhibiting PKA activation, while PKA has been shown to inhibit Sak1 kinase activity, inhibiting Snf1 activation [91, 92]. The relationship between Snf1 and TORC1 is less clear. While glucose dependent TORC1/Sch9 phenotypes have been reported in multiple publications, the mechanism by which TORC1 activity is modulated by the presence of glucose remains unclear [68]. Based on data from Hughes-Hallett et al., 2015 and Prouteau et al., 2018, we assume a direct negative interaction with Snf1 and TORC1 [87, 88]. Finally, the relationship between Sch9 and PKA is confounded by the complex and overlapping regulation of their many shared downstream targets [32]. Early studies on the PKA pathway found an apparent negative interaction between Sch9 and PKA [73], however, the exact relationship between the related kinases remains unclear. Zhang et al. claim that Sch9 interacts with, and inhibits PKA at multiple levels [74, 75], while an older report implicates the role of a heat shock protein Sse1 in the regulation of PKA activity by Sch9 [33]. Here, we assume the Bcy1 mechanism proposed by Zhang et al. 2011, and use a direct inhibition of PKA by Sch9.

3.5 Model Calibration

We curated from the literature experimental measurements relevant to the molecular species and nutrient conditions in our model. Typical experiments investigating the nutrient signaling system, characterize the cellular phenotype at the end of a long term nutrient starvation or drug treatment experiment, typically on the timescales of hours to days. This type of long term growth data is not useful for a model studying the transient, short term molecular responses after an environmental perturbation. We collected two types of short term experimental measurements in order to calibrate model behavior, (1) time series measurements, and (2) nutrient shift-dependent relative steady state levels. The former are predominantly available for the molecular regulators, while the latter are predominantly available for transcription factors.

Time series Table B.1 lists the types, sources, and reported values of time series data obtained from the literature. The methodology used to obtain the values was as follows: a plot of the time series data of interest was obtained from the figure made available on the publisher's website, or was extracted out of the article document. Next, the WebPlotDigitizer tool was used to extract the time points from the figure [118]. These were stored in the form of tab separated files. The data thus obtained has the units specific to the relevant experiment performed, and is not directly usable in our model. Therefore, the data was min-max normalized between 0 and 1, since most model variables take values between 0 and 1. The model fit to data are presented in Figure 3.1(a)-(j).

Perturbation Table B.2 list the types, sources and reported values of steady state activities of molecular species reported in the literature. The method of curation was similar to that of the time series data, except ImageJ was used to extract relative expression from blots [119]. The model fit to the curated perturbation data are presented in Figure 3.1(l).

3.5.1 Defining a goodness-of-fit cost function

In order to fit the time dynamics to available data from the literature, parameters were first hand-tuned to obtain qualitative fits to the expected time course behavior, and were then fine tuned using Levenberg-Marquardt Least Squares Optimization. The `leastsq` function in the Scipy library was used for least squares fitting. We observed that the least squares method does not scale well for a problem of this size, on the order of 125 parameters. For this, a quadratic cost function was constructed to measure the goodness of fit to the time series and the perturbation data

$$C(\mathbf{p}) = \frac{1}{N}\sum_{i}^{N}(y_i^t(\mathbf{p}) - x_i^t)^2 + \frac{1}{M}\sum_{i}^{M}(y_i^p(\mathbf{p}) - x_i^p)^2$$

where N is the number of $t$impoints and M the number of $p$erturbations. $\mathbf{p}$ is the candidate parameter vector, $y(\mathbf{p})$ is the model prediction and x is the literature-derived activity of the variable under consideration. An custom MCMC sampling scheme was carried out starting from the hand tuned parameter set, and the parameter set $\mathbf{p}^*$ associated with the least cost $C^{\min}$ was chosen. This parameter set was used as the starting point for further analysis. In order to better understand the structure of the parameter space, we then performed a global model robustness analysis. The details of the method are presented in Section 4.1.

3.5.2 Comparing simulation results with experimental time-course data.

In Figure 3.1 (a)–(j), we compare the experimental data (open circles) to time-series simulations (in red) of the model using the optimal parameter values.

Our model includes molecular species that are common readouts in experiments investigating nutrient signaling. cAMP is a standard readout for the activity of the PKA pathway. cAMP shows rapid transient changes, on the time scale of minutes, upon relief from glucose starvation. Model simulations successfully capture the cAMP dynamics in *wt* cells as well as *sch9*Δ mutant cells (Figure 3.1(a), (b)). Sch9 phosphorylation is a widely accepted readout for TORC1 activity and displays changes on the time scale of tens of minutes upon relief from nitrogen starvation [113]. Model simulations successfully capture the transient as well as steady state behavior of Sch9 phosphorylation in *wt* and in *gtr1,2*Δ strains (Figure 3.1 (c), (d), and (e)). Moreover, we are able to capture changes in Sch9 phosphorylation in response to glucose starvation and readdition (Figure 3.1(f), (g)), as reported by Prouteau et al. [88]. Snf1 kinase activity is typically measured indirectly, by examining the phosphorylation of the SAMS peptide, a Snf1 target sequence [120]. The model successfully captures the activation of Snf1 during glucose downshift (Figure 3.1(h)) leading to the expression of glucose-repressed genes. Apart from assays of biochemical activities, subcellular localization measured by fluorescence microscopy provides indirect information about the activity of transcription factors. Mig1 nuclear localization upon relief from glucose starvation was recently reported by Welkenhuysen et al. [104] (Figure 3.1(i)). Model simulations successfully capture this response, if we identify Mig1 transcriptional activity with the experimentally observed translocation of Mig1 into the nucleus. Direct information about transcription factor activity can also be obtained by measuring mRNA levels of the transcription factor's target genes. For example, *RPL*32 mRNA, encoding the large ribosomal subunit, is downregulated in response to rapamycin treatment, and our model captures this effect nicely (Figure 3.1(j)).

3.5.3 Capturing phenotypes of *wt* and mutant strains in response to nutrient shifts.

To investigate the roles of regulators in nutrient signaling, experimentalists typically characterize the phenotype of a strain in response to perturbations (nutrient shifts or drug treatments). In order to parameterize our model we collected data from direct biochemical assays of molecular responses to perturbations (see Supplementary Table 3).

An example of the type of perturbation data used to calibrate our model is illustrated in Figure 3.1(k), where model simulations are compared to data reported in Beck et al., 1999 [109], who studied the phosphorylation of Gln3 by TORC1 and Sit4. Using Western blots of Gln3 pull-downs, Beck et al. found that Gln3 is phosphorylated (inactive) in untreated

wt cells (black dashed line), while treatment with rapamycin led to Gln3 dephosphorylation. In contrast, Gln3 remained phosphorylated (inactive) in a *sit4*Δ strain even after rapamycin treatment (red dashed line). The figure shows simulated time courses of active (dephosphorylated) Gln3 under these conditions, and the markers (empty circles and empty triangles) indicate the (pre- and post-shift) steady-state values of active Gln3 used to compare against the experimental measurements.

Figure 3.1(l) visualizes such comparisons for a representative sample of experimental perturbations. For example, the results in Figure 3.1(k) are summarized in the first line of Figure 3.1(l). The second line shows the results of 3-amino-1,2,4-triazole (3AT) treatment on Gcn4 activity in *wt* cells (black) and in a *gcn2* deletion strain (red). 3AT inhibits histidine synthesis, resulting in induction of Gcn4 and an amino acid starvation response. There is no induction of Gcn4 expression in a *gcn2*Δ strain.

The second and third blocks of Figure 3.1(l) show the results of 'nitrogen shifts' and 'carbon shifts'; either a downshift (nutrient starvation) or upshift (relief from starvation). The first two rows in the 'Nitrogen Shift' block (call them N1 and N2) repeat the results in panels (c) and (e), which report the phosphorylation of Sch9 in *wt* and *gtr1*Δ*gtr2*Δ strains in response to relief from starvation (i.e., addition of glutamine to nitrogen-starved cells). Row N1 compares the pre-shift steady state (at time 0) to the post-shift steady state (at time 30 min). Row N2 compares the pre-shift steady state to the transient state (at time 4 min). Row N3 makes a similar comparison of *wt* cells to *lst4*Δ*lst7*Δ cells. Row N4 compares expression levels of RPL25 (a ribosomal protein), as measured by [73] in *wt* and *sch9*Δ strains, with model simulations.

In the 'Carbon Shift' block, the first four rows report the results of readdition of glucose to carbon-starved cells, while the last row summarizes a glucose starvation experiment. Row C1, left column, repeats the results in panel (a). The *pde1*Δ strain shows an increase in cAMP levels compared to *wt* [54][1], while the *ras1*Δ*ras2*Δ*bcy1*Δ strain shows a diminished amount of cAMP compared to *wt* [53]. Row C3 depicts trehalase (Nth1) levels after a glucose up-shift; trehalase levels remain low in a *tpk1*Δ*tpk2*Δ mutant [20]. Row C4 demonstrates that a mutation in the nitrogen signaling pathway (*sch9*Δ) has a significant effect on the carbon stress response factor Gis1 [73]. Finally, row C5 shows that Snf1 activity increases in response to a carbon down-shift in *wt* cells, but this response is absent in *sak1*Δ*tos3*Δ*elm1*Δ cells [63].

[1] Ma et al., 1999 showed that the *pde1*Δ*pde2*Δ double mutant shows a lower cAMP level than wildtype [54]. Gonzales et al., 2013 showed that a competitive inhibition model between Pde1, Pde2 and Ira1/2 explains this phenotype. However, our model does not explain this behavior since we do not explicitly model the two Pde homologs as being distinct.

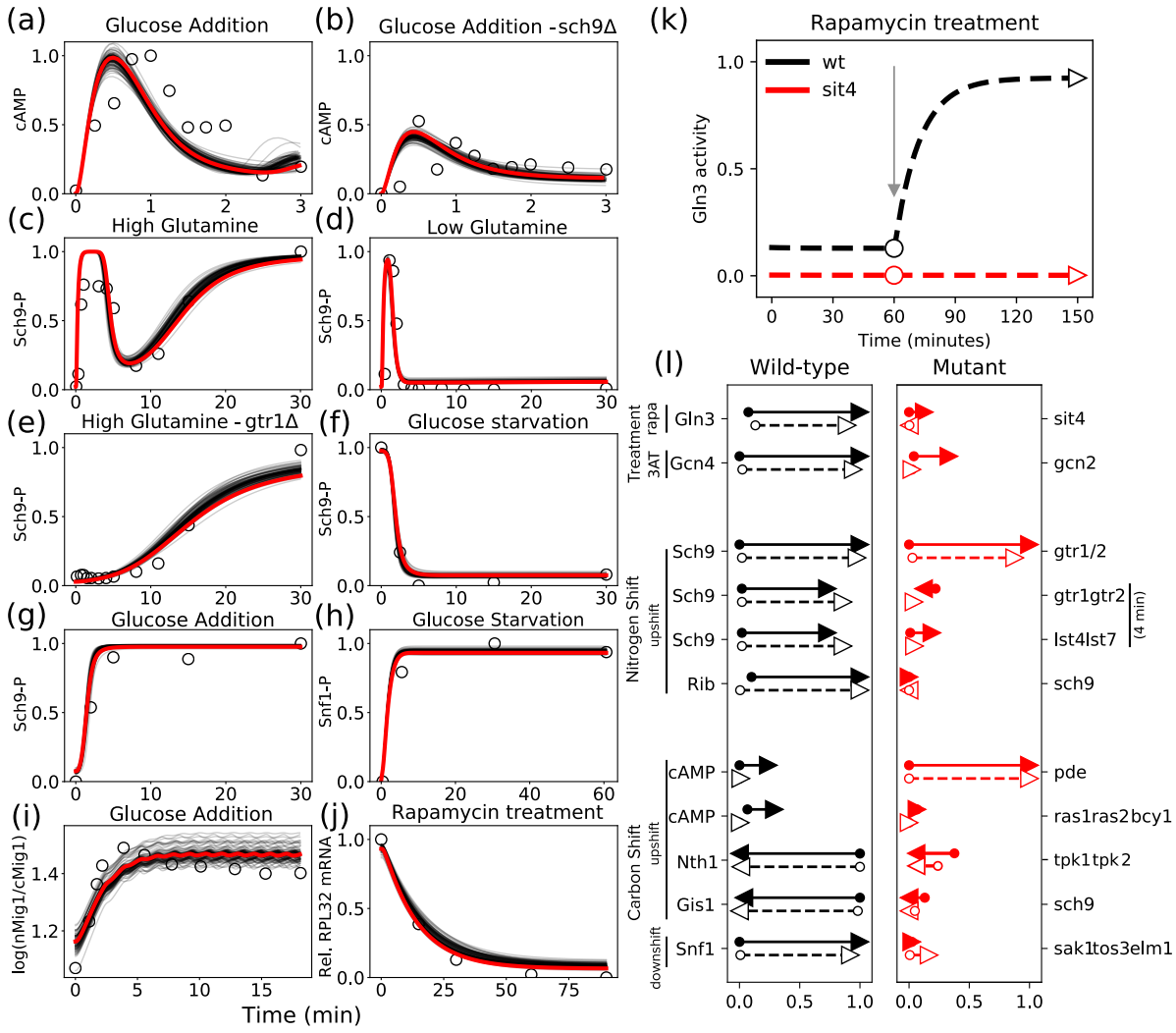


Figure 3.1: The model successfully explains experimental data. (a)-(j) Time series fits to literature-curated time-course data. The red lines represent simulated trajectories from the best-fit parameter set, while the gray lines represent simulations from 100 parameter sets with comparable sums of squared errors. The experimental measurements are shown as open circles. (a) and (b): relief from glucose starvation. (c), (d) and (e): relief from nitrogen starvation by glutamine addition. (f) and (h): glucose starvation. (g) and (i): relief from glucose starvation. (j) Rapamycin treatment of well-nourished cells. (k) Comparing model simulations with data from a shift experiment [109] measuring Gln3 phosphorylation in response to rapamycin treatment in well-nourished *wt* and *sit4*Δ strains. In the *wt* simulation (black dashed line), we calculate the steady state value of Gln3 in rich medium in the absence of rapamycin (black circle). Next, rapamycin is introduced (at the gray arrow), and the new, postshift steady state is recorded (black triangle). The same simulation was repeated for *sit4*Δ cells to yield preshift (red circle) and postshift (red triangle) steady states. *(next page)*

Figure 3.1: *(continued)* (l) Visualization of perturbation analyses. The perturbations are drug treatments or nutrient shifts carried out in *wt* cells (left column, black arrows) and mutant strains (right column, red arrows). The labels on the left indicate the molecular species being assayed, and the labels on the far right indicate the gene(s) deleted in the mutant strains. Experimental results are represented by solid arrows, and model simulations by dashed arrows. For each molecular species under consideration, the arrow's tail and head indicate (respectively) the pre- and post-perturbation steady-state values, as measured on the relative scale (0 to 1) at the bottom of the column. As an example, the rapamycin treatment in panel (k) is reproduced in the first line of panel (l). Note that the y-axis in panel (k) is now the x-axis in panel (l).

Chapter 4

Model Robustness analyis

Jalihal, A.P., Kraikivski, P., Murali, T.M., Tyson, J.J., 2020. Modeling and analysis of macronutrient signaling in budding yeast. *Under review.* Also available on *bioRxiv* (https://doi.org/10.1101/2020.02.15.950881)

In the previous chapter, quantitative data obtained from the literature was used to estimate the numerical values of the parameters in the model. Since the available data is sparse and the number of parameters is large, we expect that many parameter values will be poorly constrained by the data. Since any particular set of parameter estimates will be unreliable in this situation, we would like to obtain a representative ensemble of sets of parameter values that all give acceptable fits to the data. One approach to obtaining such an ensemble would be to randomly sample parameter values, evaluate the model for goodness-of-fit, and accept those parameter sets that meet some standard of acceptable fit. However, sampling parameter values in a completely random fashion is not an efficient way to find alternate parameter sets because small random perturbations will not effectively explore the parameter space, whereas large random perturbations are not likely to find an acceptable set of parameter values. Instead, we use the fact that the region of acceptable parameter values in a high-dimensional parameter space is characterized by 'stiff' and 'sloppy' directions [121], and a random parameter search has to respect these directions. In order to carry out this search we used a method of 'model robustness analysis' described by Tavassoly et al. [122].

In this chapter we wish to investigate the dependence of our proposed model on the model parameterization. After going over the mathematical preliminaries, we present the results of a systematic investigation of the goodness-of-fit cost surface. We outline an approach to constructing an approximation of the curvature of the cost surface. We use this approximation (the Hessian matrix) to generate alternate sets of parameter values with comparable fit to the data. We finally explore the properties of the constructed Hessians, as well as the the ensemble of parameter sets.

4.1 Robustness analysis

This section describes the details of the parameter robustness analysis to investigate the dependence of model dynamics on model parameterization. The motivation for this analysis is the concept of model sloppiness developed by Gutenkunst et al., 2007 [121] who inves-

tigated the characteristics of a cost function quantifying the deviation of model predictions from a set of experimental data. The value of the cost function depends on the values of the parameters in the model. In their investigations Gutenkunst et al. found that the cost as a function of model parameters typically exhibits a 'stiff/sloppy' structure, i.e., while one can identify a few parameter combinations that tightly constrain the cost function, a majority of parameter combinations do not significantly constrain the cost function. Here, we are interested in the so-called stiff parameter directions, which are highly constrained by data. We carry out this investigation in the following stages:

1. We define a cost function that measures the fit between the experimental data and model predictions, described in Section 4.1.1.

2. We use the cost function to improve the global fit of the model to the data using an MCMC sampling strategy described in Section 4.1.2.

3. We approximate the structure of this cost function in parameter space by computing the Hessian of this surface around the optimal parameter set using the method described in [122]. Details of this method are provided in Section 4.1.3.

4. We generate a sample of parameter sets constrained by the eigenvectors of the Hessian matrix, and iteratively refine the Hessian, as described in Sections 4.1.4 and 4.1.5.

5. Finally, we study the properties of this refined Hessian to identify the stiff and sloppy parameter directions in our model, given the set of curated experimental data used to constrain the model. This is described in Sections 4.1.6 to 4.1.9.

4.1.1 The goodness-of-fit cost function

This quadratic cost function $C(\mathbf{p})$, which includes both the time series data and the steady-state perturbation data as described in 'Results', is defined as follows:

$$C(\mathbf{p}) = \frac{1}{N}\sum_{i}^{N}(y_i^t(\mathbf{p}) - x_i^t)^2 + \frac{1}{M}\sum_{i}^{M}(y_i^u(\mathbf{p}) - x_i^u)^2, \tag{4.1}$$

where N is the number of time points (t) and M the number of perturbations (u), $\mathbf{p}$ is the candidate parameter vector, $y(\mathbf{p})$ is the model prediction and x is the literature-derived activity of the variable under consideration. While the number of time points exceed the number of perturbation data points, we do not preferentially weight one type of data over the other.

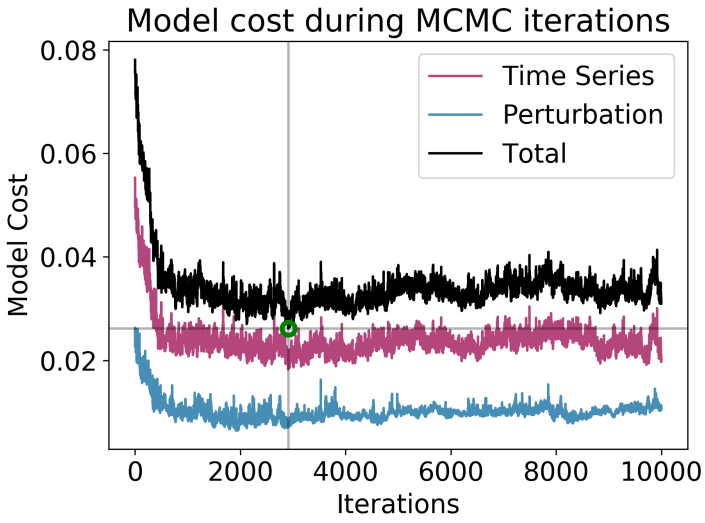


Figure 4.1: Results of MCMC sampling. The black line shows the total model cost. The contributions from the time course term and from the perturbation term are shown in pink and blue, respectively. The lowest cost, designated by the green circle at iteration 2900, defines the 'optimal' parameter set.

4.1.2 MCMC sampling to improve estimate of parameter values

Having defined the quadratic cost function, we used a Markov Chain Monte Carlo (MCMC) sampling strategy to improve the fit to the data. Briefly, in every MCMC iteration, the last accepted parameter set is perturbed as follows: for each parameter with value p in the last accepted set, a new value p' is sampled from a normal distribution $\mathcal{N}(\mu = p, \sigma = 0.025p)$. The cost function is evaluated for this new parameter set and this set is accepted with probability $e^{-\beta\Delta C}$, where $\Delta C = C(p') - C(p)$. We chose $\beta = 3.6$ based on the magnitude of the change in cost that we observed in each iteration. Starting from the hand-tuned parameter set, we repeat MCMC sampling 10,000 times. The change in cost across the iterations is shown in Figure 4.1.

4.1.3 Defining a Hessian approximation based on a sample of parameter sets

The proofs presented in this section are due to Dr. William Baumann. These results are the basis for the analysis published in Tavassoly et al., 2015, and are presented here for completeness.

In this section, we derive the expressions used to compute the Hessian.

Let $\mathbf{p}^*$ denote the parameter set that minimizes $C(\mathbf{p})$, i.e., $C(\mathbf{p}^*) = C_{\min}$, then

$$\nabla C(\mathbf{p})|_{\mathbf{p}^*} = 0.$$

Thus, for every $\mathbf{p}$ in the neighborhood of $\mathbf{p}^*$, we can carry out a Taylor series expansion around $\mathbf{p}^*$. Omitting the higher order terms:

$$C(\mathbf{p}) \approx C(\mathbf{p}^*) + (\mathbf{p} - \mathbf{p}^*)^T H(\mathbf{p} - \mathbf{p}^*) = C(\mathbf{p}^*) + \Delta\mathbf{p}^T H \Delta\mathbf{p}, \tag{4.2}$$

where $2H = \nabla^2 C$ is the Hessian of the cost function.

Next, we define some notation as introduced in Magnus and Neudecker [123]. For an $m \times n$ matrix A, the vectorization operation $\text{vec}(A)$ results in a $mn \times 1$ column vector that stacks the columns of A. If A is an $n \times n$ symmetric matrix, then the operation $\text{vech}(A)$ stacks the lower triangular columns, yielding an $\frac{n(n+1)}{2} \times 1$ column vector. There exists a unique matrix D, called the duplicator matrix, with dimensions $n \times \frac{n(n+1)}{2}$, and a unique matrix L called the eliminator matrix with dimensions $\frac{n(n+1)}{2} \times n^2$, such that

$$\begin{aligned} \text{vec}(A) &= D\,\text{vech}(A) \\ \text{vech}(A) &= L\,\text{vec}(A) \end{aligned}$$

The Kronecker product (denoted by $\otimes$) of an $m \times n$ matrix A and an $s \times t$ matrix B is an $mn \times st$ matrix

$$A \otimes B = (a_{ij})B,$$

where a_{ij} is the ij^{th} entry of A. For any three matrices A, B and C the following holds true

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$$

Suppose we choose S parameter sets $\mathbf{p}_i, 1 \leq i \leq S$ in the neighborhood of $\mathbf{p}^*$. We can construct a quadratic error function E_H which will be minimized when H approximates the true Hessian of the function.

$$E_H = \frac{1}{2}\sum_{i=1}^{S}\left(C(\mathbf{p}_i) - C(\mathbf{p}^*) - (\mathbf{p}_i - \mathbf{p}^*)^T H(\mathbf{p}_i - \mathbf{p}^*)\right)^2$$

Using $\Delta C_i = C(\mathbf{p}_i) - C(\mathbf{p}^*)$ and $\Delta\mathbf{p}_i = (\mathbf{p}_i - \mathbf{p}^*)$, we have

$$\frac{\partial E_H}{\partial H} = 0 = \sum_i^S (\Delta C_i - \Delta\mathbf{p}_i^T H \Delta\mathbf{p}_i)\Delta\mathbf{p}_i \Delta\mathbf{p}_i^T \tag{4.3}$$

We wish to solve for H given a set of parameter vectors $\mathbf{p}$. For a model with k parameters, we will have to solve for k^2 terms in H. However, we can decrease the size of this problem

by considering the fact that the Hessian should be a symmetric matrix. Thus, we need to solve only for the terms in the lower triangle.

Simplifying Equation (4.3), we have

$$\begin{aligned}
\sum_i^S \Delta C_i \Delta \mathbf{p}_i \Delta \mathbf{p}_i^T &= \sum_i^S \Delta \mathbf{p}_i^T H \Delta \mathbf{p}_i \Delta \mathbf{p}_i \Delta \mathbf{p}_i^T \\
\sum_i^S \Delta C_i \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) &= \sum_i^S (\Delta \mathbf{p}_i^T H \Delta \mathbf{p}_i) \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \\
&= \sum_i^S (\Delta \mathbf{p}_i^T \otimes \Delta \mathbf{p}_i) \text{vec}(H) \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \\
&= \sum_i^S \text{vec}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T)^T \text{vec}(H) \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \\
&= \sum_i^S \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) D^T D \text{vech(H)} \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \\
&= \sum_i^S \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T)^T D^T D \text{vech}(H) \\
&= \sum_i^S Q D^T D \text{vech}(H) \quad \text{where } Q = \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T) \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T)^T
\end{aligned}$$

If we define $R = \sum_i \Delta C_i \text{vech}(\Delta \mathbf{p}_i \Delta \mathbf{p}_i^T)$, we have that

$$R = Q D^T D \text{vech}(H)$$

In other words,

$$\text{vech}(H) = (Q D^T D)^{-1} R$$

and

$$\text{vec}(H) = D(Q D^T D)^{-1} R \tag{4.4}$$

Here, R is symmetric and D is the appropriate duplicator matrix.

4.1.4 Sampling new parameter sets constrained by the approximate Hessian

Using the approximate Hessian matrix computed as described in Equation (4.4), we next wish to use the eigenvectors of this matrix to constrain the search for new parameter vectors. Intuitively we wish to avoid the eigenvector directions corresponding to 'large' eigenvalues which are the stiff directions. We know that the dimensions of the cost ellipsoids are proportional to $\frac{1}{\lambda^{1/2}}$ where λ is an eigenvalue of the Hessian. Thus we can weight the eigenvectors

by a factor of $\frac{1}{\lambda^{1/2}}$, which favors the sloppiest eigenvector directions. In practice we want a candidate parameter vector to respect all the stiff and sloppy directions. We first generate a random vector α, which we then transform to respect the stiff and sloppy directions. We describe the transformation matrix below.

We start by translating the frame of reference our system to $\mathbf{p}^*$ so that we can generate vectors $\Delta\mathbf{p} = \mathbf{p} - \mathbf{p}^*$ which produce a relative increase in model cost $\Delta C = C - C_{\min}$. Rewriting Equation (4.2),

$$\Delta C(\mathbf{p}) \approx \Delta\mathbf{p}^T H \Delta\mathbf{p}$$

To avoid an ill-conditioned Hessian, where the eigenvalues span many orders of magnitude, we choose $\Delta l_p^T H \Delta l_p = \Delta C$ where $\Delta l_p = \log(\mathbf{p}) - \log(\mathbf{p}^*)$, with the logarithm of a vector being taken element-wise.

To sample from this new ellipsoid, we first create a random vector α which will lie inside the cost ellipsoid. we sample a vector $\tilde{\alpha}$ of random numbers drawn from $\mathcal{N}(0,1)$. Next, we compute α

$$\alpha = \frac{\tilde{\alpha}}{\sqrt{\tilde{\alpha}^T\tilde{\alpha}}} u\sqrt{\epsilon},$$

where u is a scalar drawn from the uniform distribution on [0,1] and $\epsilon = 2C_{min} = 2 \times 0.026$ is the maximum value of ΔC we wish to consider. The factor $\sqrt{\epsilon}$ scales the magnitude of the unit vector $\frac{\tilde{\alpha}}{\sqrt{\tilde{\alpha}^T\tilde{\alpha}}}$ to the edge of the ellipsoid. The final factor u ensures that the vector lies inside the ellipsoid, as we want to sample the volume, not just the surface of the ellipsoid.

The final step needed to generate a candidate parameter vector is to transform it to respect the stiff and sloppy directions. Using an eigenvalue decomposition, we write $H = V\Lambda V^T$ (where V is the eigenvalue matrix and Λ is the eigenvector matrix). (Note that we compute the absolute values of the eigenvalues and replace every eigenvalue that is less than 0.1 by 0.1. This step ensures that H is positive definite, but limits the length of the longest ellipsoid axes.) Our constrained parameter vector should take the form $\Delta l_p = V\Lambda^{-1/2}\alpha$. This will satisfy the cost constraint as follows

$$\Delta l_p^T H \Delta l_p = \alpha^T \Lambda^{-1/2} V^T V \Lambda V^T V \Lambda^{-1/2} \alpha = \alpha^T\alpha \leq \epsilon$$

Reordering terms, we have

$$\mathbf{p} = e^{\log \mathbf{p}^* + V\Lambda^{1/2}\alpha} = \mathbf{p}^* \underbrace{.*}_{\text{elem-wise multiplication}} \overbrace{e^{V\Lambda^{-1/2}\alpha}}^{\text{elem-wise exponentiation}} \tag{4.5}$$

4.1.5 Iterative refinement of the approximate Hessian

The following steps describe our methodology to initialize a approximation to the Hessian and then refine this approximation iteratively.

1. To compute an initial approximation of the Hessian, we used Latin Hypercube Sampling around 2.5%ranges around $\mathbf{p}^*$ in order to sample parameter vectors close to $\mathbf{p}^*$.

We evaluated the cost for each parameter set. For each range explored, we recorded the parameter sets with cost less than $2C_{\min}$.

2. Next, using Equation (4.4), an approximate Hessian was constructed using the accepted parameter sets from the Latin Hypercube sample. This step was designated as iteration 0.

3. Using the approximate Hessian generated in the previous step, 30,000 parameter sets were generated using Equation (4.5). The goodness-of-fit cost was evaluated for each parameter set and any parameter set satisfying a cost cutoff of $3C_{\min}$ was accepted.

4. Using these accepted parameter sets, the approximate Hessian was recomputed, a new ensemble of 30,000 parameter sets was generated, and the cost evaluation and parameter set acceptance procedure was repeated, while continually expanding the ensemble of accepted parameter sets.

5. The previous step was repeated four times.

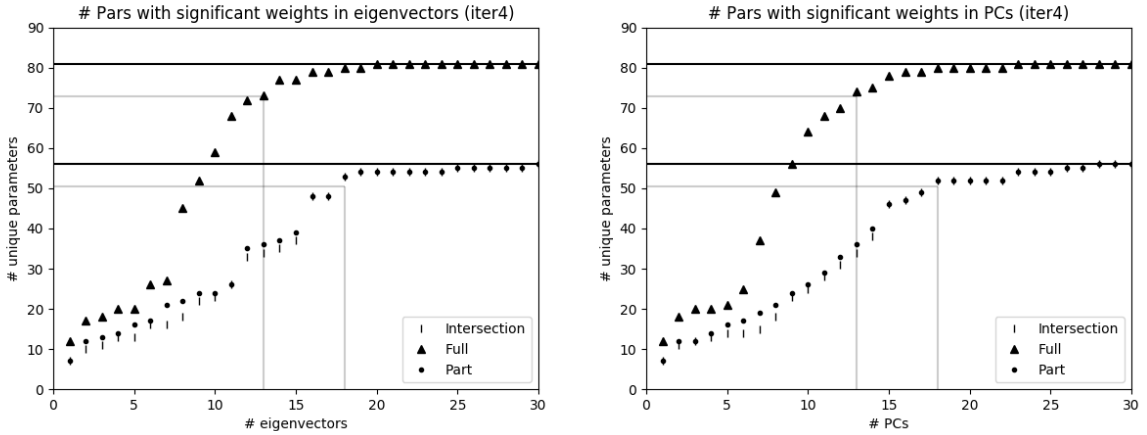
The above scheme was repeated on two different lists of parameters. In our parameter search we fix the values of the total amounts of protein, the P_T parameters to 1.0 since we currently do not have accurate abundances of the regulators in the model, accounting for 21 parameters. We also set the *sigma* parameters in each *Class II* equation to their nominal values presented in Table A.1, accounting for another 21 parameters. Lastly, 5 parameters serve as model inputs, namely ATP, Carbon, $\text{Glutamine}_{\text{ext}}$, Ammonia, and Proline. Thus we fix the values of 49 of the 128 kinetic parameters. The remaining parameters that were varied in this analysis constitute the Full set, consisting of 81 parameters. At the end of the four iterations, we obtained a total of 24,066 parameter sets in the Full set.By virtue of using the soft-Heaviside functions to represent the regulatory interactions, the switch between activation and inactivation of a target depends on the ratio of parameters governing the strengths of inputs from the regulators. We wanted to investigate the effect of fixing one ω parameter per soft-Heaviside expression, essentially exploring the range of acceptable ratios of parameters. This resulted in a second, truncated list with 56 parameters, constituting the Partial set. At the end of the 4 iterations, the size of the ensemble for the Partial set was 37,116.

4.1.6 Fewer than 16% of the eigenvectors are required to capture all the stiff directions

We studied the eigenvector of the refined-approximate Hessian H in order to identify the parameters that contribute to the stiff directions. We deemed a parameter as making a substantial contribution to an eigenvector if the absolute value of its 'loading' (i.e., its weight in the eigenvector) was greater than one standard deviation of all the loadings across

all eigenvectors. Figure 4.2(a) shows the number of unique parameters with substantial contributions to the ordered eigenvectors of H. Considering the matrix for the Full (Partial) run, we obtain 81 (56) eigenvectors. We observe that we require 13 (18) eigenvectors to capture 90% of the parameters, and 21 (25) eigenvectors to capture all parameters with significant weights in at least one eigenvector. Thus, all parameters contribute significantly in the first 16% (42 %) of the stiff directions.

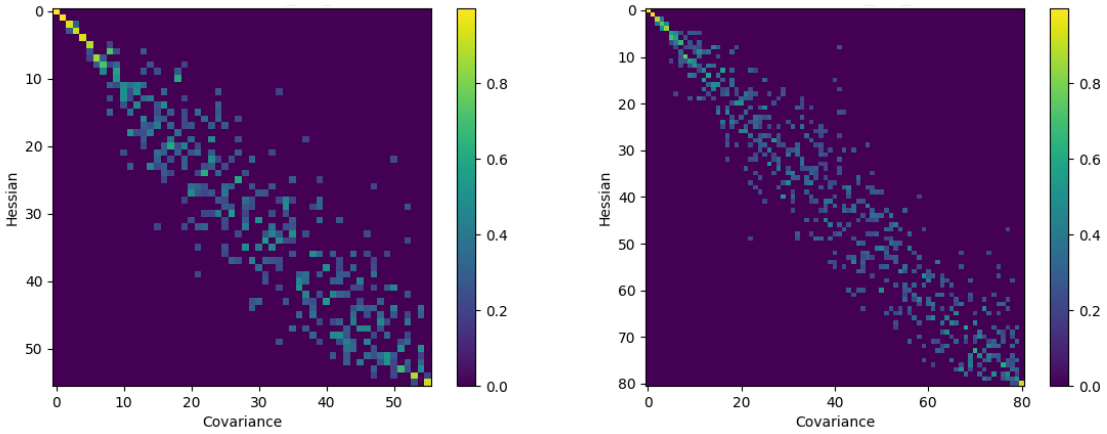
Finally,x we observe an initial slump in the number of unique parameters, implying that in the Full run, only 28 parameters contribute substantially to the first seven stiff directions, i.e., around 34% of the parameters contribute to the stiffest directions, as shown in Figure 4.2(a). Interestingly, these parameters were very similar to the Partial run, as indicated by the fact that the intersection is close to the circles in Figure 4.2 (a).



(a) Eigenvectors of H sorted in decreasing order of their corresponding eigenvalue

(b) PCs sorted in increasing order of their explained variance ratio

Figure 4.2: Comparison between the parameters with high weights in eigenvectors. On the x-axis are the sorted eigenvectors, such that the eigenvector corresponding to the largest eigenvalue gets a rank of 1. Each plot shows the number of unique parameters with substantial coefficients in the each eigenvector. The plots show the results for the two runs of our iterative scheme on the Full (81) and the Partial (56) list of parameters as described in the text, with the horizontal black lines marking the number of parameters in each list. The gray lines mark the number of eigenvectors required to capture 90% of the parameters. The black horizontal line marks the total number of parameters varied in our analysis, namely 81 and 56 in the Full and Partial run respectively. The vertical lines mark the size of the intersection between the sets of parameters between the two runs for each eigenvector. The plot on the left is derived from the eigenvectors from the approximate Hessian, whereas the plot on the right represents the results from the principal components resulting from carrying out PCA on the collection of parameter sets, i.e., the eigenvectors of the covariance matrix sorted by the inverse of their eigenvalues.



(a) Partial set (56 parameters) (b) Full set (81 parameters)

Figure 4.3: We compared the eigenvectors of the approximate Hessian with the principal components of the ensemble of parameter sets. The eigenvectors were sorted in descending order of their corresponding eigenvalue. The PCs were sorted in ascending order of their corresponding explained variance ratio (which is proportional to their eigenvalue). The heat-map shows the dot product of the eigenvectors and the PCs, with a brighter color implying a higher value.

4.1.7 Comparison between the eigenvectors of the Hessian and the inverse covariance matrix

We also compared the quality of the approximate Hessian derived from our iterative scheme, with the inverse of the covariance matrix (Σ^{-1}) of the ensemble of parameter sets [124]. Since the eigenvectors of the Σ^{-1} matrix are identical to those of Σ, we obtained the principal components (PCs) of the ensemble of parameter sets (i.e. the eigenvectors of the covariance matrix). We then sorted the PCs in increasing order of the explained variance such that the PC with the smallest explained variance (proportional to its eigenvalue) received a rank of 1. Figure 4.2 (b) shows the parameters with substantial weights in these sorted eigenvectors. We note that the trends displayed by the PCs are qualitatively similar to those of the eigenvectors of the approximate Hessian in Figure 4.2 (b). We next examined if the eigenvectors and PCs were actually identical by studying their pairwise dot products. Figure 4.3 presents a heat-map of the pairwise dot products. A brighter color indicates a number closer to 1, indicating greater similarity. There is a very good correspondence between the first 8-10 eigenvectors and PCs, indicating that our iterative scheme is able to confidently estimate the stiffest directions in parameter space.

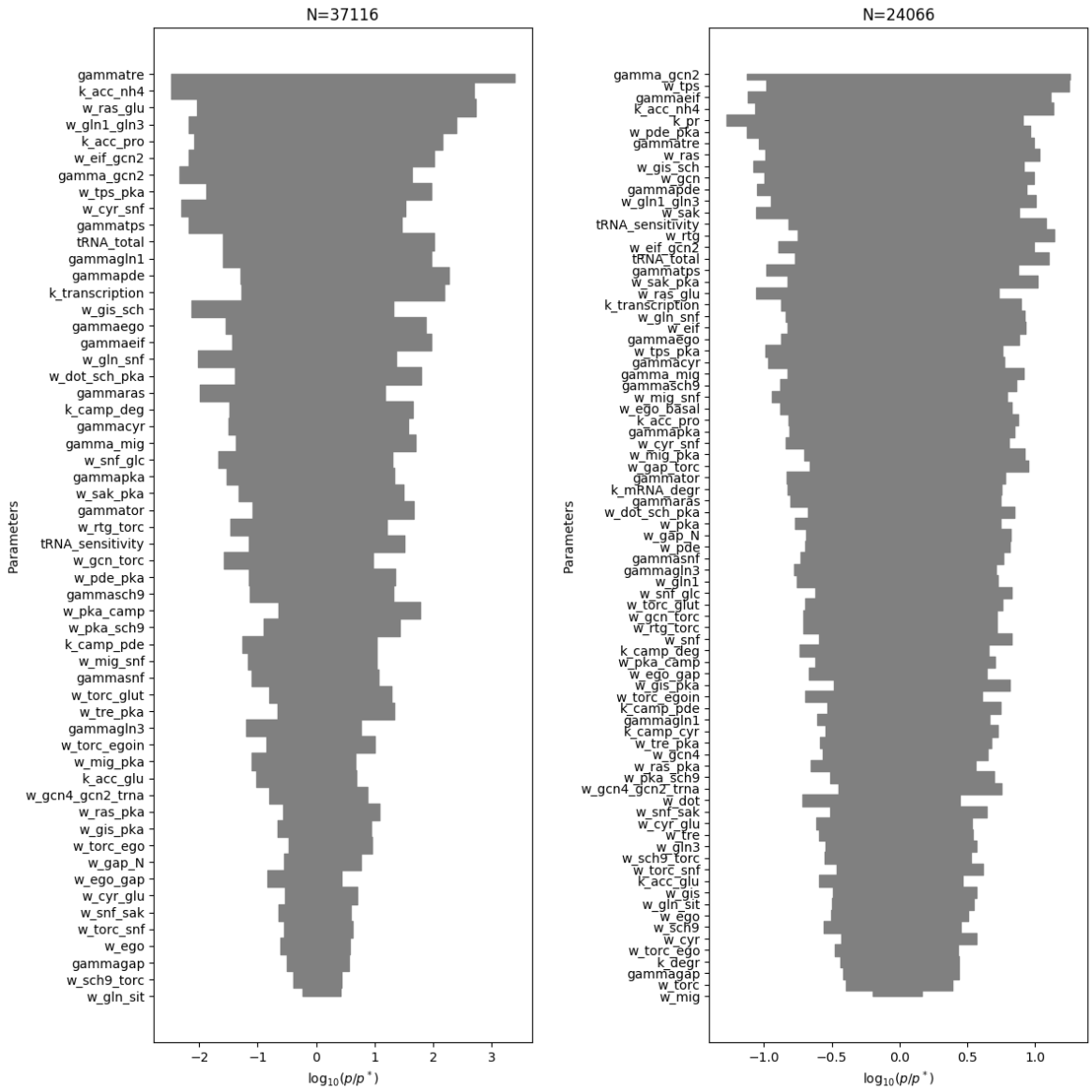
4.1.8 The relative ranges explored agree with the amount of data used to constrain the corresponding variable

Motivated by the success of our iterative Hessian-directed search in identifying the stiff directions of the cost function, we next examined the relationship between the stiff parameters and the data constraining the model. We first studied the relative ranges of parameter values explored for each parameter. These are visualized as ratios of parameter values with respect to $\mathbf{p}^*$ on a log10 scale in Figure 4.4. The parameters are sorted according to the relative ranges explored. The ensemble of parameter sets was further analyzed. The parameter ranges vary from$<$ 1 to 2 orders of magnitude. Figure 4.4 summarizes these parameter ranges. One feature that stands out from this sorting is that the gamma parameters which govern the time scales are mostly found in the top half of the plot, with broad ranges. This is likely a consequence of the lack of time series data used to constrain most of the variables in the model. Examining the parameters with narrower ranges, at the bottom of the plot, we notice that while many parameters do occur in the equations of variables that are constrained by data, this is not the case for other parameters. Examples include ω_{cyr} (regulates basal dynamics Cyr1), $\omega_{\text{torc\_ego}}$ (regulates the stimulation of TORC1 by Gtr1/2) and ω_{gis} (regulates basal dynamics of Gis1) (Figure 4.4) While the ranges in the figure indicate that the model is very sloppy in general, the occurrence of these 'unconstrained' parameters at the bottom of the figure was surprising. In order to investigate the influence these parameters had on the model, we decided to carry out a detailed parameter perturbation analysis.

4.1.9 Model structure exerts an important influence on the stiff/sloppy classification of parameters

In order to study the relationships among parameters, the model structure and the model constraints, we first ranked the parameters in the model by their contribution to the stiff directions. For this, an arbitrary cutoff of one standard deviation of the distribution of weights for parameters across all eigenvectors was chosen. Then, the parameters with absolute weights greater than the chosen cutoff were designated to contribute substantially to a given eigenvector. Finally, a cumulative list of parameters was constructed, where the rank of a given parameter is the eigenvector number in which it first appears with significant weight.

From the comparison of the eigenvectors of the covariance matrix and the approximate Hessian, it can be observed that there is a good correspondence between the first 8 or 9 eigenvectors (Figure 4.3). These were designated the high confidence directions, and the parameters occurring in these directions are marked in bold in Table C.1 respectively. A striking finding from this table is that among the top-ranked parameters, many are not constrained by data, i.e., they do not appear in equations whose dynamics are constrained by data.



(a) Partial set (56 parameters) (b) Full set (81 parameters)

Figure 4.4: Ranges of parameters explored over all parameter sets at the end of four iterations. The smallest and largest value of each parameter over the ensemble were chosen, and the log10 value of the ratio with respect to the p* value was used to define the range.

While the ranks of parameters in Table C.1 indicate a complex relationship between model constraints and model structure influencing the structure of the cost surface, we were interested in the distribution of parameters that do appear in equations constrained by data. To investigate this distribution, starting from $\mathbf{p}^*$ we picked each parameter one at a time and perturbed its value in a $\pm 2.5\%$, $\pm 10\%$, and ± 10-fold range and obtained the fitting cost in

each instance. We also measured the model cost when the parameter was set to 0. Figure 4.5 shows the results of this analysis. The color of the heat-map is the log 10 fold increase in cost over $C_{\min}$. The cost values are truncated to a 10-fold increase so that smaller costs are visible. We observe that while a perturbation of up to $\pm 10\%$ has little effect on model cost, a 10-fold change produces a dramatic increase in model cost for the parameters on top of the ranked list, while those at the bottom of the list show a decreased effect, in agreement with the ranking of stiffness. The arrows in Figure 4.5 indicate the parameters which occur in equations constrained by data. We observe that these parameters do not exhibit any type of clustering, and are distributed across the entire list.

These observations indicate that, despite a small amount of data available to constrain the model, the model structure (in particular the pathway crosstalk and feedback interactions) might play an important role in indirectly constraining other parts of the model that are not directly constrained by data.

4.2 Perspective

Examining the ranked parameters in Tables C.2 and C.1, we notice that there is no relationship between presence of experimental data constraining the dynamics of an equation, and the ranks of parameters appearing therein. In fact, in Table C.2, the first 6 stiffest parameters appear in an equation (TORC1) that is unconstrained by data. This suggests that the relationship between model constraints and parameter sloppiness is nonlinear. In the original investigation into parameter sloppiness by Gutenkunst et al., synthetic continuous time course data is generated to constrain all model variables, and the conclusions about universal parameter sloppiness are drawn from this analysis. In our work, with a sparsely constrained cost function, it is unclear if we are identifying the 'true' sloppy directions. Consequently the the reason for a parameter to be designated as sloppy can arise from both structural reasons as well as the lack of model constraints. Moreover, the approach of sloppy model investigation is fundamentally based on the assumption that the model topology is a good representation of the 'true' system topology. The relationship between model constraints, model topology, and parameter sloppiness remains unclear and warrants a systematic investigation.

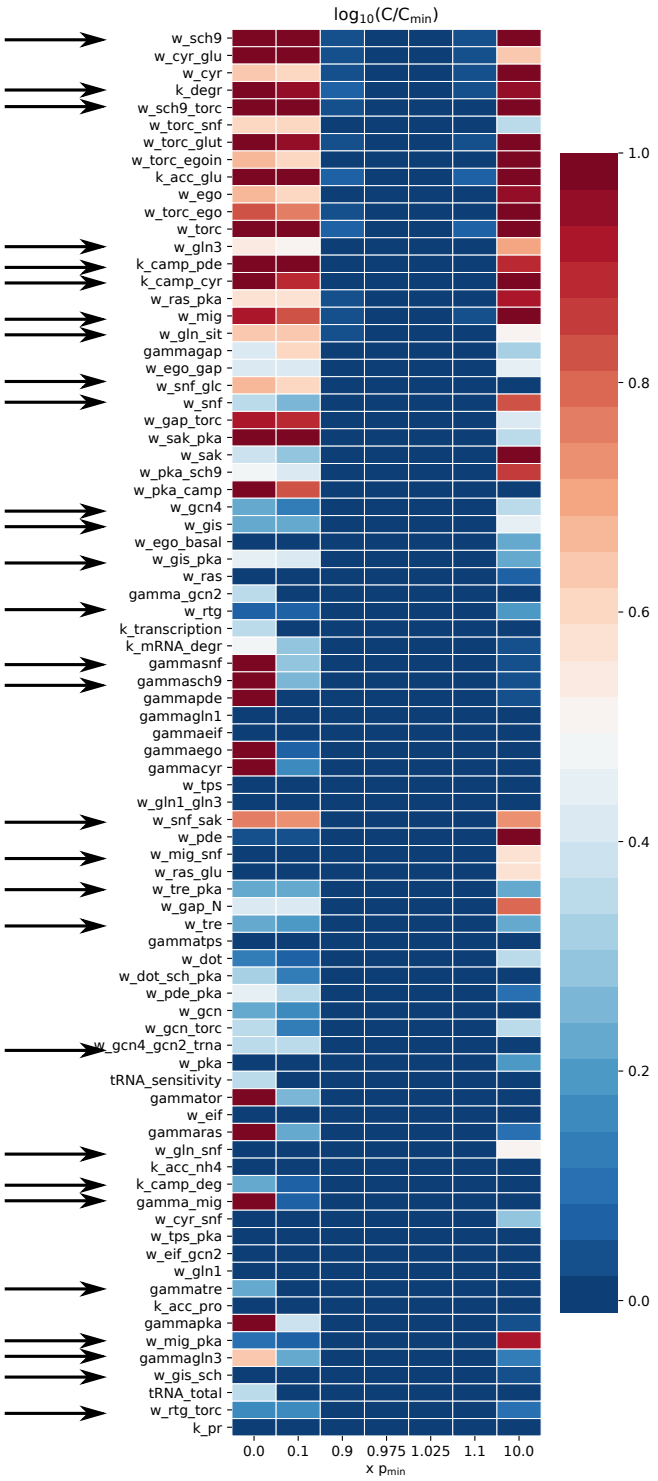


Figure 4.5: Parameter robustness is governed by model topology and experimental data. The plot shows a heat map of relative increase in model cost when parameters are perturbed one at a time. The parameters along the y-axis are ordered by their rank, i.e., their contribution to the stiff directions. The arrows indicate the parameters that occur in equations constrained by data.

Chapter 5

Model predictions

Jalihal, A.P., Kraikivski, P., Murali, T.M., Tyson, J.J., 2020. Modeling and analysis of macronutrient signaling in budding yeast. *Under review.* Also available on *bioRxiv* (https://doi.org/10.1101/2020.02.15.950881)

In the previous chapter, we described the approach we employed to use the quantitative literature curated data to create a collection of alternate parameter sets. In this chapter, we will describe our investigations of using the calibrated nutrient signaling model to make novel, experimentally testable predictions.

5.1 Testing the model against observed phenotypes of mutant strains

In this section, we test the behavior of the parameterized model against qualitative experimental observations of the responses of *wt* and mutant strains to environmental perturbations. These 'test case' data were not used in the previous section to constrain the parameters.

From the literature we collected a list of 40 experiments (provided in Supplementary Table 4, Appendix D) describing the qualitative phenotypes of mutant strains, typically colony growth phenotypes. We selected strains whose cellular states under nutrient shifts can be adequately characterized in terms of the transcription factors (TFs) in our model. To compare these experiments with our model predictions, we used the following strategy. For each colony growth experiment, we interpreted the observed phenotype in terms of the state (ON or OFF) of the nutrient responsive TFs in our model. Next we simulated each nutrient-shift or drug-treatment protocol for the specific mutant strain. We recorded the predicted steady-state values of the six TFs in our model, and converted these values into binary form (0 or 1) using thresholds corresponding to the half-maximal values in the *wt* simulation; thereby obtaining a Boolean vector of predicted TF states. Finally, we compared the predicted states of the relevant TFs to those interpreted from the experiment.

An important assay in the investigation of the nutrient signaling response is rapamycin treatment, constituting 22 of the 40 experiments in our collection. In *wt* cells, rapamycin treatment inhibits TORC1 activity, consequently activating nitrogen stress responses via TFs

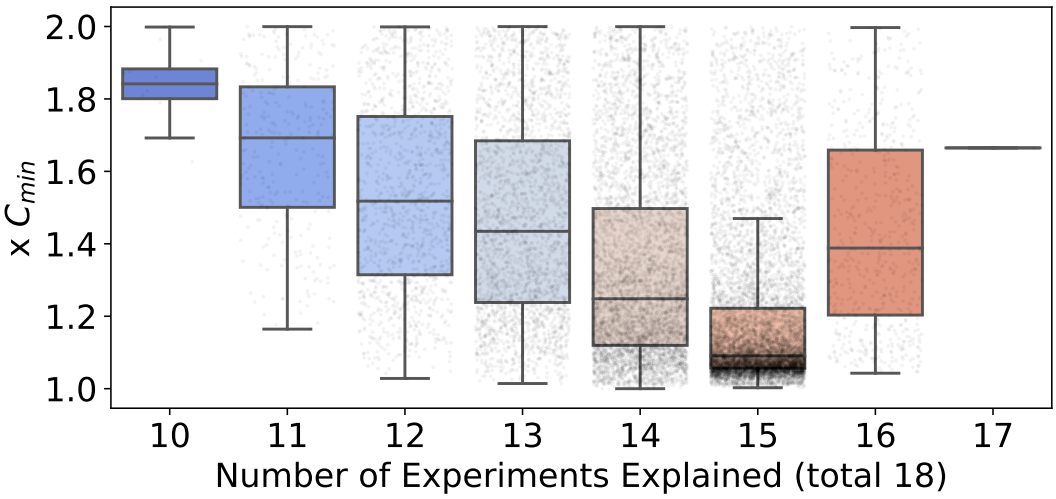


Figure 5.1: Dependence of model cost on explanatory capacity, across the entire collection of 18,000 alternative sets of parameter values. Model predictions of the 40 qualitative phenotype experiments were repeated for the collection of acceptable parameter sets. Each boxplot shows the distribution of cost-of-fit (to the quantitative measurements) for a given level of explanatory capacity (i.e. the number of qualitative phenotypes explained). Each point represents a parameter set. The cost, on the y-axis, is reported as a multiple of $C_{\min}$, the best observed cost across all parameter sets. The boxplot shows the median cost, and the whiskers extend to 1.5 times the interquartile range (IQR), or the last data point less than 1.5xIQR. Note that only parameter sets with a cost less than or equal to twice the $C_{\min}$ are reported.

like Gln3, Gcn4, and Rtg1/3 (via the Tap42-Sit4 branch), inhibiting ribosome biogenesis via transcriptional repressors like Dot6/Tod6 (via the Sch9 branch), and indirectly causing a cell cycle arrest in the G1 phase. It appears that decoupling either one of these branches from TORC1 is sufficient to confer rapamycin resistance (cf. rapamycin treatment of *gln3*Δ*gcn4*Δ double mutant [125], and the Sch9^{DE} and *tap41-11* TORC1 insensitive strains [126]). Taking these observations into account, we attempted to carry out *in silico* experiments to explain our collection of experiments by examining different potential effects of rapamycin treatment, including upregulation of nitrogen adaptation responses and downregulation of ribosome biogenesis. Using either of these definitions, the model could explain only about half of the rapamycin experiments that we have collected. We present this investigation in Section 5.6. Below, we focus on experiments not involving rapamycin treatment.

In order to characterize our confidence in model predictions, we recorded statistics on the number of times an experiment is predicted correctly across the representative collection of parameter sets (Table 5.1). Across all parameter sets, the model is able to correctly simulate up to 17 of 18 experiments. Five of these 18 experiments are explained by all of the parameter

Exp ID	Confidence	Exp ID	Confidence
26-gcn2 snf1	100.0	2-rho0	96.12
12-gln3 gat1	100.0	29-rph1 gis1	94.68
24-gcn2	100.0	3-rtg1	80.64
41-sch9 gis1	100.0	23-wt	76.6
31-mig1 snf1 pde2	100.0	1-rho0	72.39
33-mig1 snf1 pde2	99.99	11-gln3	67.31
32-mig1 snf1 pde2	99.94	40-sch9	8.23
30-rph1 gis1	96.89	5-snf1	9.28
13-wt	96.12	4-rtg1	0.94

Table 5.1: Confidence in qualitative predictions expressed as the percentage of parameter sets that make the correct prediction. The experiments are ordered in decreasing order of prediction confidence. The *in silico* experiments corresponding to the experiment IDs are presented in Chapter D.

sets, and three of 18 are predicted by fewer than half of the parameter sets. The following sections discuss some of the experimental observations that our model succeeds in explaining and some that our model fails to explain. The predictions described here are made using the optimal parameter set. Details of simulations are provided in Chapter D.

5.2 Nitrogen adaptation responses.

An important function of the tricarboxylic acid cycle (TCA cycle) in central carbon metabolism is diverting the carbon backbone to amino acid biosynthesis pathways via α-ketoglutarate. Since the genes coding for the TCA cycle enzymes are present in the mitochondrial genome, mitochondrial dysfunction can severely affect respiration, resulting in the so-called ρ^0 petite strains. Conceivably, these strains have a diminished capacity to divert carbon flux into amino acid biosynthesis. In yeast, the Rtg1/3 transcription factors control the expression of the *CIT2* gene, a TCA cycle enzyme encoded by the nuclear genome, which allows the diversion of carbon flux to amino acid biosynthesis pathways, under nuclear control (the so called retrograde pathway). Liu et al. studied *wt* (ρ^+) and ρ^0 strains to characterize the activity of the Rtg1/3 transcription factors by systematically varying the carbon and nitrogen content in the growth medium [127]. While the ρ^0 strains can grow on media containing glucose only, raffinose only, and glucose supplemented with glutamate, the *rtg* strains in a ρ^0 background only show growth on glucose + glutamate medium.

We simulated *wt* and *rtg* strains in such ρ^0 petite strains. We assume that ρ^0 strains have a lower level of glutamine pool replenishment than ρ^+ strains. The *rtg1*Δ, *rtg3*Δ and *rtg1*Δ*rtg3*Δ strains in ρ^0 backgrounds do not grow on media lacking glutamate, while supplementing the medium with glutamate results in *wt* growth [128]. Model predictions

recapitulate the observations made by Liu et al. in media with and without glutamate (Sections D.1 to D.3).

5.3 The Snf1 pathway.

While investigating the regulation of Gis1, Balciunas et al. studied the *mig1*Δ*snf1*Δ*pde2*Δ triple-mutant strain in various carbon sources and found that this strain is viable when grown on glucose and is inviable when grown on raffinose [129]. In order for this strain to grow on raffinose, Gis1 will have to be expressed. Our model predicts that Gis1 will be inactive during growth on both glucose and raffinose in this strain, indicating that the strain will show no growth on raffinose (Sections D.31 and D.32).

5.4 Model mismatches

The majority of parameter sets ($> 50\%$) in our model succeed in explaining 15 of the 18 'test-case' experiments (Table 5.1). We discuss the reasons for the three model mismatches below.

Gasmi et al. [130] reported that a *snf1*Δ strain cultured on minimal medium supplemented with ethanol grows slowly. In our model the glucose repression factor Mig1 will remain active in a *snf1*Δ strain, inhibiting any carbon adaptation responses, indicating no growth (Section D.5). Liu et al. [127] showed that an *rtg1*Δ single deletion strain does not grow on medium containing a low amount of glutamate (0.02%). Our model predicts that this strain will still mount an Rtg1/3 response (Section D.4). This might indicate a problem with the representation of nitrogen sufficiency in the model. Roosen et al. [32] showed that an *sch9*Δ strain grows when cultured on medium containing glycerol as a carbon source. In order to grow on glycerol, the Gis1 transcription factor must be activated via inhibition of PKA. Since our model assumes an inhibition of PKA via Sch9, an *sch9*Δ strain, with hyperactive PKA, will result in repression of Gis1 (Section D.40).

5.5 Predictions of global cellular responses to nutrient states

Motivated by our model's success in predicting the qualitative phenotypes of 15 out of 18 mutant yeast strains, we used it to predict global cellular responses to nutrient conditions in a variety of mutant strains. We chose a list 16 mutant strains affecting all three nutrient signaling pathways: thirteen of these strains are gene deletion mutants including

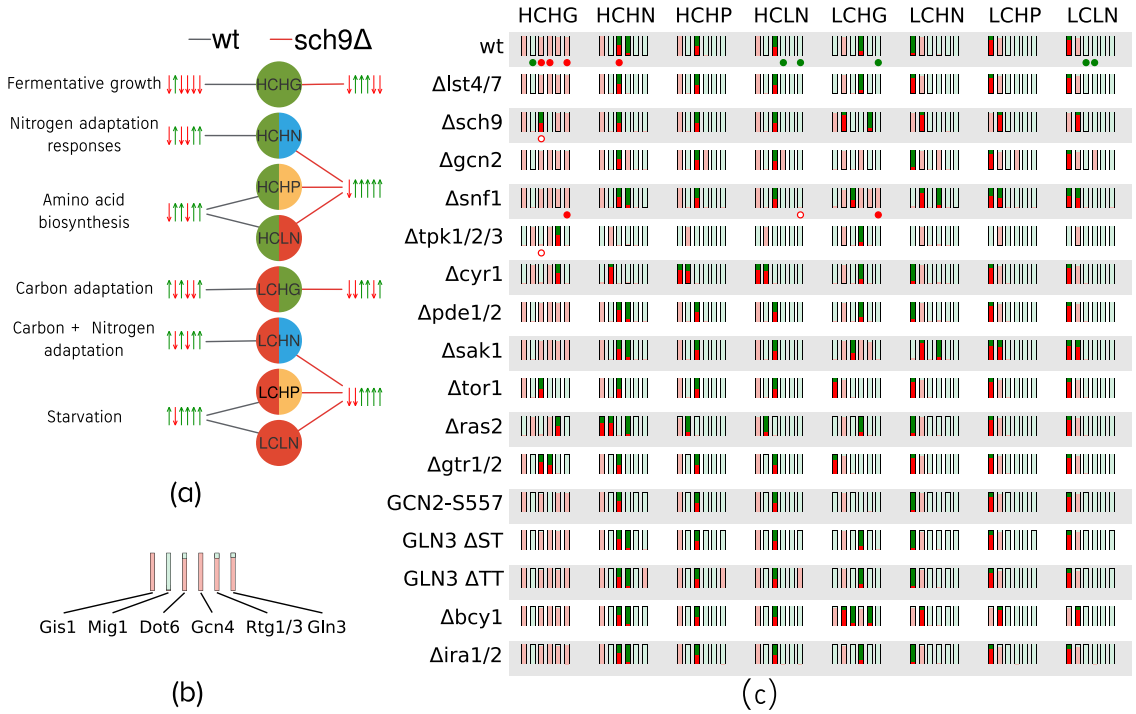


Figure 5.2: Global cellular responses to nutrient conditions. (a) Illustration of cellular states in response to various nutrient environments. The nutrient input is shown in the circles: L, H indicate low and high, respectively; C, G, N, P represent carbon, glutamine, ammonia and proline, respectively. The cellular state, defined by the set of transcription factor (TF) readouts, is shown by the upward pointing green (on) and downward pointing red arrows (off). The order of TFs is: Gis1, Mig1, Dot6, Rtg1/3, Gcn4, Gln3. Each nutrient input is connected to a cellular state by an edge. The figure shows a comparison between the predicted states for two strains, *wt* (black edges) and *sch9*Δ (red edges), using the optimal parameter set. Qualitative descriptions of the cellular states in the *wt* simulations are provided on the far left. (b) Representation of the consensus prediction of cellular state across 18,000 alternative parameter sets. The example shows the prediction for *wt* cells under HCHG condition. The height of a red/green bar represents the fraction of parameter sets that predicted the corresponding state to be off and on, respectively. We consider the prediction to be robust if greater than 90% of parameter sets are in agreement, shown using light green or red. As shown, most parameter sets are in agreement regarding the states of the readouts for *wt* cells under high carbon, high glutamine. (c) The robustness of global state predictions across 17 strains and 8 nutrient conditions. Fragile predictions are indicated in bright green and red. Additionally, direct measurements of TF activities obtained from the literature are indicated using filled and empty circles below the corresponding readouts; a filled circle indicates that model prediction agrees with experimental data while empty circles represent a model mismatch. Green and red circles indicate that the readout is on or off, respectively.

*lst4*Δ*lst7*Δ, *sch9*Δ, *gcn2*Δ, *snf1*Δ, *tpk1*Δ*tpk2*Δ*tpk3*Δ, *cyr1*Δ, *pde1*Δ*pde2*Δ, *sak1*Δ, *tor1*Δ, *ras2*Δ, *gtr1*Δ*gtr2*Δ, *bcy1*Δ, *ira1*Δ*ira2*Δ. and three are protein-sequence mutants: *GCN2-S577A* has lost the phosphorylation site for TORC1, and the two hypothetical strains *GLN3-*Δ*ST* and *GLN3-*Δ*TT* lack the phosphorylation sites for Snf1 and TORC1, respectively (see Figure 5.2).

We characterized each strain's response to eight nutrient conditions (qualitatively low and high inputs for 'Carbon', 'Glutamine', 'Ammonia', and 'Proline'). In Figure 5.2, these qualitative states are denoted using 'H' and 'L' respectively for high and low, and the nutrient input is denoted using 'C', 'G', 'N', and 'P' for Carbon, Glutamine, Ammonia and Proline, respectively. (Note that 'LN' represents nitrogen starvation.) For each simulation, we recorded the steady state values of the six transcription factors (TFs) in our model and thresholded the TF activities using their half-maximal values in the *wt* simulations. We thus obtained a Boolean vector describing the phenotypic state of a cell under different nutrient conditions, with each TF taking a value 0 (off) or 1 (on). The predicted states for the 17 strains (including *wt*) over all nutrient conditions (as predicted by the model using the optimal parameter set) are provided in Supplementary Table S5.

Figure 5.2(a) provides an example of the predicted cellular states for *wt* and *sch9*Δ cells for each of the eight nutrient conditions. We observed that, for each nutrient input, the state of the *wt* strain was different from that of the *sch9*Δ strain. Repeating these simulations across the 17 strains and the 8 nutrient conditions yielded a total of 136 predicted cellular states. In order to gain confidence regarding model predictions, we repeated these *in silico* experiments for each of the 18,000 alternative parameter sets. Using two parameter sets, if the predicted activity of a particular TF in a given experiment is identical, the two sets are considered to agree on the prediction. We define fragile predictions to be the experiments where fewer than 90% of the parameter sets were in agreement about the outcome of a simulation. The results of this analysis are summarized in panel (c) of Figure 5.2, where the consensus of the predictions across all tested parameter sets is indicated by the brightness of the bars. We observed that, in the HCHG simulations, the predictions exhibited consensus across most strains for all TFs. Note that in Figure 5.2(c) there is significant disagreement between the predictions made by the parameter sets regarding the state of Dot6, the repressor of RIBI, in the case of high carbon (HC) irrespective of the nitrogen sufficiency. This is likely due to the lack of experimental data constraining Dot6 dynamics. Furthermore, we observe that the HP and LN columns are identical for both HC and LC, implying that the model fails to distinguish between the high proline and nitrogen starvation conditions, whereas proline, a poor source of nutrients would still be expected to serve as a substrate in order to maintain intracellular amino acid reserves. We believe that this failure is a consequence of the fact that our model does not include the effects of metabolic feedback on nutrient signaling, thus rendering the two poor-nitrogen conditions indistinguishable.

Finally we observe that the alternative parameter sets disagree about the state of Gis1 across most strains during nitrogen stress under carbon starvation (LCHN, LCHP and LCLN columns in Figure 5.2(c)). Gis1 is jointly regulated by Sch9 and PKA. However, the data

used to calibrate the model come from experiments related to the PKA pathway only [73]. In order to refine our predictions for Gis1, the model will need to be recalibrated with Gis1 data obtained from nitrogen-shift experiments, when they become available.

5.6 Alternative interpretations of rapamycin treatment

As discussed in the section "Testing the model against observed phenotypes of mutant strains", 22 experiments that we collected from the literature relate to rapamycin treatment in various mutant strains. There are three ways of interpreting the immediate effect of rapamycin on a strain grown in rich medium. First, via the Sch9 branch of TORC1 signaling, rapamycin can lead to an inhibition of ribosome biogenesis. In our model, this would be represented as an upregulation of Dot6 activity. Second, via the Tap42-Sit4 branch of TORC1 signaling, the activities of Gln3 and Gcn4 can result in an upregulation of nitrogen starvation and adaptation responses. Finally, via mechanisms not currently present in the model, TORC1 can directly or indirectly impinge on the cell cycle machinery to cause a G1 arrest [131]. We considered the first two definitions of the effect of rapamycin on cells. As shown in Figure 5.3, both the Dot6 and Gln3Gcn4 definitions show a trend of decreasing median cost until 11 of 22 experiments are explained. However, using the Dot6 definition, a large number of high cost parameter sets succeed in explaining up to 16 of 22 experiments.

We first describe the causes of model mismatch across both definitions of rapamycin. Finally, we investigate the basis of the 2000 parameter sets that explain up to 16 experiments using the Dot6 definition.

Model mismatches Eleven of the 22 experiments are explained by less than 50% of the parameter sets using either definition of rapamycin treatment (Table 5.2). Four of the 11 rapamycin treatment experiments involve strains carrying mutations downstream of TORC1. $SCH9^{DE}$ encodes a constitutively active Sch9 kinase. *tap42-11* is a temperature sensitive allele of TAP42, which encodes a protein involved in transmitting the TORC1 signal to Sit4 and other stress response TFs. The single mutant strains $SCH9^{DE}$ and *tap42-11* are slightly resistant to rapamycin treatment, whereas the double mutant strain $SCH9^{DE}$*tap42-11* is fully resistant [84]. Using our definition of rapamycin treatment based on Gln3 and Gcn4 activities, strains involving *tap42-11* are predicted to be rapamycin resistant. However, since neither of these TFs are regulated by Sch9, the $SCH9^{DE}$ strain is predicted to be rapamycin sensitive. Additionally, three strains, namely *gln3*Δ*gat1*Δ, *gcn4*Δ, and an overexpression mutant 2μ URE2 are all predicted to be rapamycin sensitive as a consequence of our chosen definition (Sections D.6, D.7 and D.28).

Seven of the 11 rapamycin treatment experiments that we have curated include mutants of the carbon signaling pathway. These experiments clearly indicate that mutations affecting the carbon signaling pathways influence the nitrogen adaptation response, and the model's failure to explain these results give us insight into the crosstalk between carbon and nitro-

gen pathways. As mentioned in the description of the PKA pathway in the Results section of the main text, our model supports some results from Schmelzle et al., but not from Zurita-Martinez et al., originating from strain specific differences. These observations account for three of the eight mismatches ('14-bcy1', '15-ira1', and '16-ira1ira2' described in Sections D.14 to D.16). Schmelzle et al. examined three hyperactivating PKA strains in a *gln3*Δ*gat1*Δ background, ('19-RAS2v19gln3gat1', '20-TPK1gln3gat1', and '22-bcy1gln3gat' described in Sections D.19, D.20 and D.22). These strains were shown to be rapamycin resistant. However, our model predicts that these strains are sensitive to rapamycin. Our model does not currently include a direct interaction between PKA and TORC1. Similarly, the last rapamycin treatment mismatch relates to a *snf1*Δ strain which was observed to show rapamycin resistance [89]. While our model assumes that Snf1 inhibits TORC1, Snf1 will be inactive during growth on YPD, hence the model predicts that a *snf1* deletion will not affect Gln3 activity in this nutrient condition. Further mechanistic details of crosstalk between PKA, Snf1 and TORC1 will be needed in order to resolve these mismatches.

The mechanism of Dot6 regulation explains the distribution of number of experiments explained We investigated the cause of the spread of number of experiments using the Dot6 definition, shown in Figure 5.3. We termed the parameter sets that explained 11 or fewer experiments as the "low" set, and those that explained more than 11 as the "high" set. We observed that the "low" set failed to explain the experiments starting from row 3 ('20-TPK1 gln3 gat1') in the first column of Table 5.2. These experiments constitute the model mismatches pertaining to claims by Schmelzle et al. and Zurita-Martinez et al. regarding the role of the PKA pathway in rescuing the rapamycin treatment induced growth arrest phenotype. We hypothesized that the cause for this bimodal distribution of experiments explained was related to the representation of Dot6 in the model. Dot6 is regulated by both Sch9 and PKA. In order for rapamycin treatment to be sufficient for Dot6 activation, the Sch9 and PKA must regulate Dot6 via an AND gate. If this condition is not satisfied, the fact that rapamycin treatment experiments are carried out in glucose medium will imply that PKA can independently repress Dot6. The parameters that regulate Dot6 are w_dot, the basal activation term, and w_dot_sch_pka, the strength of inhibition by the sum of PKA and Sch9 activities. Since PKA and Sch9 take a maximum value of 1.0, Dot6 is maximally repressed for any combination of parameters satisfying the following relation

$$\frac{\text{w\_dot}}{\text{w\_dot\_sch\_pka}} < 2$$

Figure 5.4 shows that qualitatively a majority of the parameter sets in the "high" set indeed satisfy this relationship.

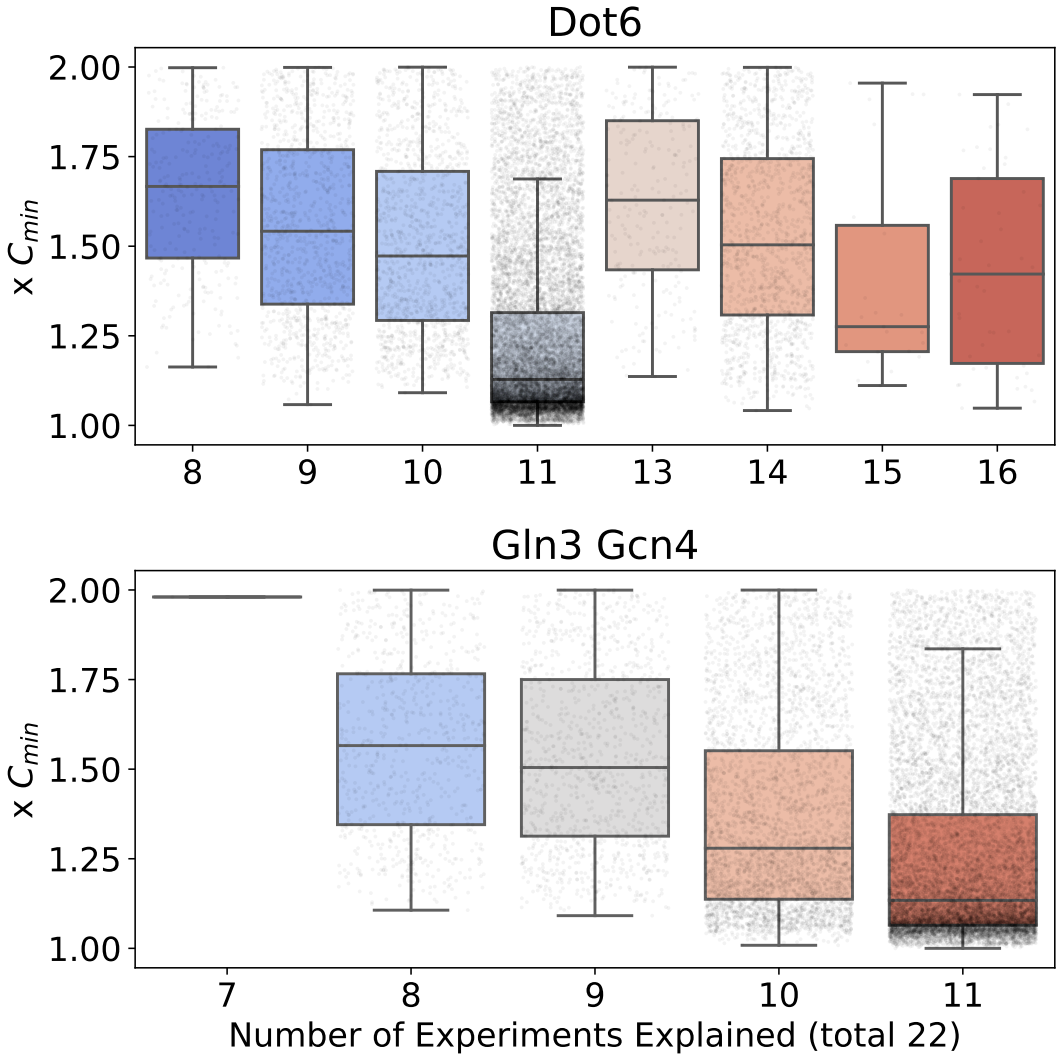


Figure 5.3: Dependence of model cost on explanatory capacity, across the entire collection of alternative sets of parameter values. Rapamycin experiments are defined using Dot6 as a model readout. The collection of parameter sets explaining less than or equal to 11 experiments are termed the "low" set, and those explaining more than 11 are termed the "high" set.

5.7 Discussion

Since the default state of yeast cells in the presence of nutrients is exponential growth, the coordination of various nutrient levels within the cell is of paramount importance. The integration of extracellular nutrient signals is distributed broadly across the cAMP/PKA (glu-

Exp ID	Dot6	Gln3Gcn4
20-TPK1 gln3 gat1	13.13	0.74
6-gln3 gat1	13.63	0.74
7-2μ URE2	13.63	0.7
37-tap42-11	13.63	71.41
34-snf1	13.63	0.0
27-gln3 gcn4	13.63	100.0
14-bcy1	13.64	0.0
22-bcy1 gln3 gat1	13.65	0.74
19-RAS2v19 gln3 gat1	13.65	0.74
16-ira1 ira2	13.65	0.0
15-ira1	13.65	0.0
10-gat1	85.44	84.81
28-gcn4	86.37	0.0
9-gln3	86.37	84.81
35-reg1	86.37	99.26
36-ure2	86.37	99.26
8-wt	86.37	98.91
18-tpk1	86.67	99.06
17-ras2	86.78	99.17
21-bcy1	86.87	99.22
38-SCH9^{DE}	88.18	0.18
39-SCH9^{DE} tap42-11	88.18	100.0

Table 5.2: List of rapamycin treatment experiments, with confidence in model predictions. Confidence in qualitative predictions is expressed as the percentage of parameter sets that make the correct prediction. The *in silico* experiments corresponding to the experiment IDs are presented in Chapter D.

cose signaling), Snf1 (carbon adaptation) and TORC1 (nitrogen signaling) pathways, along with many crosstalk interactions among members of these pathways and their downstream targets. These interactions comprise the nutrient-signaling and decision-making system in budding yeast. In this paper we have proposed a mathematical model of this system, based on well-studied interactions of the individual pathways and less-studied crosstalk between the pathways (Figure 2.3). Our results demonstrate that the proposed mechanism can account for many aspects of nutrient signaling in budding yeast, as observed in diverse nutrient-shift experiments (Figure 3.1).

Global cellular responses to nutrient shifts Two experimental observations highlight the crosstalk between nutrient signaling pathways: (1) cells facing nitrogen stress in the context of carbon abundance can redirect carbon flux from the TCA cycle into amino acid biosynthesis [132], and (2) acute nitrogen starvation results in down-regulation of glucose

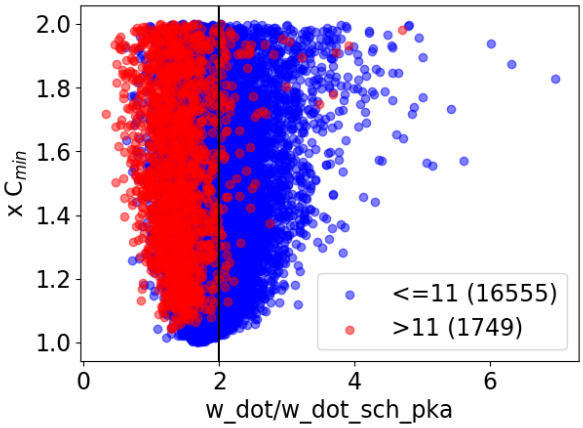


Figure 5.4: Distribution of ratios of kinetic parameters of Dot6. The "high" set is represented by the red points, and the "low" set is represented by the blue points. The scatter plot shows the ratio of the Dot6 parameters on the x-axis, and the cost associated with the parameter set on the y-axis

fermentation [133]. These interactions are regulated by metabolic enzymes that are under the regulation of various transcription factors (TFs) included in our model. To get a sense of the cell's response to complex nutrient shifts, we thus need to examine the activities of the TFs spanning carbon and nitrogen responses.

In our model, we use the Boolean states of the TFs to define the state of a cell. Each of these predicted states is deemed 'robust' or 'fragile' based on the consensus among predictions from a representative collection of 18,000 alternative sets of parameter values (see Figure 5.2 (c)). Robust predictions that are subsequently shown to be incorrect indicate parts of the model that need further refinement. Fragile predictions, on the other hand, indicate those parts of the model that are poorly constrained by available data. Experimental tests of a fragile prediction can invalidate those 'alternative' parameter sets that make the incorrect prediction. Hence, both robust and fragile predictions suggest potentially informative experiments that can be used to further constrain the model. In the following paragraphs, we propose experiments that we deem informative based on some fragile predictions.

The Rtg1/3 transcription factors play the crucial role of diverting carbon flux into amino acid biosynthesis. Rtg1/3 are expected to be activated during nitrogen stress in carbon-rich conditions. Figure 5.2 (c) indicates that the model makes fragile predictions for Rtg1/3 activity in the HCHG condition in three mutant strains, namely *tpk1/2/3*Δ, *cyr1*Δ, and *ras2*Δ. In these three strains PKA is inactive. Because, in our model, PKA activates TORC1 (via Sak1 and Snf1), TORC1 activity should be low in these mutants. However, since high glutamine activates TORC1, this protein receives contradictory signals from the carbon and nitrogen pathways. As a result, Rtg1/3 may be either active or inactive, depending on the values of the parameters in our model simulations. Hence, data pertaining to Rtg1/3 activity

in these strains will help to narrow down the acceptable sets of parameter values.

The Mig1 transcriptional repressor implements glucose repression, i.e., in the presence of glucose, Mig1 inhibits expression of genes involved in carbon adaptation responses. During glucose starvation, Snf1 inhibits Mig1 by localizing it to the cytoplasm. Interestingly, our model predictions for the state of Mig1 are fragile in LCHN, LCHP, and LCLN conditions in two separate strains, *sch9*Δ and *bcy1*Δ. (Note: Sch9 inhibits PKA via Bcy1.) In this case, the fragile predictions of Mig1 activity are consequences of contradictory signals received by Snf1. In LC, Snf1 should be active and Mig1 inactive. However, in these mutants, despite LC conditions, PKA is active, Sak1 is inactive, and there is little or no activation of Snf1. Hence, whether Snf1 is active or inactive depends sensitively on precise parameter values. Therefore, measurements of Mig1 activity under glucose starvation conditions in the *bcy1*Δ strain will also help to narrow down the acceptable sets of parameter values.

An important determinant of cell growth under diverse nutrient conditions is ribosome biogenesis, which in our model is regulated by the Dot6/Tod6 repressors. Interestingly, Figure 5.2 (c) shows that Dot6 predictions are fragile under HCHN, HCHP, and HCLP conditions across most mutant strains, indicating that the strength of crosstalk between carbon and nitrogen signaling pathways is crucial in determining Dot6 activity. Measurements of Dot6 activity in the mutant strains identified by Figure 5.2 (c) will provide important constraints on these crosstalk interactions. However, we must bear in mind that other factors controlling ribosome biogenesis, such as Sfp1, are not yet included in our model.

Nutrient adaptation responses and global metabolic feedback Our proposed mechanism successfully recapitulates many features of the nutrient response system in budding yeast over short timescales (30 minutes). During nutrient stress, the cell mounts a variety of adaptation responses in order to maintain intracellular nutrient pools. Our model captures the activation of stress TFs immediately after a nutrient shift. However, on longer time scales of about 50 minutes, data from Granados et al. reveal that many of these stress TFs are eventually turned off [134]. Our model simulations do not capture this feature. How can this inactivation of TFs be explained? The stress TFs activate various biosynthetic pathway enzymes which restore the flux of metabolites through anabolic processes. Consequently, the starvation signals are turned down, inactivating the stress TFs themselves. Metabolic processes are known to play important roles in cell growth and division [90]. Because the mechanisms by which metabolic responses feedback on the nutrient signaling network are complex and poorly understood, we have not yet tried to include these effects in our model (Figure 5.5).

In general, how to incorporate metabolic fluxes in nutrient signaling networks is a difficult challenge. Three aspects stand out: (1) the accurate representation of intracellular nutrient pools, (2) the metabolic fluxes spanning carbon and nitrogen metabolism, and (3) the dynamics of regulation of metabolic enzymes by the upstream signaling network. By providing a mechanistic model of the nutrient signaling system in yeast, our model is well poised to serve as a bridge between global metabolism and a wide range of cellular adaptation

responses ranging from ribosome biogenesis to autophagy.

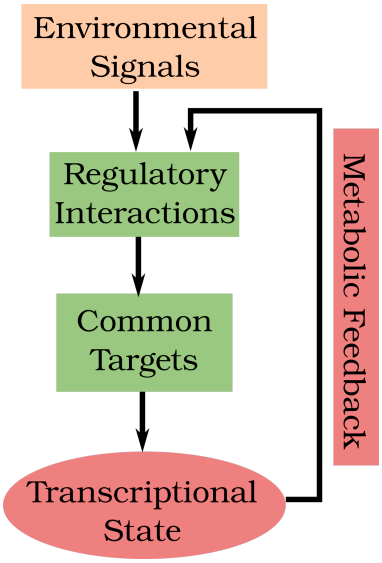


Figure 5.5: During nutrient stress, upstream sensing mechanisms transduce stress signals to downstream transcription factors. Stress signals can originate from both the extracellular environment and intracellular nutrient reserves. The latter are sensed indirectly via the levels of uncharged tRNAs (in the case of amino acids) or metabolic intermediates that are directly sensed by signaling molecules. The drops in flux through metabolic pathways result in upregulation of specific biosynthetic pathways and other adaptation responses, indicated in red at the bottom of the figure. Subsequently, metabolic fluxes are remodeled, leading to inactivation of stress responses. This 'metabolic feedback', which is currently not included in our model, determines the long-term responses of the nutrient signaling network.

Chapter 6

BoolODE

Pratapa, A., **Jalihal, A.P.**, Law, J.N., Bharadwaj, A. and Murali, T.M., 2019. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17, 147–154 (2020). https://doi.org/10.1038/s41592-019-0690-6

6.1 Background

Improvements in sequencing technologies in the past decade have led to a surge in the availability of single cell RNAseq (scRNAseq) data of high quality and throughput. scRNAseq data, while increasing the resolution of measurement to the single cell level, are typically characterized by noise and 'dropouts'. The promise and challenges of scRNAseq data have caused a renewed interest in the problem of gene regulatory network (GRN) inference, where the transcription factor (TF) interactions underlying a biological process are inferred from transcriptomic data. More than ten new algorithms have been proposed since 2015 to tackle this challenge, which accept a scRNAseq dataset as input and return a network of interactions as output (LEAP [135], PIDC [136], SINCERITIES [137], SCNS [138], GRNVBEM [139], SCRIBE [140], GRNBOOST2 [141], GRISLI [142], SINGE [143]). However, since there is no commonly accepted framework for assesing the the performance of these algorithms, end users have no criterion to select one method over the other. There is thus a need for benchmarking the performance of these algorithms for the accuracy of their proposed reconstructions. There are two commonly used strategies in benchmarking GRN inference methods:

1. Evaluation on a published scRNAseq dataset. Since the "true" TF-TF interactions underlying the measured biological process are usually unknown, a database of global TF-TF interactions (curated from experiments such as ChIP-seq) is used to create a "ground truth" network. The problem with this approach is that many of the interactions in such a network might not be relevant to the process under study [144].

2. Starting from some TF-TF interaction network, and creating a synthetic dataset from this network. This approach has the advantage that the "ground truth" network is known, which raises the confidence of any evaluation performed using these datasets. The challenge in this approach is to find an appropriate method to create a synthetic scRNAseq dataset starting from a TF-TF interaction network.

Here, we use the second approach. Before we describe the details of our proposed framework for generating a benchmarking dataset, we briefly review other methods that have attempted to use this simulation-based approach.

6.2 Survey of published methods

Before describing the various frameworks proposed to simulate both bulk- and scRNAseq datasets, we will first define some terms that are commonly used in the scRNAseq literature.

Dimensionality reduction This is the process of converting a high-dimensional data set to a low (typically two or three) dimensional dataset to aid in visualization or analysis [145]. An scRNAseq dataset takes the form of a cell×gene matrix. Each sequencing experiment can measure on the order of thousands of cells and hundreds to thousands of genes. Dimensionality reduction using methods such as principal component analysis, t-distributed stochastic neigbor embedding (tSNE) [146], multidimensional scaling (MDS) [147] or uniform manifold approximation and projection (UMAP) [148] are frequently used to make the analysis of these large datasets more tractable.

Trajectory Cells undergoing a biological process can be ordered based on the similarities in their gene expression patterns [149]. This ordering is referred to as a gene expression trajectory. Using trajectory inference methods, the gene expression trajectories can be computed from scRNAseq datasets. Dimensionality reduction of these trajectories typically reveal continuous, branching trajectories, corresponding to the various cellular differentiation processes.

Pseudotime A biological sample may contain cells at different points along a gene expression trajectory, representing different stages of the biological process. Cells can thus be ordered along a trajectory, and can be positioned 'early' or 'late' in the biological process depending on their position in the trajectory. This ordering, analogous to the notion of "real" time in a biological process, is defined as pseudotime.

RNA velocity The single cell resolution of scRNAseq datasets now allows for the differentiation of mature and unprocessed transcripts and their abundances in each cell. This knowledge allows the estimation of relative rates of transcript maturation. In a low dimensional projection of the vector of transcript maturation rates defines a vector for each cell called RNA velocity. Estimation of RNA velocity transforms the static "trajectory" view of low dimensional embeddings to a more dynamic "flux" view, where the future gene expression state of each cell can be estimated.

Importantly, in order for a method to successfully generate a synthetic scRNAseq dataset, a low dimensional representation of such a dataset must exhibit trajectory structures that reflect the traversal of cells from "early" to "late" stages in a process.

GeneNetWeaver (GNW). This is original evaluation framework used in the DREAM4 community challenge for GRN inference methods using bulk RNAseq data [150]. GNW relies on the the background TF-TF interaction networks from *E. coli* and *S. cerevisiae* in order to sample "ground truth" networks of different topologies. These sampled networks are directed graphs. Next, the TF-TF interaction network is represented as a set of coupled Ordinary Differential Equations (ODEs). The form of the equations proposed by GNW is used by other methods, including BOOLODE, and is described in detail in Section 6.3. Importantly, the background networks do not carry any information regarding Boolean logic rules governing the regulation of a given gene. GNW solves this problem by assigning a Boolean rule to a set of regulators. The resulting dynamical model is parameterized by sampling kinetic parameter values from predefined ranges, and this model is simulated as a system of Stochastic Differential Equations. GNW defines the notion of an external perturbation, where a randomly chosen variable in the model receives an input for the first half of the numerical simulation. In summary, GNW relies on random sampling at four stages:

1. Sampling of the background network,
2. Assignment of Boolean logic to combinations of regulators,
3. Sampling of kinetic parameter values,
4. Selection of the variable that receives an external perturbation.

The consequence of these decisions is that it is unclear if the synthetic dataset generated using GNW can correspond to a biologically meaningful time course experiment. Furthermore, there is no reason to believe that the output generated from GNW captures any of the characteristics of scRNAseq datasets, such as discernible trajectories and distinct steady state clusters, even qualitatively.

SPLATTER. Single cell RNAseq datasets are characteristically noisy, and the origins of this noise have been decomposed broadly as biological noise (originating from noisy gene expression, causing cell to cell variability) and technical noise (arising from problems with sequence capture). An understanding of the statistical distributions of these various noise contributions can be used to create synthetic datasets. Zappia et al. [151] combined previous approaches such as Lun, scDD, and BASIcS, along with a novel method called Splat, to create the SPLATTER framework which provides a unified interface to simulate statistically realistic scRNA seq datasets. (For a comprehensive review of other previous statistical approaches, please see Zappia et al., 2017.) The major drawback of using the SPLATTER is that it is not possible to represent complex regulatory relationships between genes. For example, to create trajectories between (user defined) steady state clusters, SPLATTER interpolates between defined gene expression states by using a Brownian bridge. While, this method can artificially create "paths", the variation in gene expression along such trajectories does not capture any regulatory relationship between genes.

DYNGEN. In a recent benchmarking effort, Saelens et al. adapted the GNW framework in order to generate synthetic scRNAseq datasets [152]. The objective of this benchmarking effort was to evaluate the performance of algorithms that create a "pseudotime" ordering of cells in an scRNAseq datasets. In short, each cell in the dataset is ordered based on measures of similiarity defined in gene expression space such that cells with similar expression profiles are closer in the final ordering. In practice, the scRNAseq datasets are first transformed from high dimensional gene expression space into a low dimensional projection using techniques like tSNE or UMAP [153]. Such a projection often reveals "trajectories", namely structures corresponding to the temporal progression through a biological process, for example through a developmental program. A challenge in pseudotime inference is handling branching trajectories. DYNGEN solved the problem of creating synthetic datasets that exhibit diverse trajectory structures by modifying the first stage of the GNW framework: instead of sampling a subnetwork from a background network, DYNGEN relies on a set of manually created "toy" networks, each of which is constructed such that visualizing the simulation output after dimensionality reduction yields trajectories displaying various qualitative structures (linear, linear long, cycle, bifurcating, bifurcating converging and trifurcating, show in Figure 6.1).

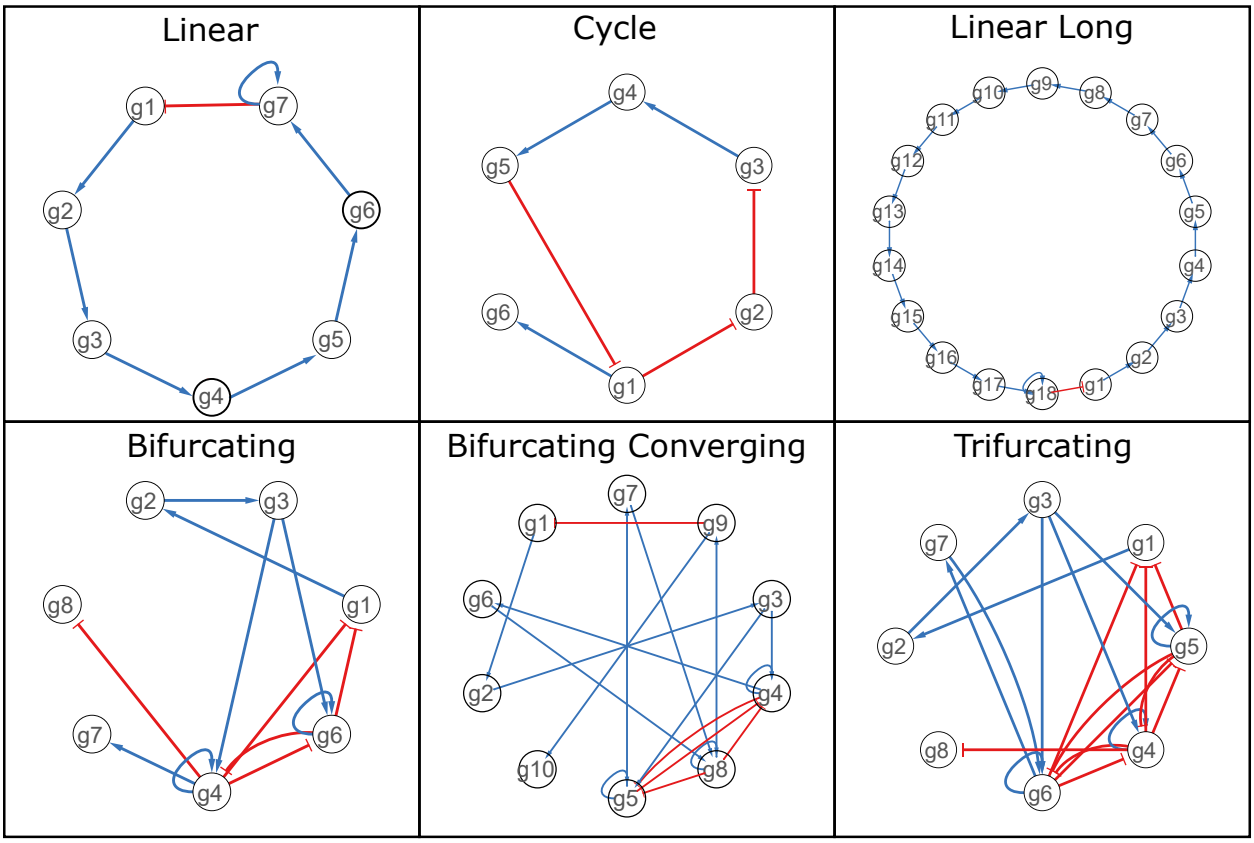


Figure 6.1: The six synthetic networks from DYNGEN. These manually created networks were originally used to create qualitatively different trajectories to benchmark pseudotime inference methods.

Beyond the approach to sampling the initial network, there are other differences in implementation between DYNGEN and GNW:

1. DYNGEN carries out SSA (the Stochastic Simulation Algorithm [154]) simulations, while GNW carries out SDE (Stochastic Differential Equations) simulations.

2. DYNGEN treats the nodes in toy networks as node groups. Each node group can be composed of multiple 'genes', each of which inherit the interactions attributed to the node group. In this manner, DYNGEN can generate regulatory networks with arbitrary numbers of genes while still guaranteeing that the resulting trajectories still display the appropriate qualitative structures.

In terms of applications, DYNGEN claims that the datasets generated using their framework can be used to evaluate methods for predicting RNA velocity, inferring pseudotime, as well as inferring gene regulatory networks. A major drawback of the approach used by DYNGEN is that it assumes a limited set of network topologies, each of which is constructed by hand. Each network is defined by a set of interactions between nodes specified by interaction sign and interaction strength. In the default setting, each node is interpreted as a gene, and the network topology then defines a "ground truth" interaction network. More generally, users can specify arbitrary numbers of genes to be simulated. In this case, each node is interpreted as a module, which can be composed of multiple genes. "Genes" are randomly assigned to each of these modules until the regulatory network has the user specified size. (For more details, please see Saelens et al., 2020.) The rationale for such a module centric approach is based on the observation that real GRNs maybe be composed of gene modules. However the drawback of this approach is that DYNGEN-defined modules are not (and cannot be) exhaustive when it comes to network topologies, and many different topologies can achieve similar trajectories. There is no easy way that DYNGEN provides for adding new network topologies.

SERGIO. Dibaeinia et al. propose SERGIO as a framework in anticipation of benchmarking new methods that estimate RNA velocity from scRNAseq data [155]. This framework is identical to GNW, except instead of representing each node in the initial interaction network as a gene and a protein variable, the node is represented as an unprocessed and mature transcript respectively. Since their approach to sampling subnetworks from the background and creating Boolean rules between regulators is the same as that used by GNW, this approach suffers from the same drawbacks as GNW, even though the model parameters used in SERGIO result in the discernible trajectories in the low dimensional projections.

6.3 Methodology

We start by describing the framework proposed in GNW. This method starts with a network of regulatory interactions among transcription factors (TFs) and their targets. It computes a

connected, dense subnetwork around a randomly selected seed node and converts this network into a system of differential equations. To express this network in the form of Ordinary Differential Equations (ODEs), it assigns each node i in the network a 'gene' variable (x_i) representing the level of mRNA expression and a 'protein' variable (p_i) representing the amount of transcription factor produced by protein translation as follows:

$$\frac{d[x_i]}{dt} = mf(R_i) - l_x[x_i]$$
$$\frac{d[p_i]}{dt} = r[x_i] - l_p[p_i],$$

where m is the mRNA transcription rate, l_x is the mRNA degradation rate, r is the protein translation rate, and l_p is the protein degradation rate. In the first equation, R_i denotes the set of regulators of node i. The non-linear input function $f(R_i)$ captures all the regulatory interactions controlling the expression of node i [150];we specify it below.

If there are N regulators for a given gene, there are 2^N possible configurations of how the regulators can bind to the gene's promoter. Considering cooperative effects of regulator binding, the probability of each configuration $S \in 2^{R_i}$, the powerset of R_i, is given by the following equation [156]:

$$\Pr(S) = \frac{\prod_{q \in S} ([q]/k)^n}{1 + \sum_S \prod_{q \in S} ([q]/k)^n}, \tag{6.1}$$

where k and n are the Hill threshold and Hill coefficient respectively. Here, we use q to denote a regulator in the configuration S. In the summation in the denominator of this equation, the set S ranges over all members of the powerset 2^{R_i} other than the empty set.

The product in the numerator ranges over all regulators that are present (bound) in a specific configuration S, and the sum in the denominator runs over all configurations in the powerset, other than the empty set.

GNW further introduces a randomly-sampled parameter $\alpha_S \in [0, 1]$ to specify the efficiency of transcription activation by a specific configuration S of bound regulators. Thus, the function $f(R_i)$ takes the following form:

$$f(R_i) = \sum_{S \in 2^{R_i}} \alpha_S \Pr(S) \tag{6.2}$$

Next, GNW adds a noise term to each equation to mimic stochastic effects in gene expression [157]. In addition, in order to create variations among individual experimental samples, GNW recommends adopting a multifactorial perturbation [157] that increases or decreases the basal activation of each gene in the GRN simultaneously by a small, randomly-selected value. GNW removes this perturbation after the first half of the simulation. Simulating this system of SDEs generates the requisite gene expression data.

BOOLODE uses Boolean models to create simulated datasets. In order to generate simulated time course data for our analysis, we used the GNW framework with one critical difference and one minor variation. The form of the equations used by BOOLODE is identical to that of GNW. The critical difference is that we do not sample the α_S parameters in Equation 6.2 randomly, i.e., we do not combine the regulators of each gene using a random logic function. Instead, we use the fact that in both the artificial networks and the literature-curated models, we know the Boolean function that specifies how the states of the regulators control the state of the target genes. Moreover, we can express any arbitrary Boolean function in the form of a truth table relating the input states (i.e., activities of transcription factors) to the output state (the activity of target gene). For a gene with N regulators in its Boolean function, we explore all 2^N combinations of transcription factor states and evaluate the transcriptional activity of each specific regulator configuration. Since the value of the Boolean function is the logical disjunction ('or') of all these values, we set the α value to one (respectively, zero) for every configuration that evaluates to 'on' (respectively, 'off'). The following example illustrates our approach. Consider a gene X with two activators (P and Q) and one inhibitor (R), represented by the following rule:

$$X = (P \vee Q) \wedge \neg(R)$$

The truth table corresponding to this rule along with the α parameters appears below.

P	**Q**	**R**	**X**	Parameter name	Parameter value
0	0	0	0	α_0	0
1	0	0	1	α_P	1
0	1	0	1	α_Q	1
0	0	1	0	α_R	0
1	1	0	1	α_{PQ}	1
1	0	1	0	α_{PR}	0
0	1	1	0	α_{QR}	0
1	1	1	0	α_{PQR}	0

Therefore, the ODE governing the time dynamics of gene X is

$$\begin{aligned}\frac{d(X/k)^n}{dt} &= m\left(\frac{\alpha_0 + \alpha_P(P/k)^n + \alpha_Q(Q/k)^n + \alpha_R(R/k)^n + \alpha_{PQ}(P/k)^n(Q/k)^n + \\ \alpha_{PR}(P/k)^n(R/k)^n + \alpha_{QR}(Q/k)^n(R/k)^n + \alpha_{PQR}(P/k)^n(Q/k)^n(R/k)^n}{1 + (P/k)^n + (Q/k)^n + (R/k)^n + (P/k)^n(Q/k)^n \\ + (P/k)^n(R/k)^n + (Q/k)^n(R/k)^n + (P/k)^n(Q/k)^n(R/k)^n}\right) - l_x(X/k)^n \\ &= m\left(\frac{(P/k)^n + (Q/k)^n + (P/k)^n(Q/k)^n}{1 + (P/k)^n + (Q/k)^n + (R/k)^n + (P/k)^n(Q/k)^n \\ + (P/k)^n(R/k)^n + (Q/k)^n(R/k)^n + (P/k)^n(Q/k)^n(R/k)^n}\right) - l_x(X/k)^n,\end{aligned}$$

since only α_P, α_Q, and α_{PQ} have the value one and every other parameter has the value zero.

Next, we discuss the minor variation of BOOLODE from GNW, which is in how we sample kinetic parameters. The GNW equations use four kinetic parameters: one each for mRNA tran-

scription, protein translation, and mRNA and protein degradation rates. Saelens et al. [149] sample them uniformly from parameter-specific intervals.

Independently for every dataset, we sample each parameter from a normal distribution using the value shown in Table 6.1 as the mean and a standard deviation of 10% of this mean value. Within a single dataset and for all simulations for that dataset, we fix each parameter (e.g., mRNA degradation rate) for all genes.

Parameter	Symbol	value
mRNA transcription rate	m	20
mRNA degradation rate	l_x	10
Protein translation rate	r	10
Protein degradation rate	l_p	1
Hill threshold	k	10
Hill coefficient	n	10

Table 6.1: Kinetic parameters used in BOOLODE.

In order to create stochastic simulations, we use the formulation proposed by Saelens et al. [149] to modify the ODE expressions as follows:

$$\frac{d[x_i]}{dt} = mf(R_i) - l_x[x_i] + s\sqrt{[x_i]}\Delta W_t$$
$$\frac{d[p_i]}{dt} = r[x_i] - l_p[p_i] + s\sqrt{[p_i]}\Delta W_t$$
$$\Delta W_t = \mathcal{N}(0, h),$$

where s is the noise strength. We use $s = 10$ in our simulations. We use the Euler-Maruyama scheme for numerical integration of the SDEs with time step of $h = 0.01$.

Defining a single cell. We define the vector of gene expression values corresponding to a particular time point in a model simulation as a single cell. For every analysis, we sample one time point, i.e. one cell from a single simulation. Using this procedure, for a dataset generated from 5000 simulations, we can obtain up to 5000 cells.

Creating GNW simulations for comparison with BOOLODE. In order to simulate a synthetic network using GNW, we used its the edge list as the input network to GNW. In order to create the simulations, we used the default options of the noise parameter (0.05) and multifactorial perturbations. We only performed wildtype simulations and used the DREAM4 time series output format for comparison with the BOOLODE output.

Alternative formulation of activation function using a soft-Heaviside expression. While BOOLODE re-implements the GNW approach of representing the gene activation function f as a combination of Hill functions of its regulators, a potential drawback of this formulation is that the number of terms constituting the converted ODE will grow as $2^{|R|}$, where $|R|$ is the number of regulators of a given gene. This exponential growth in the

number of terms makes this formulation practically infeasible for two reasons when working with Boolean models of TF-TF interaction networks inferred from ChiP-seq experiments or large models of signal transduction pathways (eg [158]).

1. The enumeration of all $2^{|R|}$ states is computationally intensive to the point that even the time for conversion of a system of Boolean rules to the system of ODEs grows exponentially in the number of regulators of genes.

2. The large number of Hill expressions with a high value of the Hill exponent ($n = 10$) contribute to numerically instability.

An alternative approach to representing a Boolean rule as an ODE is to use soft-Heaviside functions [105]. This formulation does not make assumptions about the underlying mechanism of promoter-binding of the regulators, while accomplishing the nonlinear activation dynamics exhibited by the Hill functions. Moreover, the representation of the Boolean rule for each gene does not require the enumeration of each of the $2^{|R|}$ configurations regulators; rather, the Boolean expression is represented directly as a polynomial. The 'AND', 'OR', and 'NOT' relationships would then be represented as follows:

Boolean Operator	Example	Polynomial form
'or'	$x_1 \vee x_2$	$1-(1-x_1)(1-x_2)$
'and'	$x_1 \wedge x_2$	$x_1 x_2$
'not'	$\neg x_1$	$1-x_1$

Using these forms, a Boolean function $\mathbf{B}$ can be represented as a polynomial $P(\mathbf{B})$ In this formulation, Equation (6.2) takes the form[1]

$$f(R_i) = \mathcal{H}(\sigma P(\mathbf{B})) \qquad (6.3)$$
$$\text{where} \quad \mathcal{H}(x) = \frac{1}{1+e^{-x}}$$

Here, σ is a scaling factor that governs the steepness of the response. This formulation guarantees that the polynomial representing the Boolean rule will take values between 0 and 1 as long as the continuous variables x (representing the activities of proteins) are appropriately scaled to this range. Thus, under the appropriate conversion of a Boolean rule to a polyomial, this formulation provides a practical alternative to the Hill function formulation for larger Boolean systems. BOOLODE provides users with the option of using either formulation.

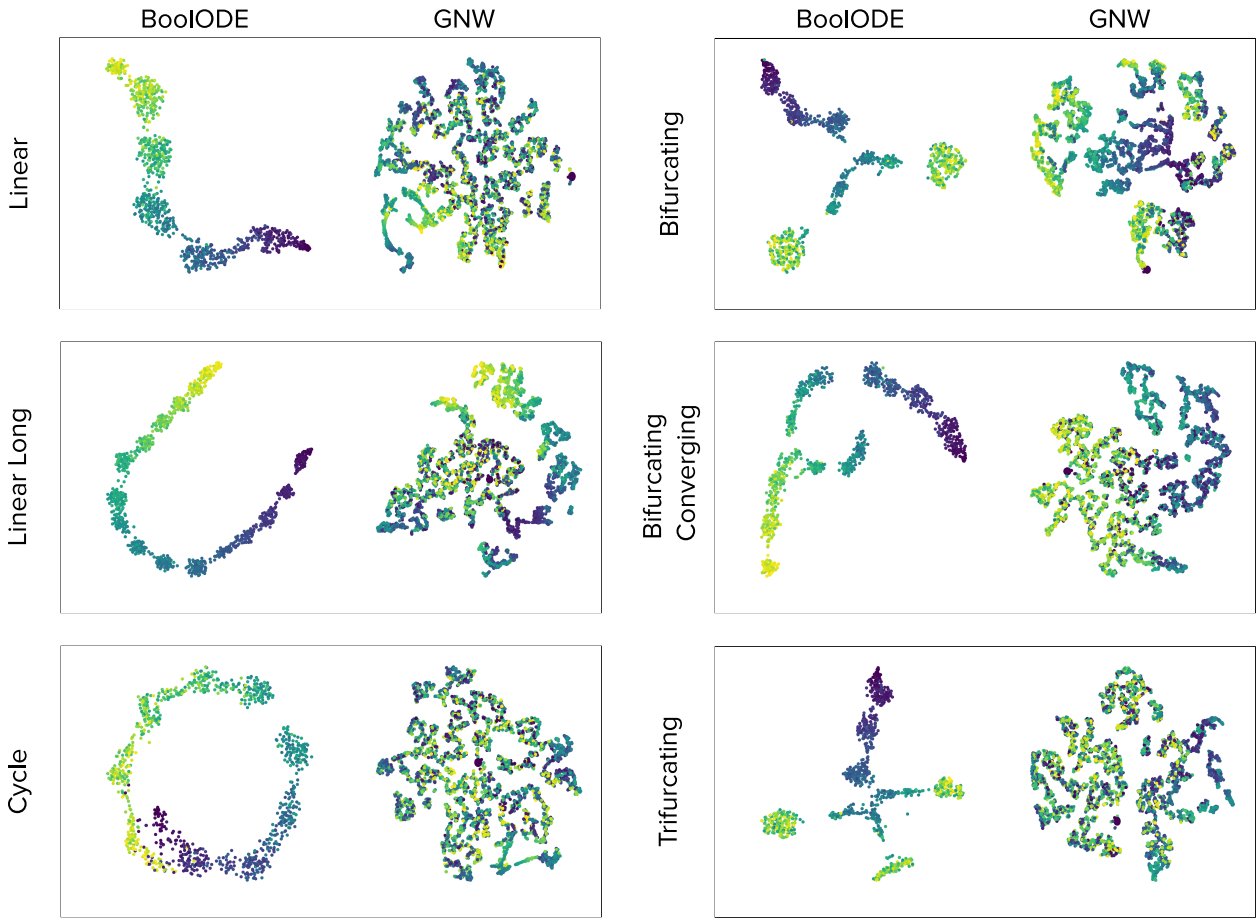


Figure 6.2: Comparison of BoolODEoutput and GNWoutput for the synthetic networks

6.4 Simulation results

Comparison to GNW. We first wished to compare the output from BoolODE and GNW. We used six synthetic networks from DYNGEN, shown in Figure 6.1. The linear and linear-long models display one steady state. The bifurcating and bifurcating-converging models display bifurcations due to a mutual inhibition motif, but exhibit two and one steady states respectively. The trifurcating model exhibits three steady states. The cycle model exhibits a limit cycle. We used the following rule to convert the synthetic networks to Boolean models: the target gene is turned on if all its activators are on and none of the inhibitors are on. We provided BoolODE and GNW with the same Boolean models as input. We ran GNW using the default settings. Finally, we visualized the simulation output from both methods by carrying out dimensionality reduction using tSNE. The comparison is show in Figure 6.2. The dark colors (blue) correspond to early time points in the simulation, while the light colors (yellow) represent late timepoints.

[1]There are various approaches to expressing a Boolean function as a polynomial, cf.[159].

We observe that GNW does not meaningfully capture the qualitative trajectories from the synthetic networks. We attribute this to the fact that GNW uses the interaction network from the input, but assigns random logic to interactions. This result provides a strong motivation to use an approach such as BOOLODE which can capture the qualitative trajectories expected from an underlying network.

Performance on Boolean models. Having shown that BOOLODE can capture qualitative trajectories from synthetic networks, we next investigated if the same holds true for Boolean models whose steady states have been characterized. We used four Boolean models of TF-TF interactions in various developmental contexts curated from the literature. The details of these models are provided in Table 6.2.

Name	**Source**	**# Genes**	**# Steady States**
mammalian Cortical Area Development (mCAD)	Giacomantonio et al., 2010[160]	5	2
Ventral Spinal Cord Development (VSC)	Lovrics et al., 2014 [161]	8	5
Hematopoeitic Stem Cell Development (HSC)	Krumsiek et al., 2011[162]	11	4
Gonadal Sex Determination (GSD)	Rios, et al. 2015 [163]	19	2

Table 6.2: Summary of Boolean models

For each of these models, we first verified if the BOOLODE simulations could capture the appropriate number of steady states as reported in the original publication. Figure 6.3 shows the results of the BOOLODE simulation. In each of the four models, we found that BOOLODE was able to recover the number of steady state clusters as the number of Boolean steady states reported in the original publication. In Figure 6.3 (b), the simulation time is indicated by the color - darker colors indicate early time, and bright (yellow) colors indicate late time. Examining the yellow- green clusters, we find that counting the number of clusters captures the number of steady states expected from each model. In order to confirm the above observation, we additionally carried out k-means clustering on the 'experiments', i.e., each individual time course simulation, with k set to the expected number of steady states. The results of this clustering are shown in Figure 6.3(c). This visualization confirms trajectories originating at the early time points (the initial conditions), which diverge into the final steady state clusters. Finally, we wished to see if the output datasets from BOOLODE can be used as input to trajectory inference methods. For this, we picked the best-performing trajectory inference method, SlingShot, from a recent benchmarking study [149]. Figure 6.3(d) shows the results of running the SlingShot pseudotime computation method on the datasets generated from the Boolean models. We find that the pseudotime ordering computed by SlingShot is in good agreement with the true simulated time points.

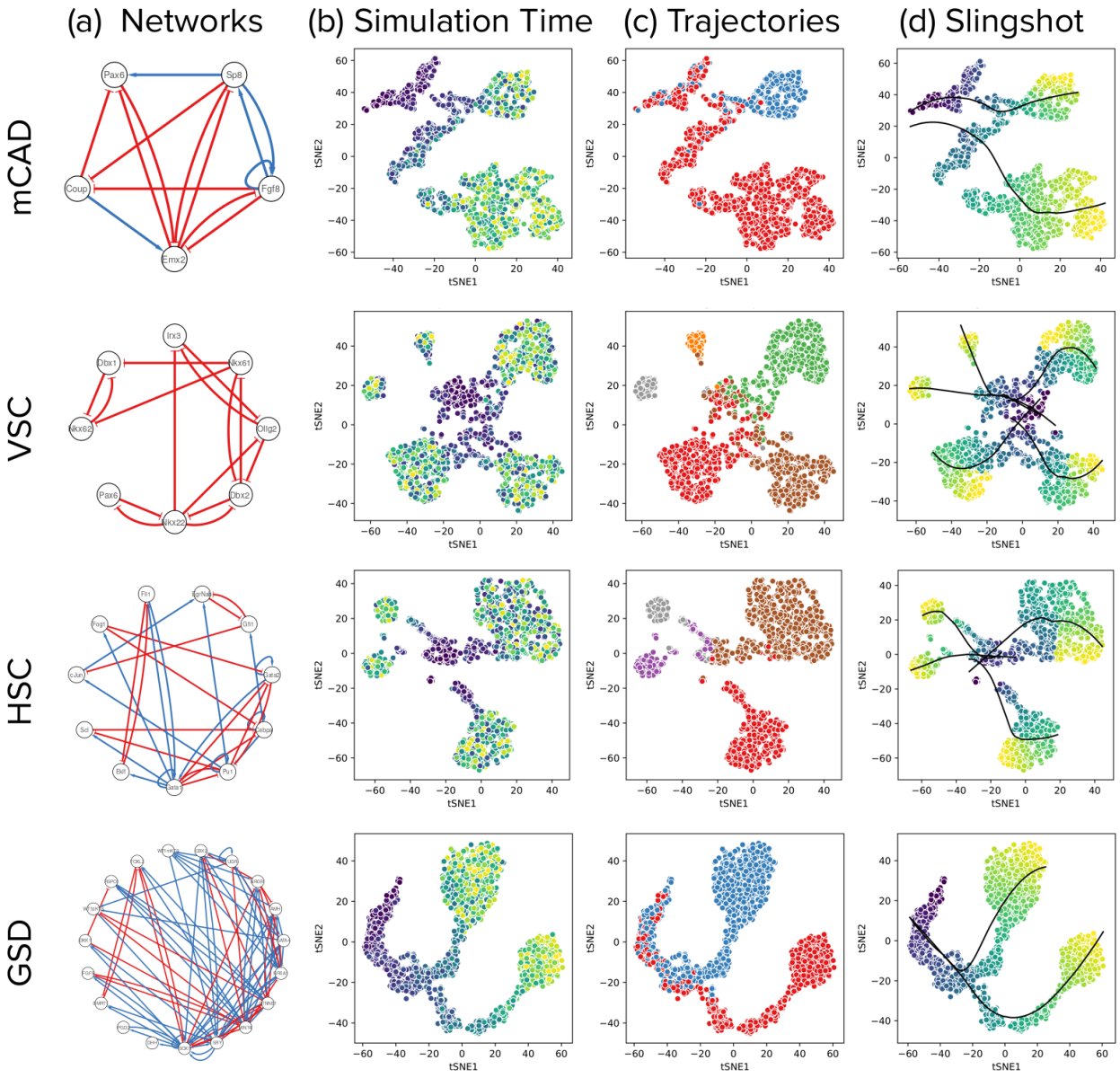


Figure 6.3: BoolODE output from published Boolean models. (a) The interaction network corresponding to the Boolean model. (b) tSNE visualization of BoolODE output. (c) Simulations clustered by the expected number of steady states. (d) Slingshot output on the dataset generated by BoolODE. mCAD is the mammalian Cortical Area Development model, VSC is the Ventral Spinal Cord development model, HSC is the Hematopoeitic stem cell development model, GSD is the Gonadal Sex Determination model.

Correspondence of steady states. Having observed that the number of steady state clusters from BoolODE simulations recapitulated the number of steady states under asynchronous Boolean updates, we next asked whether the steady state clusters corresponded

with the biologically meaningful steady states reported in the Boolean models. Specifically, we wished to see if the qualitative gene expression patterns (low or high expression) in the steady state clusters corresponded to the Boolean gene expression patterns from the Boolean model steady states (OFF or ON, respectively). We plotted the strength of gene expression for each gene in a model. The results are shown in Figures 6.4 to 6.7. We find that the gene expression patterns reflect those reported in the original Boolean models in every case.

Application to gene regulatory network inference algorithms. The BOOLODE framework was used in the BEELINE benchmarking study to generate simulated scRNAseq datasets. Datasets of varying number of cells, gene expression dropouts, and shuffled pseudotimes were generated as benchmarking datasets from the synthetic and boolean models described above. These datasets were used as input to various GRN inference algorithms. The major conclusion of the benchmarking effort was that the performance of most methods was close to random.

6.5 Outlook

This chapter has outlined a novel framework for harnessing expert knowledge on biological processes expressed as Boolean models in order to automatically generate simulated datasets. While this framework currently focuses on Boolean models of TF-TF interaction networks, models of other molecular interaction processes can easily be used.

Some of the ways in which BOOLODEcan be improved are:

- BOOLODE currently carries out SDE simulations. However, stochastic chemical reactions are more realistically simulated using the SSA algorithm [154]. BOOLODE can be extended to use this simulation framework.
- There are various formulations to convert a set of Boolean rules to ODE equations. BOOLODE currently offers a Hill function and a soft-Heaviside function formulations. This framework can be extended by incorporating other approaches to representing Boolean expressions.
- An observation from Figures 6.4 to 6.7 is that the number of cells in the steady state clusters varies. This is likely an outcome of the model parameterization. The kinetic parameters used in BOOLODE are currently fixed. The dependence of the size of steady state clusters on the kinetic parameters remains to be investigated.

BOOLODE is available at https://github.com/murali-group/BoolODE

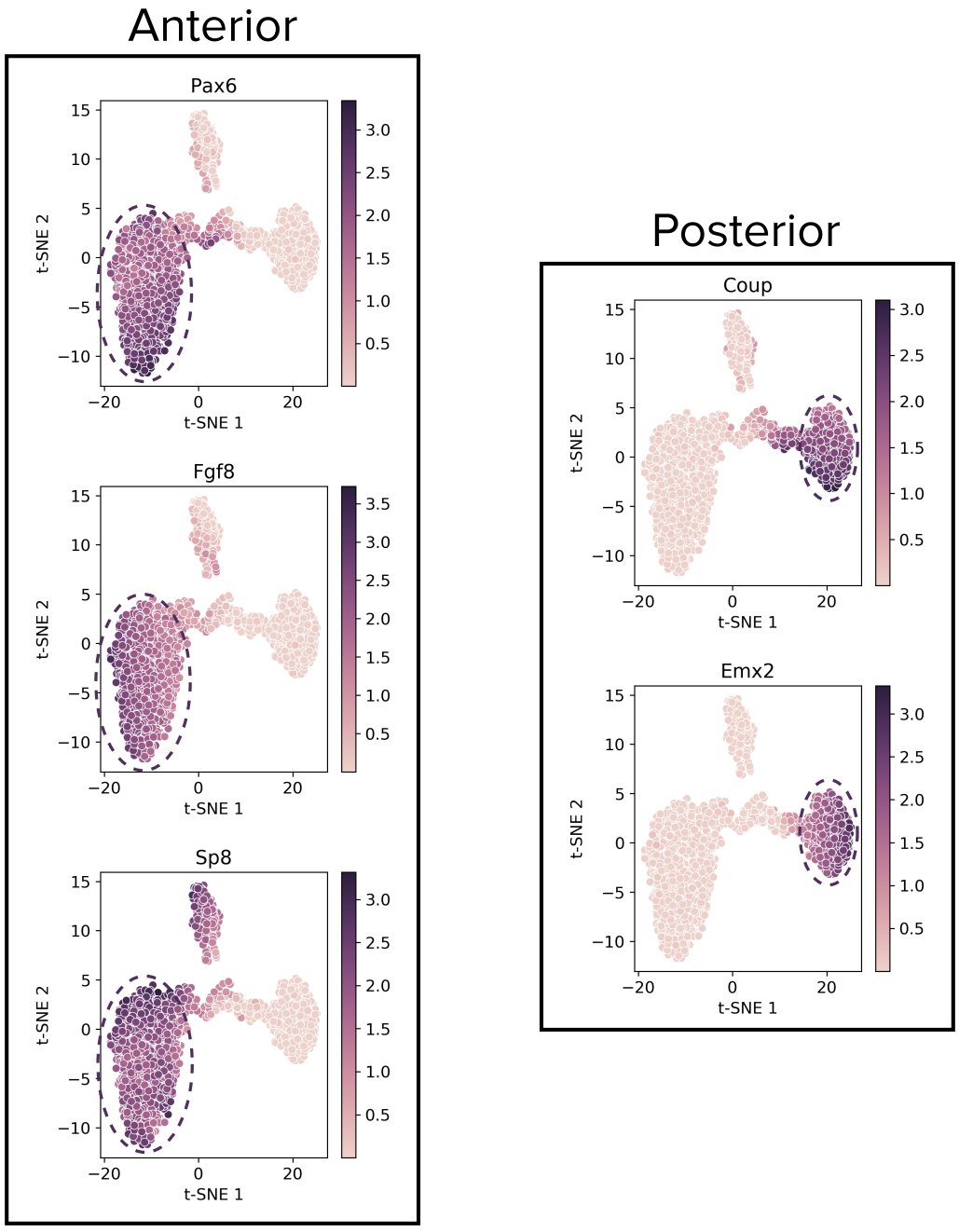


Figure 6.4: t-SNE visualizations of simulated trajectories from the mCAD model. The bar on the right of each plot displays the mapping between the expression value of a gene and the corresponding color. The model has two steady states corresponding to the anterior and posterior compartments during cortical area development[160]. The anterior compartment is characterized by high transcriptional activity of Fgf8, Pax6 and Sp8, while the posterior compartment is characterized by the high transcriptional activity of Coup-tfi and Emx2. The steady states corresponding to the two compartments appear as two distinct clusters in the t-SNE visualizations.

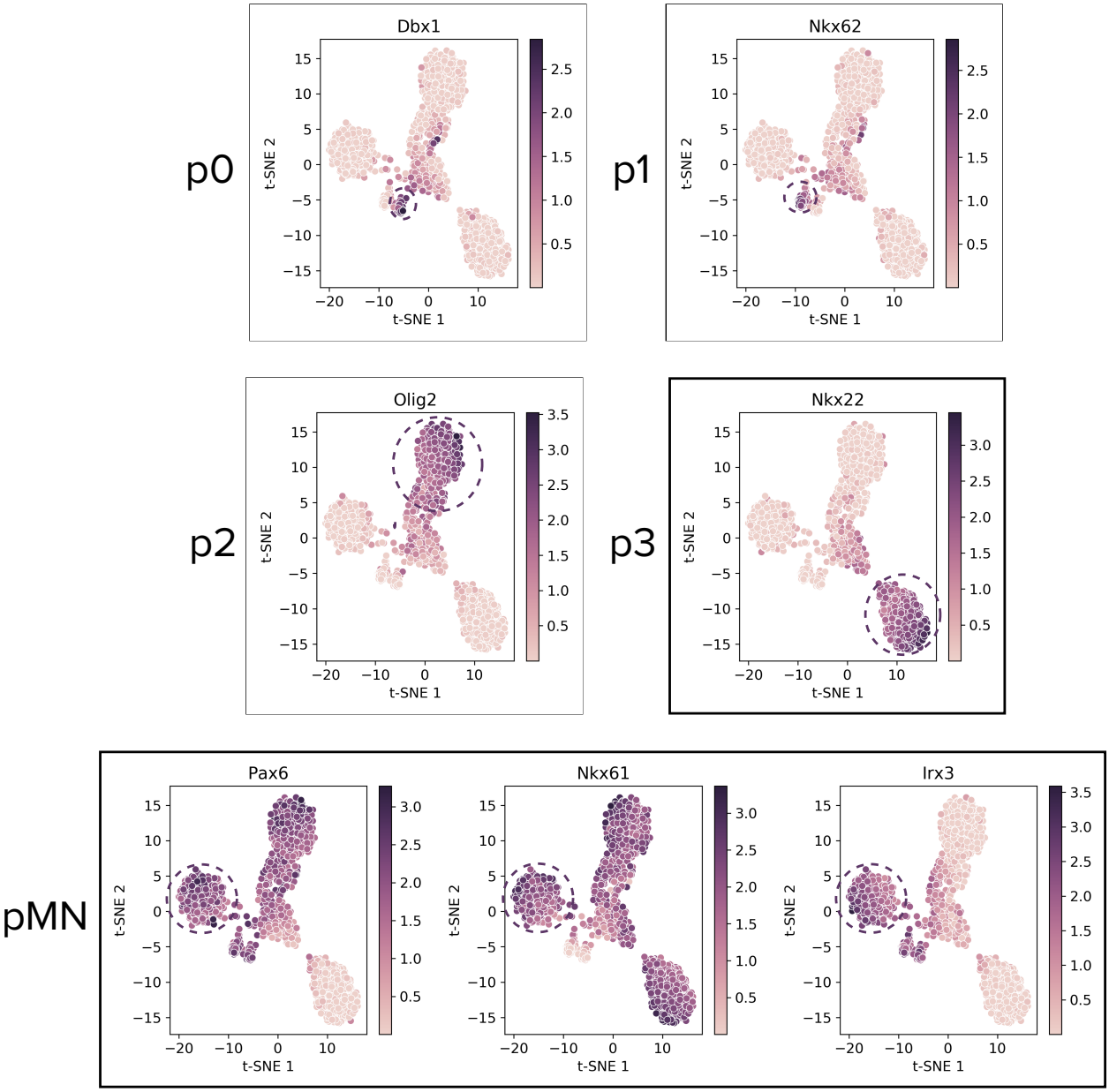


Figure 6.5: t-SNE visualizations of simulated trajectories from the VSC model. This model exhibits five distinct steady states corresponding to five distinct cell types in the neural tube[161]. Each of these cell types appears as a distinct cluster in the tSNE visualizations shown above. The transcription factors specific to the p0 (Dbx1), p1 (Nkx6.2), p2 (Olig2), and p3 (Nkx2.2) states are shown in the four plots on top. The pMN state shown in the lower plot is characterized by high activity of Nkx6.1, Pax6 and Irx3.

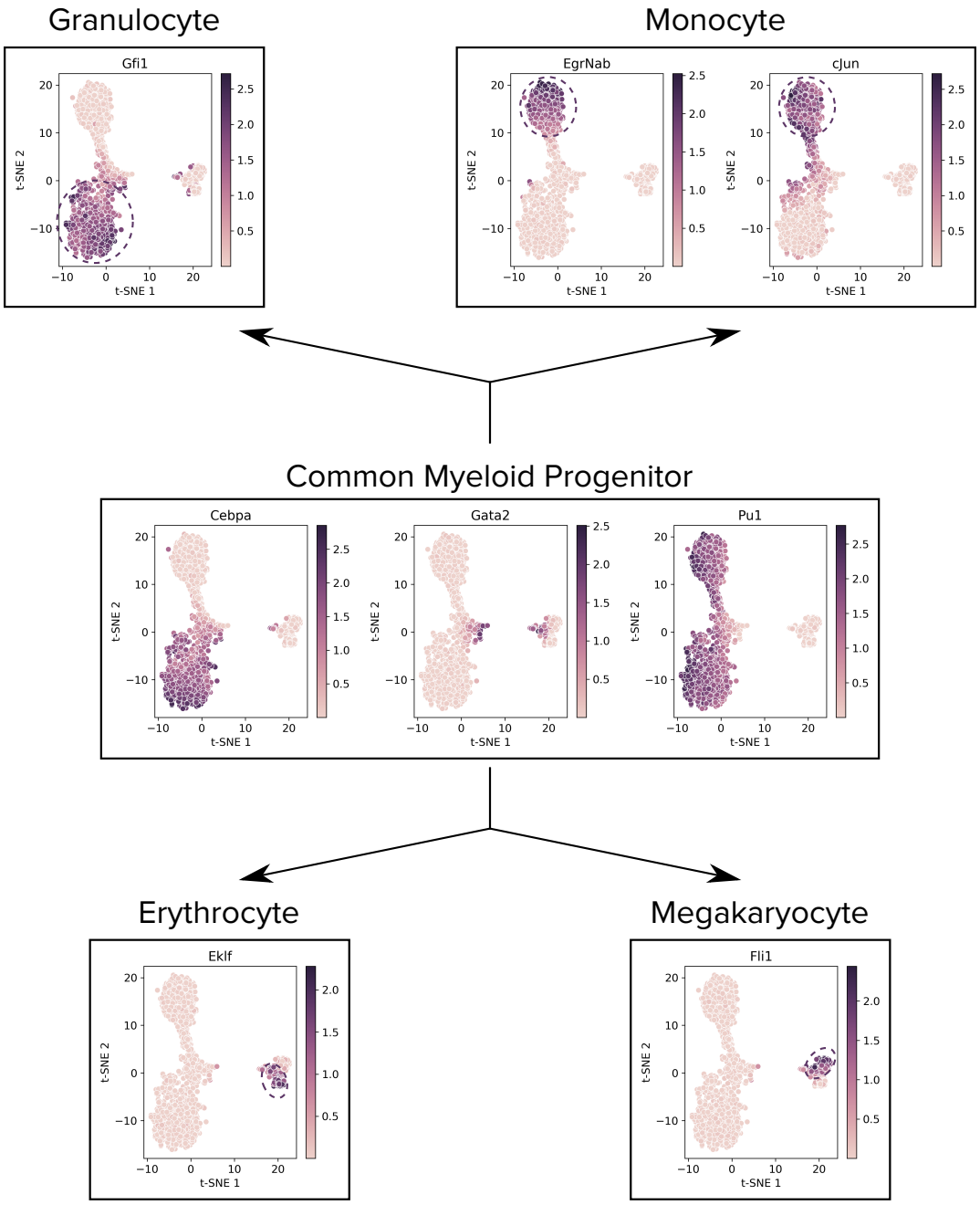


Figure 6.6: t-SNE visualizations of simulated trajectories from the HSC model. In the initial state, which corresponds to the Common Myeloid Progenitor (CMP), the activities of GATA-2. PU.1 and C/EBPα are high[162]. There are two main branches of differentiation. The branch leading to the monocyte-granulocyte progenitors is characterized by the inactivation of GATA-2, followed by the activation of either Gfi1 (granulocyte) or cJun and EgrNab (monocyte) [162]. The clusters in the t-SNE visualization that correspond to these states are shown in the plots on top. Similarly, the branch leading to the megakaryocyte-erythrocyte progenitors is characterized by the inactivation of C/EBPAα and PU.1 along with the concomitant activation of GATA1, FOG1 and SCL [162]. The subsequent inactivation of GATA2 along with activation of Fli1 leads to the megakaryocyte state while that wit the activation of EKLF leads to the erythrocyte state [162]. These states are shown at the bottom.

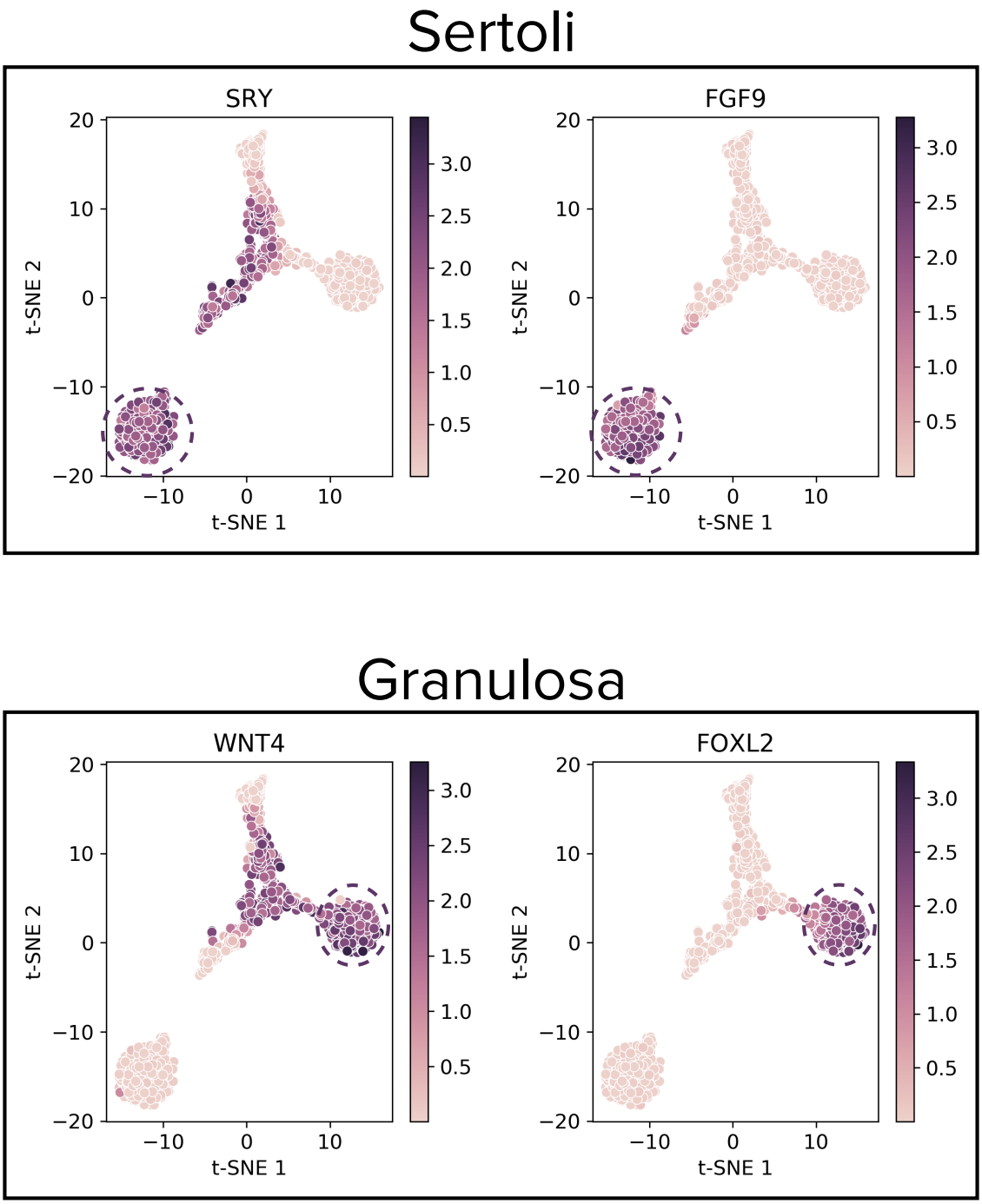


Figure 6.7: t-SNE visualizations of simulated trajectories from the GSD model. This model has two steady states corresponding to Sertoli cells (male gonads) or Granulosa cells (female gonads)[163]. The clusters on the tSNE plots correspond to these two states. SRY and FGF9 show male-gonad-specific activity while WNT4 and FOXL2 show female gonad-specific activity. Except for GATA4 and WT1mKTS, which are active in both branches, all other transcription factors in the model belong to one of the two clusters shown above.

Appendices

Appendix A

Model Parameters

Name	Value	Name	Value	Name	Value
Cyr1_T	1.00e+0	k_degr	8.98e-2	w_gis	1.30e+0
Dot6_T	1.00e+0	k_mRNA_degr	7.30e-2	w_gis_pka	3.30e+0
EGO_T	1.00e+0	k_pr	2.02e-2	w_gis_sch	8.42e-1
EGOGAP_T	1.00e+0	k_transcription	2.36e-1	w_gln_sit	8.61e-1
Gcn2_T	1.00e+0	sigma_cyr	3.50e+0	w_gln_snf	3.90e+0
Gcn4_T	1.00e+0	sigma_dot	2.00e+1	w_gln1	2.22e-1
Gis1_T	1.00e+0	sigma_ego	5.00e+0	w_gln1_gln3	5.20e-1
Gln1_T	1.00e+0	sigma_eif	1.00e+0	w_gln3	6.39e-1
Gln3_T	1.00e+0	sigma_gap	1.00e+0	w_mig	1.06e+1
Mig1_T	1.00e+0	sigma_gcn2	2.00e+1	w_mig_pka	2.31e+0
PDE_T	1.00e+0	sigma_gcn4	5.00e+0	w_mig_snf	1.21e+0
PKA_T	1.00e+0	sigma_gis1	1.00e+1	w_pde	3.83e-1
Ras_T	1.00e+0	sigma_gln	1.00e+1	w_pde_pka	2.89e+0
Rtg13_T	1.00e+0	sigma_gln1	1.00e+0	w_pka	5.81e-2
Sak_T	1.00e+0	sigma_mig1	2.70e-1	w_pka_camp	1.02e+2
Sch9_T	1.00e+0	sigma_pde	1.90e+0	w_pka_sch9	1.75e+1
Snf1_T	1.00e+0	sigma_pka	1.00e+0	w_ras	2.08e-2
TORC1_T	1.00e+0	sigma_ras	1.00e+0	w_ras_glu	2.07e-1
Tps1_T	1.00e+0	sigma_rtg	1.00e+1	w_ras_pka	1.87e+0
Trehalase_T	1.00e+0	sigma_sak	2.00e+1	w_rtg	1.86e-1
eIF_T	1.00e+0	sigma_sch9	8.00e+0	w_rtg_torc	8.77e-1
gamma_gcn2	4.71e+0	sigma_snf	3.00e+0	w_sak	2.05e-1
gamma_mig	6.56e-1	sigma_tor	5.00e+0	w_sak_pka	3.75e-1
gammacyr	8.96e+0	sigma_tps	5.00e+0	w_sch9	5.65e-1
gammaego	5.07e+1	sigma_trehalase	1.00e+1	w_sch9_torc	1.96e+0
gammaeif	4.71e-1	tRNA_scale	7.45e+1	w_snf	5.38e-1
gammagap	5.62e-1	tRNA_total	2.47e+0	w_snf_glc	1.15e+0
gammagln1	6.35e-2	w_cyr	1.35e+0	w_snf_sak	1.52e+0
gammagln3	8.09e-2	w_cyr_glu	5.13e+0	w_torc	5.39e-1
gammapde	2.82e-1	w_cyr_snf	1.19e-1	w_torc_ego	8.77e-1
gammapka	2.68e+0	w_dot	2.93e-1	w_torc_egoin	3.03e-1
gammaras	1.82e+0	w_dot_sch_pka	1.63e-1	w_torc_glut	8.63e-1
gammasch9	4.63e+0	w_ego	2.84e-1	w_torc_snf	4.37e-1
gammasnf	8.20e-1	w_ego_basal	1.10e-2	w_tps	5.30e-2
gammator	7.55e+0	w_ego_gap	2.21e+0	w_tps_pka	5.74e-1
gammatps	4.71e-1	w_eif	3.73e+0	w_tre	1.07e+0
gammatre	3.42e-1	w_eif_gcn2	2.76e-1	w_tre_pka	3.07e+0
k_acc_glu	4.92e-2	w_gap_N	7.76e+0	ATP	1.0
k_acc_nh4	1.47e-3	w_gap_torc	8.83e+1	Carbon	1.0
k_acc_pro	2.15e-4	w_gcn	1.15e-1	Glutamine_ext	1.0
k_camp_cyr	1.09e+1	w_gcn_torc	1.29e+0	NH4	0.00e+00
k_camp_deg	8.38e-2	w_gcn4	7.43e-1	Proline	0.00e+00
k_camp_pde	1.41e+1	w_gcn4_gcn2_trna	1.53e+0		

Table A.1: Parameter values constituting the optimal parameter set, obtained from 10,000 steps of MCMC sampling. This set of values is used to define the simulation of a *wt* strain

Appendix B

Model Calibration

Source	tunits	vunits	Time	Values
Fig3 [120]	min	Kinase (nmol min-1 g-1)	[0.39, 5.55, 30.56, 60.64]	[0.03, 0.26, 0.32, 0.302]
Fig2 [73]	s	cAMP nmol/(g wet wt)	[0.16, 15.27, 30.66, 44.88, 59.95, 74.84, 89.72, 104.2, 119.75, 149.39, 180.06]	[0.079, 0.632, 0.822, 1.197, 1.227, 0.928, 0.617, 0.617, 0.632, 0.209, 0.282]
Fig2 [73]	s	cAMP nmol/(g wet wt)	[0.15, 14.97, 30.09, 44.76, 59.71, 74.74, 89.51, 104.54, 119.59, 149.79, 180.28]	[0.052, 0.112, 0.67, 0.259, 0.484, 0.375, 0.264, 0.279, 0.301, 0.275, 0.259]
Fig4B [113]	min	%Sch9 phosphorylation	[0.06, 0.36, 0.73, 1.05, 3.07, 4.12, 5.06, 8.05, 11.03, 15.07, 30.06]	[5.92, 10.23, 34.29, 41.14, 40.59, 39.82, 32.99, 13.09, 17.25, 35.4, 52.57]
Fig4B [113]	min	%Sch9 phosphorylation	[0.46, 1.09, 1.57, 2.04, 3.01, 4.05, 5.03, 8.05, 11.04, 15.01, 29.98]	[10.23, 49.55, 45.84, 27.66, 6.55, 5.49, 5.22, 5.27, 4.82, 4.89, 5.23]
Fig6B [113]	min	%Sch9 Phosphorylation	[0.15, 0.7, 1.04, 1.51, 2.06, 3.05, 4.04, 5.06, 8.06, 11.0, 15.04, 30.02]	[7.96, 8.48, 8.48, 7.44, 7.44, 7.31, 7.44, 7.96, 9.64, 12.49, 25.7, 51.72]
Fig1F [88]	min	Rel. Sch9 Phosphorylation	[0.0,2.5,5.0,15.0,30.0]	[1.0,0.34,0.13,0.15,0.20]
Fig1F [88]	min	Rel. Sch9 Phosphorylation	[0.0,2.0,5.0,15.0,30.0]	[0.20,0.63,0.92,0.91,1.0]
Fig1A [99]	s	log(nMig1/cMig1)	[0.0, 67.0, 104.0, 130.0, 233.0, 334.0, 466.0, 606.0, 771.0, 919.0, 1088.0]	[1.070, 1.233, 1.363, 1.428, 1.490, 1.465, 1.432, 1.425, 1.416, 1.400, 1.402]
Fig2B [164]	min	Rel. RPL32 mRNA	[0.0,15.0,30.0,60.0,90.0]	[1.00,0.43,0.19,0.09,0.07]
Fig3A [109]	min	Gln3P (1-Gln3 from Image)	[0,5,10,15,20]	[0.0, 1048.5, 7846.2, 9644.2, 12018.4]

Table B.1: Literature curated time series data used for model calibration. The table mentions the source of the data from the respective publication, and time points at which measurements were carried out, the measured values along with their units, and the maximum or minimum value of the measured species as mentioned in other parts of the original publication.

Description	WT (pre)	wt (post)	Mut (pre)	Mut (post)	Source	Units
rapamycin	1214.134	38081.039	38081.039	38081.039	Fig 3A [109]	pixels free Gln3 pull-down
$\Delta sch9$	25	700	10	100	Fig 6B [73]	% CTT1 gene expression
$\Delta sch9$	600	150	150	150	Fig 9B [73]	% RPL25 expression
$\Delta gln3$	0.04	1.55	0.03	0.14	Table 2 [16]	Glutamine Synthase sp act
$\Delta snf1$	0.026	0.35	0.026	0.026	Fig 1 [120]	Kin activity nmol min-1 mg-1
$\Delta gcn2$	0.0	1.0	0.0	0.0	[165]	
$\Delta gcn4$	0.0	1.0	0.0	0.0	[166]	
$\Delta rtg1$	43	320	30	30	Table 2 [167]	beta galactosidase nmol/min per mg
Δpde	1.1	0.25	3.8	0.25	Fig 1A [54]	cAMP nmol/g wet wt
$\Delta tpk3$	100	25	50	25	Fig 5 [20]	Trehalase sp act nmol/(min mg prtn)
$\Delta ras1/2$	1.1	0.25	0.25	0.25	Fig1 [53]	Assumed cAMP nmol/g wet wt
$\Delta gtr1/2$ transient	98	2	18	22	Fig 1 [168]	% TORC1 activity as Sch9-P
$\Delta lst4/7$ transient	98	2	21	1	Fig 1 [168]	% TORC1 activity as Sch9-P
$\Delta sak1$ $\Delta tos3$ $\Delta elm1$	0.05	2.0	0.05	0.05	Fig 2C [63]	Kin activity nmol min-1 mg-1
rapa:$\Delta sit4$	2221	5311	0	891	Fig 3 [109]	GST-URE2 Pull-down
$\Delta gtr1/2$ steady state	55	5	55	5	Fig 1 [113]	% TORC1 activity as Sch9-P

Table B.2: Literature curated perturbation data used to calibrate the model. The table contains the citation and figure from which data is extracted. The data includes the measured activity in the units of the original publication at steady state before and after a shift or a treatment, which is then min-max normalized so that the values can be compared to model predictions.

Appendix C

Ranked parameters in stiff directions

#	Parameter	Rank	Data?
1	**w_sch9**	1	Yes (T,P)
2	**w_cyr_glu**	1	No
3	**w_cyr**	1	No
4	**k_degr**	1	Yes (T,P)
5	**w_sch9_torc**	1	Yes (T,P)
6	**w_torc_snf**	1	No
7	**w_torc_glut**	1	No
8	**w_torc_egoin**	1	No
9	**k_acc_glu**	1	No
10	**w_ego**	1	No
11	**w_torc_ego**	1	No
12	**w_torc**	1	No
13	**w_gln3**	2	Yes (T,P)
14	**k_camp_pde**	2	Yes (T,P)
15	**k_camp_cyr**	2	Yes (T,P)
16	**w_ras_pka**	2	No
17	**w_mig**	2	Yes (T,P)
18	**w_gln_sit**	3	Yes (T,P)
19	**gammagap**	4	No
20	**w_ego_gap**	4	No
21	**w_snf_glc**	6	Yes (T,P)
22	**w_snf**	6	Yes (T,P)
23	**w_gap_torc**	6	No
24	**w_sak_pka**	6	No
25	**w_sak**	6	No
26	**w_pka_sch9**	6	No
27	**w_pka_camp**	7	No
28	**w_gcn4**	8	Yes (P)
29	**w_gis**	8	Yes (P)
30	**w_ego_basal**	8	No
31	**w_gis_pka**	8	Yes (P)
32	**w_ras**	8	No
33	**gamma_gcn2**	8	No
34	**w_rtg**	8	Yes (P)
35	**k_transcription**	8	No
36	**k_mRNA_degr**	8	No
37	**gammasnf**	8	Yes (T,P)
38	**gammasch9**	8	Yes (T,P)
39	**gammapde**	8	No
40	**gammagln1**	8	No
41	**gammaeif**	8	No
42	**gammaego**	8	No
43	**gammacyr**	8	No
44	**w_tps**	8	No
45	**w_gln1_gln3**	8	No
46	**w_snf_sak**	9	Yes (T,P)
47	**w_pde**	9	No
48	**w_mig_snf**	9	Yes (T)
49	**w_ras_glu**	9	No
50	**w_tre_pka**	9	Yes (P)
51	**w_gap_N**	9	No
52	**w_tre**	9	Yes (P)
53	**w_gcn**	10	No
54	**w_dot**	10	No
55	**gammatps**	10	No
56	**w_pde_pka**	10	No
57	**w_dot_sch_pka**	10	No
58	**w_gcn_torc**	10	No
59	**w_gcn4_gcn2_trna**	10	Yes (P)
60	w_pka	11	No
61	k_acc_nh4	11	No
62	gammaras	11	No
63	k_camp_deg	11	Yes (T,P)
64	tRNA_sensitivity	11	No
65	gammator	11	No
66	w_gln_snf	11	Yes (T,P)
67	w_eif	11	No
68	gamma_mig	11	Yes (T)
69	w_cyr_snf	12	No
70	w_eif_gcn2	12	No
71	w_tps_pka	12	No
72	w_gln1	12	No
73	gammatre	13	Yes (P)
74	gammagln3	14	Yes (T,P)
75	w_gis_sch	14	Yes (P)
76	k_acc_pro	14	No
77	tRNA_total	14	No
78	gammapka	16	No
79	w_mig_pka	16	Yes (T)
80	w_rtg_torc	18	Yes (P)
81	k_pr	20	No

Table C.1: *(next page)*

Table C.1: Ranked list of parameters for the 'Full' parameter run with 81 parameters. The 'Data?' column indicates whether or not the equation in which a parameter appears is contrained by any type of data. If it is constrained, the column further indicates the type of data, 'T' for timecourses or 'P' for perturbations. The parameters highlighted in bold indicate those appearing in the first 10 stiff directions.

#	Parameter	Rank	Data?
1	$\mathbf{w_{ego}}$	1	No
2	$\mathbf{w_{torcsnf}}$	1	No
3	$\mathbf{w_{torcglut}}$	1	No
4	$\mathbf{w_{torcegoin}}$	1	No
5	$\mathbf{w_{torcego}}$	1	No
6	$\mathbf{k_{accglu}}$	1	No
7	$\mathbf{w_{sch9torc}}$	1	Yes (T,P)
8	$\mathbf{k_{camppde}}$	2	Yes (T,P)
9	$\mathbf{w_{glnsit}}$	2	Yes (T)
10	$\mathbf{w_{egogap}}$	2	No
11	$\mathbf{w_{raspka}}$	2	No
12	$\mathbf{w_{snfsak}}$	2	Yes (T,P)
13	**gammagap**	2	No
14	$\mathbf{w_{cyrglu}}$	2	No
15	$\mathbf{w_{snfglc}}$	4	Yes (T,P)
16	$\mathbf{w_{pkasch9}}$	5	No
17	$\mathbf{w_{pkacamp}}$	6	No
18	$\mathbf{w_{gcn4gcn2trna}}$	6	Yes (P)
19	$\mathbf{w_{sakpka}}$	7	No
20	$\mathbf{w_{pdepka}}$	7	No
21	$\mathbf{w_{gln1gln3}}$	7	No
22	$\mathbf{w_{gapN}}$	7	No
23	$\mathbf{w_{eifgcn2}}$	7	No
24	$\mathbf{w_{rtgtorc}}$	7	Yes (P)
25	$\mathbf{gamma_{gcn2}}$	7	No
26	**gammacyr**	7	No
27	**gammaras**	7	No
28	**gammasch9**	7	Yes (T,P)
29	$\mathbf{k_{campdeg}}$	7	Yes (T,P)
30	$\mathbf{w_{gispka}}$	8	Yes (P)
31	**gammatps**	9	No
32	$\mathbf{gamma_{mig}}$	9	Yes (T)
33	**gammator**	9	No
34	$\mathbf{w_{gissch}}$	9	Yes (P)
35	$\mathbf{w_{trepka}}$	9	Yes (P)
36	$\mathbf{k_{accnh4}}$	10	No
37	$w_{dotschpka}$	11	No
38	gammagln3	11	Yes (T,P)
39	gammapka	11	No
40	$tRNA_{total}$	11	No
41	w_{cyrsnf}	11	No
42	w_{glnsnf}	11	Yes (T,P)
43	gammasnf	13	Yes (T,P)
44	w_{migpka}	13	Yes (T,P)
45	gammaeif	13	No
46	w_{rasglu}	14	No
47	gammaego	14	No
48	$w_{gcntorc}$	15	No
49	gammagln1	15	No
50	$tRNA_{sensitivity}$	15	No
51	w_{migsnf}	17	Yes (T,P)
52	$k_{transcription}$	18	No
53	w_{tpspka}	18	No
54	k_{accpro}	19	No
55	gammatre	19	Yes (P)
56	gammapde	22	No

Table C.2: *(next page)*

Table C.2: Ranked list of parameters for the 'Partial' parameter run with 56 parameters. The 'Data?' column indicates whether or not the equation in which a parameter appears is contrained by any type of data. If it is constrained, the column further indicates the type of data, 'T' for timecourses or 'P' for perturbations. The parameters highlighted in bold indicate those appearing in the first 10 stiff directions.

Appendix D

Comparison to qualitative experimental data

This section records the curated qualitative experiments, along with the model simulations using the optimal parameter set. Each experiment has a unique experiment ID, a table containing the interpreted and predicted states of transcription factors, along with the simulated values at steady state, a description of the experiment containing a reference to the original publication, the strain used, and the experiment performed, and the parameters used to represent the strain and the shift experiment in the model. Finally, the simulated time courses of the six model readouts and the interpretation of the simulation are described.

As described in Section 5.6, we use two definitions of rapamycin experiments. The sample simulations here use the Gln3Gcn4 definition of rapamycin treatment. Experiments involving rapamycin treatment are thus indicated using the text **Readout used is Gln3 Gcn4**. For the general trend of predictions using the alternate definition using Dot6, please see Table 5.2. (Note that the results shown here only correspond to one parameter set of our collection of 18,000 alternate sets of parameter values.)

D.1 1-rho0

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.487
Dot6	-	-	0.312
Gcn4	-	-	0.024
Rtg13	Off	Off	0.274
Gln3	-	-	0.983

Description: Liu et al, 1999 studied a *rho0* strain (PSY142 ρ^0) grown in YP + 5% glucose.

Representation:

Preshift Parameters		*Postshift Parameters*	
ATP	0.1	ATP	1.0
Carbon	0.1	Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Initial Conditions

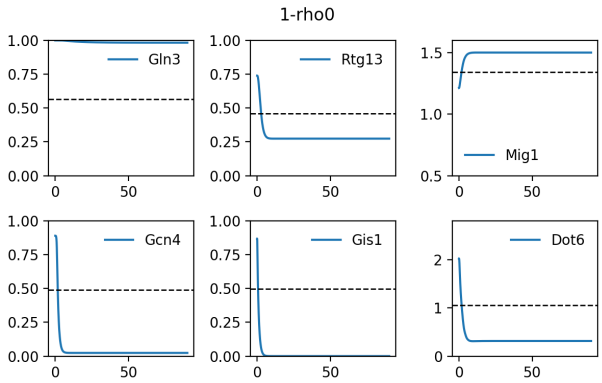


Figure D.1: Qualitative state comparison - *rho*^{0} petites in YP + 5% glucose

Mutant definition

Parameters	
k_{accglu}	0.0369
k_{accpro}	0.0002
k_{accnh4}	0.0011

Initial conditions

Model agrees with experiment.

D.2 2-rho0

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.482
Dot6	-	-	1.121
Gcn4	-	-	0.722
Rtg13	On	On	0.683
Gln3	-	-	0.999

Description: Liu et al, 1999 studied a *rho0* strain (PSY142 ρ^0) grown in YP + 2% raffinose.

Representation:

Preshift Parameters		*Postshift Parameters*	
ATP	0.1	ATP	0.7
Carbon	0.1	Carbon	0.7
$Glutamine_{ext}$	1.0	$Glutamine_{ext}$	0.5

Postshift Initial Conditions

Mutant definition

Parameters	
k_{accglu}	0.0369
k_{accpro}	0.0002
k_{accnh4}	0.0011

Initial conditions

Model agrees with experiment.

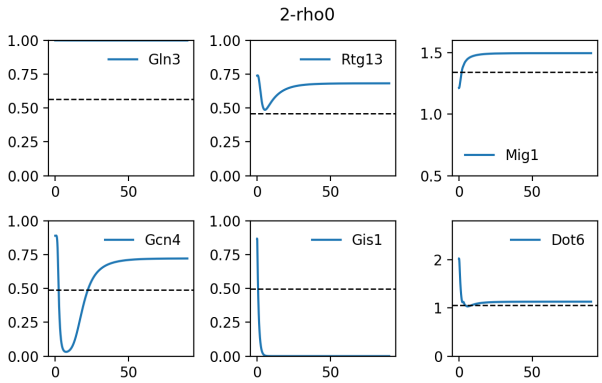


Figure D.2: Qualitative state comparison - ρ^0 in YP + 2% raffinose

D.3 3-rtg1

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.498
Dot6	-	-	0.532
Gcn4	-	-	0.024
Rtg13	Off	Off	0.279
Gln3	-	-	0.992

Description: Liu et al, 1999 studied a *rtg1* strain (PSY142 ρ^0) grown in YNBD + 0.02% Glutamate.

Representation:

Preshift Parameters		*Postshift Parameters*	
ATP	1.0	ATP	1.0
Carbon	1.0	Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	0.9

Postshift Initial Conditions

Mutant definition

Parameters	
Rtg13_{T}	0.75
k_{accglu}	0.0369
k_{accpro}	0.0002
k_{accnh4}	0.0011

Initial conditions

Model agrees with experiment.

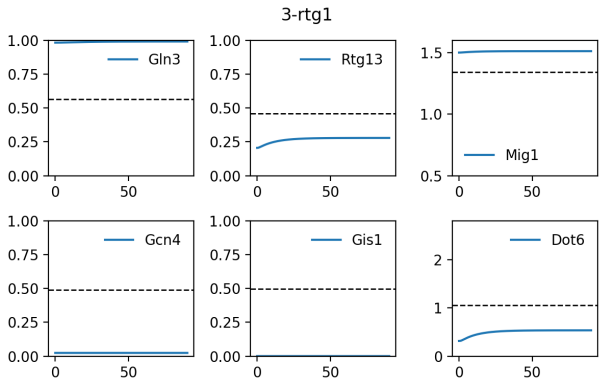


Figure D.3: Qualitative state comparison - *rtg1* in YNBD + 0.02% glutamate

D.4 4-rtg1

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.094
Gcn4	-	-	0.945
Rtg13	Off	On	0.613
Gln3	-	-	1.000

Description: Liu et al, 1999 studied a *rtg1* strain (PSY142 ρ^0) grown in YNBD.

Representation:

Preshift Parameters		*Postshift Parameters*	
ATP	1.0	ATP	1.0
Carbon	1.0	Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	0.0

Postshift Initial Conditions

Mutant definition

Parameters	
Rtg13_{T}	0.75
k_{accglu}	0.0369
k_{accpro}	0.0002
k_{accnh4}	0.0011

Initial conditions

Model does not agree with experiment.

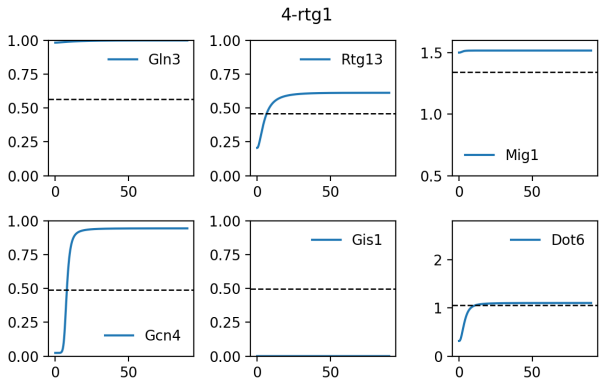


Figure D.4: Qualitative state comparison - *rtg1* in YNBD.

D.5 5-snf1

TF	Interpreted	Simulated	Simulation
Gis1	Off	On	0.884
Mig1	On	Off	1.325
Dot6	-	-	1.977
Gcn4	-	-	0.872
Rtg13	-	-	0.728
Gln3	-	-	0.999

Description: Gasmi et al, 2014 studied a *snf1* strain (BY4741) grown in Minimal + (0.2%casa) + 2% Ethanol.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	0.1
Carbon	0.1
$\text{Glutamine}_{\text{ext}}$	0.2

Postshift Initial Conditions

Mutant definition

Parameters

Snf1_{T}	0.0

Initial conditions

Snf1	0.0

Model does not agree with experiment.

D.6 6-gln3 gat1

Readout used is Gln3 Gcn4

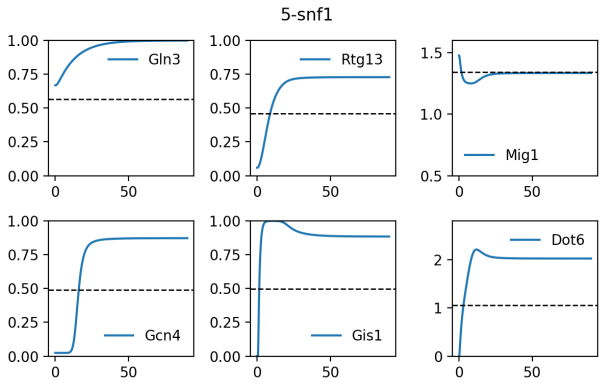


Figure D.5: Qualitative state comparison - *snf1* in Minimal + (0.2%casa) + 2% Ethanol.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	Off	0.000

Description: Beck et al, 1999 studied a *gln3 gat1* strain (wt) grown in YPD + rapamycin.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.0

Initial conditions

Gln3	0.0

Model does not agree with experiment.

D.7 7-2μ URE2

Readout used is Gln3 Gcn4

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	On	0.999

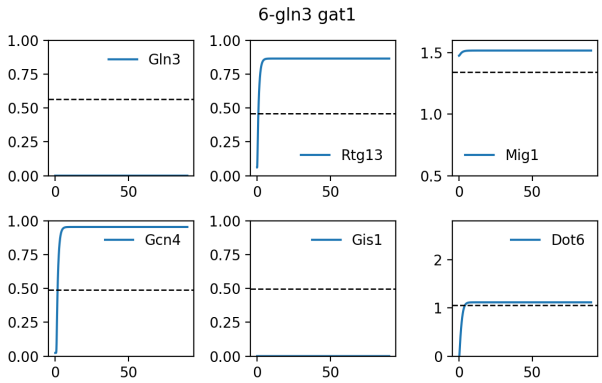


Figure D.6: Qualitative state comparison - *gln3 gat1* in YPD + rapamycin.

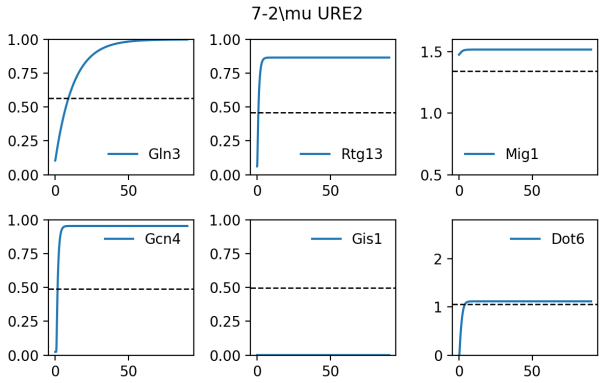


Figure D.7: Qualitative state comparison - *2μ URE2* in YPD + rapamycin.

Description: Beck et al, 1999 studied a *2μ URE2* strain (wt) grown in YPD + rapamycin.
Representation:

Preshift Parameters	
ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters	
ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions	
TORC1	0

Mutant definition

Parameters	
w_{gln3}	1.2784

Initial conditions

Model does not agree with experiment.

D.8 8-wt

Readout used is Gln3 Gcn4

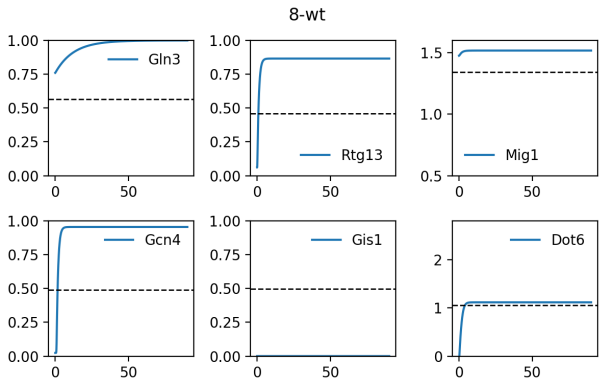


Figure D.8: Qualitative state comparison - *wt* in YPD + rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Beck et al, 1999 studied a *wt* strain (wt) grown in YPD + rapamycin.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters *Initial conditions*

Model agrees with experiment.

D.9 9-gln3

Readout used is Gln3 Gcn4

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	0.650

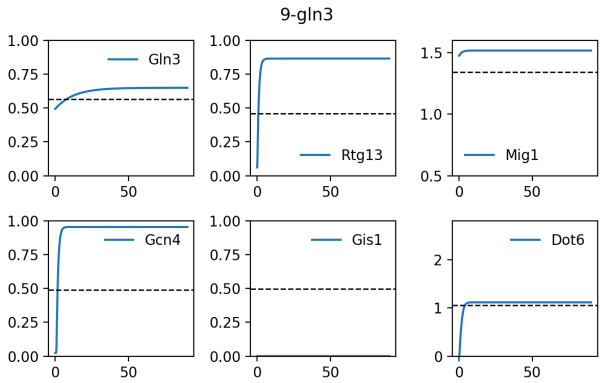


Figure D.9: Qualitative state comparison - *gln3* in YPD + rapamycin.

Description: Beck et al, 1999 studied a *gln3* strain (wt) grown in YPD + rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.65

Initial conditions

Gln3	0.0

Model agrees with experiment.

D.10 10-gat1

Readout used is Gln3 Gcn4

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	0.650

Description: Beck et al, 1999 studied a *gat1* strain (wt) grown in YPD + rapamycin.

Representation:

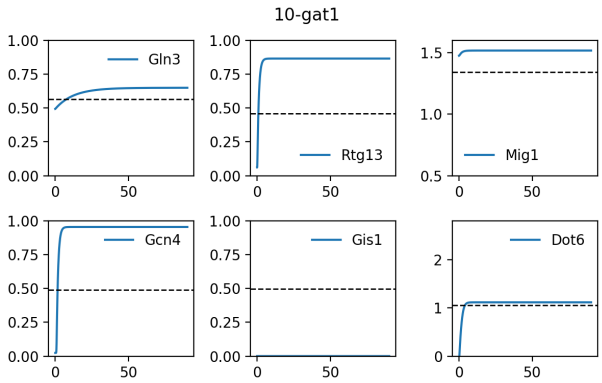


Figure D.10: Qualitative state comparison - *gat1* in YPD + rapamycin.

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.65

Initial conditions

Gln3	0.0

Model agrees with experiment.

D.11 11-gln3

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.037
Gcn4	-	-	0.806
Rtg13	-	-	0.702
Gln3	Off	On	0.649

Description: Crespo et al, 2002 studied a *gln3* strain (TB123) grown in SD + 1mM MSX.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
k_{accglu}	0.0

Postshift Initial Conditions

Mutant definition

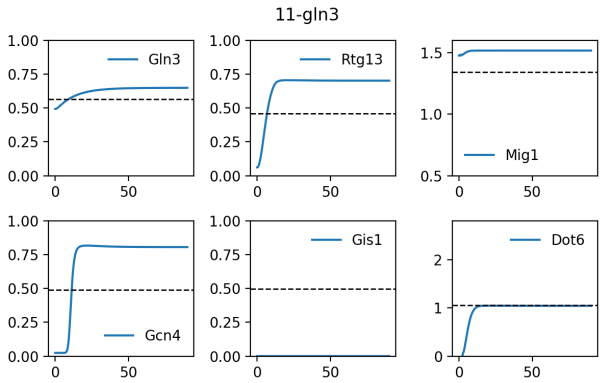


Figure D.11: Qualitative state comparison - *gln3* in SD + 1mM MSX.

Parameters	
$Gln3_T$	0.65

Initial conditions	
Gln3	0.0
Glutamine	0.01

Model does not agree with experiment.

D.12 12-gln3 gat1

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.037
Gcn4	-	-	0.806
Rtg13	-	-	0.702
Gln3	Off	Off	0.000

Description: Crespo et al, 2002 studied a *gln3 gat1* strain (TB123) grown in SD + 1mM MSX.

Representation:

Preshift Parameters	
ATP	1.0
Carbon	1.0
$Glutamine_{ext}$	1.0

Postshift Parameters	
ATP	1.0
Carbon	1.0
$Glutamine_{ext}$	1.0
k_{accglu}	0.0

Postshift Initial Conditions

Mutant definition

Parameters	
$Gln3_T$	0.0

Initial conditions	
Gln3	0.0
Glutamine	0.01

Model agrees with experiment.

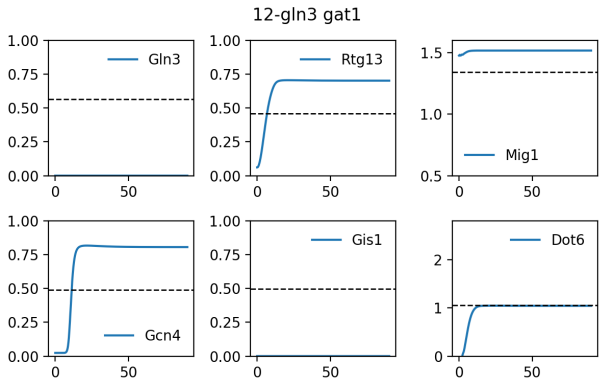


Figure D.12: Qualitative state comparison - *gln3 gat1* in SD + 1mM MSX.

D.13 13-wt

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.037
Gcn4	-	-	0.806
Rtg13	-	-	0.702
Gln3	On	On	0.999

Description: Crespo et al, 2002 studied a *wt* strain (TB123) grown in SD + 1mM MSX.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
k_{accglu}	0.0

Postshift Initial Conditions

Mutant definition

Parameters *Initial conditions*

Model agrees with experiment.

D.14 14-bcy1

Readout used is Gln3 Gcn4

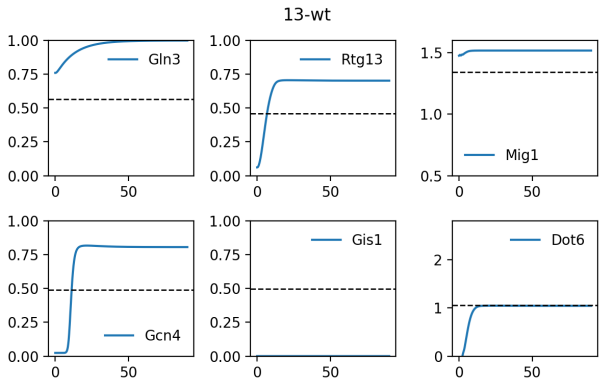


Figure D.13: Qualitative state comparison - *wt* in SD + 1mM MSX.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	On	1.000

Description: Zurita-Martinez et al, 2005 studied a *bcy1* strain (S1278b) grown in YP Glucose + 50nM rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{pkasch9}	0.0

Initial conditions

Model does not agree with experiment.

D.15 15-ira1

Readout used is Gln3 Gcn4

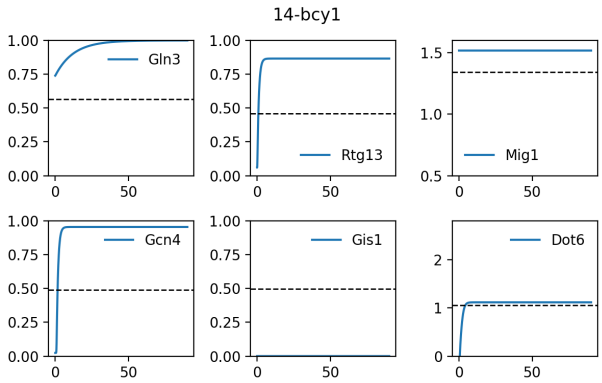


Figure D.14: Qualitative state comparison - *bcy1* in YP Glucose + 50nM rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	On	1.000

Description: Zurita-Martinez et al, 2005 studied a *ira1* strain (S1278b) grown in YP Glucose + 50nM rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{raspka}	0.9362

Initial conditions

Model does not agree with experiment.

D.16 16-ira1 ira2

Readout used is Gln3 Gcn4

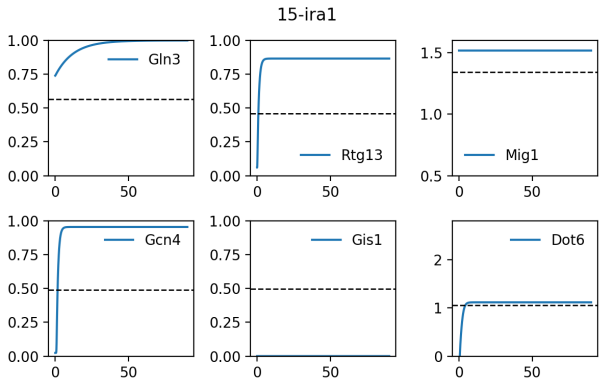


Figure D.15: Qualitative state comparison - *ira1* in YP Glucose + 50nM rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	On	1.000

Description: Zurita-Martinez et al, 2005 studied a *ira1 ira2* strain (S1278b) grown in YP Glucose + 50nM rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{raspka}	0.1872

Initial conditions

Model does not agree with experiment.

D.17 17-ras2

Readout used is Gln3 Gcn4

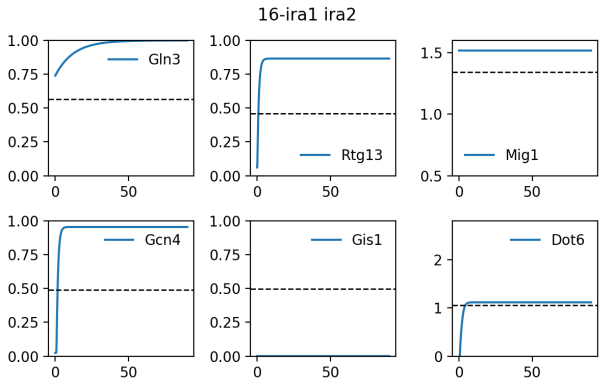


Figure D.16: Qualitative state comparison - *ira1 ira2* in YP Glucose + 50nM rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.001
Mig1	-	-	1.399
Dot6	-	-	1.637
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Zurita-Martinez et al, 2005 studied a *ras2* strain (MLY41a) grown in YP Glucose + 50nM rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Ras_{T}	0.0

Initial conditions

Model agrees with experiment.

D.18 18-tpk1

Readout used is Gln3 Gcn4

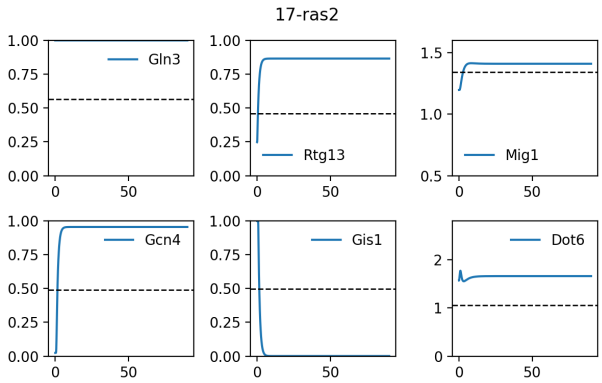


Figure D.17: Qualitative state comparison - *ras2* in YP Glucose + 50nM rapamycin.

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.411
Dot6	-	-	1.578
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Zurita-Martinez et al, 2005 studied a *tpk1* strain (MLY41a) grown in YP Glucose + 50nM rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

PKA_{T}	0.66

Initial conditions

Model agrees with experiment.

D.19 19-RAS2v19 gln3 gat1

Readout used is Gln3 Gcn4

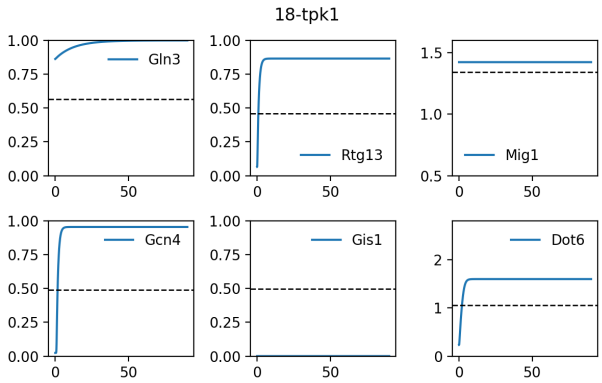


Figure D.18: Qualitative state comparison - *tpk1* in YP Glucose + 50nM rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	Off	0.000

Description: Schmelzle et al, 2003 studied a *RAS2v19 gln3 gat1* strain (TB50a) grown in YPD + 200ng/mL rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.0
w_{raspka}	0.0

Initial conditions

Model does not agree with experiment.

D.20 20-TPK1 gln3 gat1

Readout used is Gln3 Gcn4

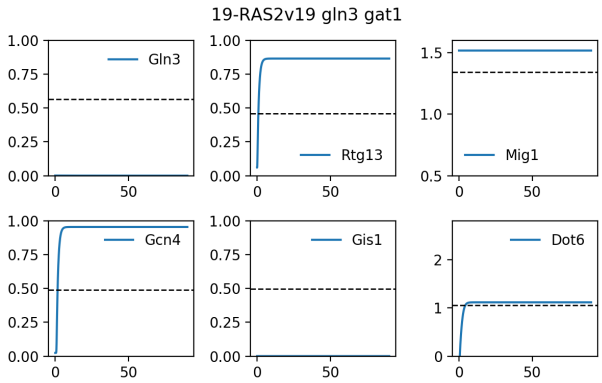


Figure D.19: Qualitative state comparison - *RAS2v19 gln3 gat1* in YPD + 200ng/mL rapamycin.

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.177
Mig1	-	-	1.338
Dot6	-	-	1.874
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	Off	0.000

Description: Schmelzle et al, 2003 studied a *TPK1 gln3 gat1* strain (TB50a) grown in YPD + 200ng/mL rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.0
w_{pkacamp}	0.0

Initial conditions

Model does not agree with experiment.

D.21 21-bcy1

Readout used is Gln3 Gcn4

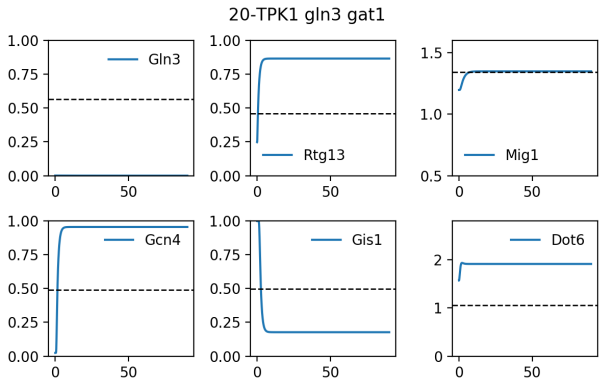


Figure D.20: Qualitative state comparison - *TPK1 gln3 gat1* in YPD + 200ng/mL rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.177
Mig1	-	-	1.338
Dot6	-	-	1.874
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Schmelzle et al, 2003 studied a *bcy1* strain (TB50a) grown in YPD + 200ng/mL rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{pkacamp}	0.0

Initial conditions

Model agrees with experiment.

D.22 22-bcy1 gln3 gat1

Readout used is Gln3 Gcn4

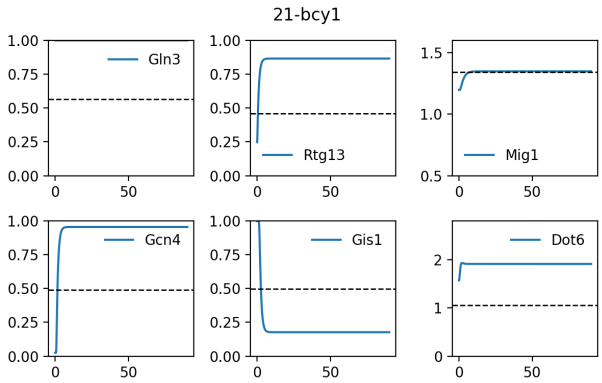


Figure D.21: Qualitative state comparison - *bcy1* in YPD + 200ng/mL rapamycin.

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	Off	0.000

Description: Schmelzle et al, 2003 studied a *bcy1 gln3 gat1* strain (TB50alpha) grown in YPD + 200ng/mL rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gln3_{T}	0.0
w_{pkacamp}	1021.0983

Initial conditions

Model does not agree with experiment.

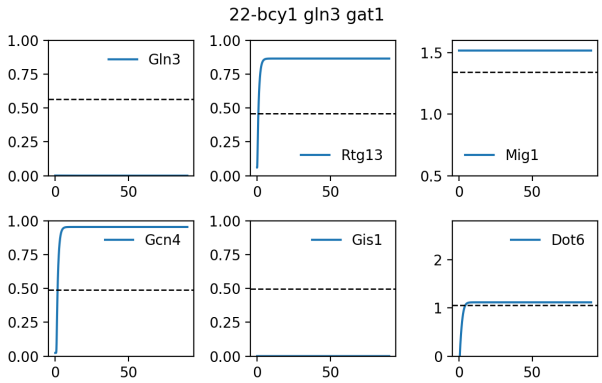


Figure D.22: Qualitative state comparison - *bcy1 gln3 gat1* in YPD + 200ng/mL rapamycin.

D.23 23-wt

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.037
Gcn4	On	On	0.806
Rtg13	-	-	0.702
Gln3	-	-	0.999

Description: Cherkasova et al, 2010 studied a *wt* strain (H1642) grown in SC + 10mM 3AT.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0
k_{accnh4}	0.0001

Postshift Initial Conditions

Mutant definition

Parameters *Initial conditions*

Model agrees with experiment.

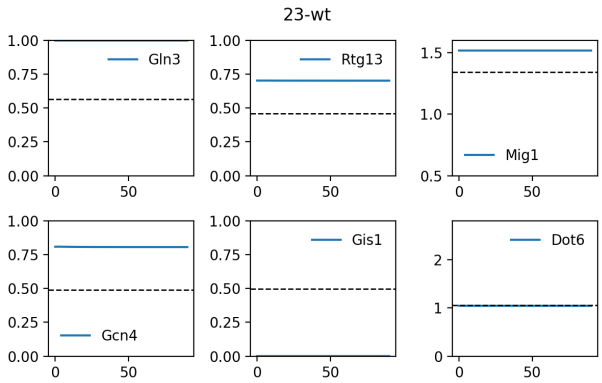


Figure D.23: Qualitative state comparison - *wt* in SC + 10mM 3AT.

D.24 24-gcn2

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.037
Gcn4	Off	Off	0.024
Rtg13	-	-	0.702
Gln3	-	-	0.999

Description: Cherkasova et al, 2010 studied a *gcn2* strain (H1895) grown in SC + 10mM 3AT.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0
k_{accnh4}	0.0001

Postshift Initial Conditions

Mutant definition

Parameters

Gcn2_{T}	0.0

Initial conditions

Model agrees with experiment.

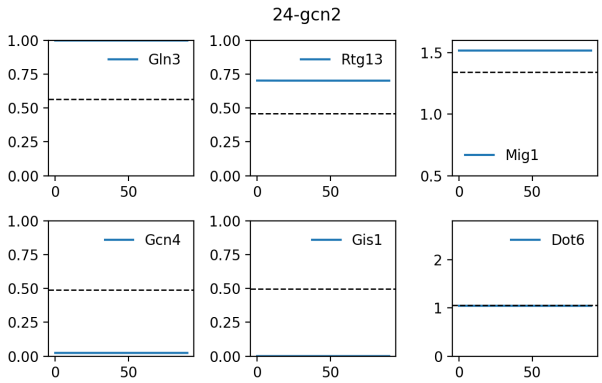


Figure D.24: Qualitative state comparison - *gcn2* in SC + 10mM 3AT.

D.25 25-snf1

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.504
Dot6	-	-	1.037
Gcn4	Off	On	0.806
Rtg13	-	-	0.702
Gln3	-	-	0.999

Description: Cherkasova et al, 2010 studied a *snf1* strain (HQY343) grown in SC + 10mM 3AT.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0
k_{accnh4}	0.0001

Postshift Initial Conditions

Mutant definition

Parameters

Snf1_{T}	0.0

Initial conditions

Model does not agree with experiment.

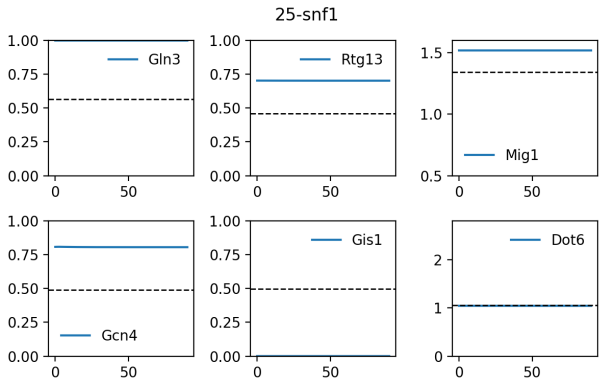


Figure D.25: Qualitative state comparison - *snf1* in SC + 10mM 3AT.

D.26 26-gcn2 snf1

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.504
Dot6	-	-	1.037
Gcn4	Off	Off	0.024
Rtg13	-	-	0.702
Gln3	-	-	0.999

Description: Cherkasova et al, 2010 studied a *gcn2 snf1* strain (HQY344) grown in SC + 10mM 3AT.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0
NH4	2.0
k_{accnh4}	0.0001

Postshift Initial Conditions

Mutant definition

Parameters

Gcn2_{T}	0.0
Snf1_{T}	0.0

Initial conditions

Model agrees with experiment.

D.27 27-gln3 gcn4

Readout used is Gln3 Gcn4

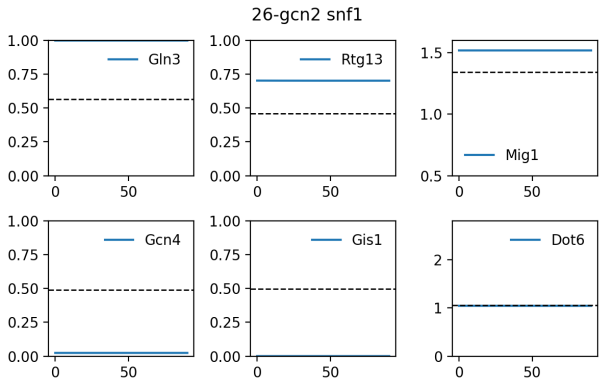


Figure D.26: Qualitative state comparison - *gcn2 snf1* in SC + 10mM 3AT.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	Off	0.000
Rtg13	-	-	0.866
Gln3	Off	Off	0.000

Description: Valenzuela et al, 2001 studied a *gln3 gcn4* strain (CLA-303) grown in YPD 200ng/mL rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gcn4_{T}	0.0
Gln3_{T}	0.0

Initial conditions

Model agrees with experiment.

D.28 28-gcn4

Readout used is Gln3 Gcn4

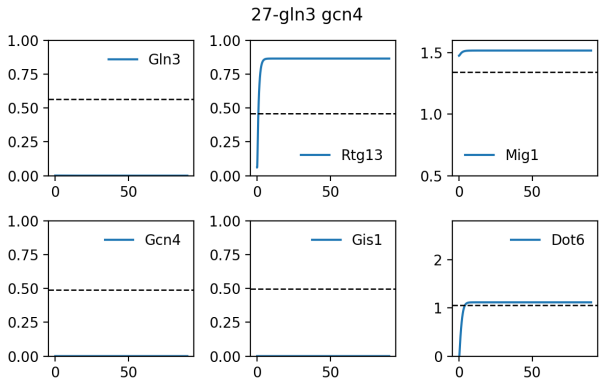


Figure D.27: Qualitative state comparison - *gln3 gcn4* in YPD 200ng/mL rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	On	Off	0.000
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Valenzuela et al, 2001 studied a *gcn4* in YPD 200ng/mL rapamycin.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	0.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Gcn4_{T}	0.0

Initial conditions

Model does not agree with experiment.

D.29 29-rph1 gis1

TF	Interpreted	Simulated	Simulation
Gis1	Off	Off	0.000
Mig1	On	On	1.465
Dot6	-	-	-0.048
Gcn4	-	-	0.024
Rtg13	-	-	0.061
Gln3	-	-	0.755

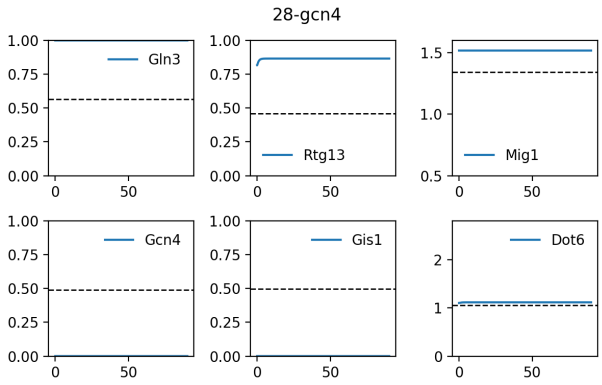


Figure D.28: Qualitative state comparison - *gcn4* in YPD 200ng/mL rapamycin.

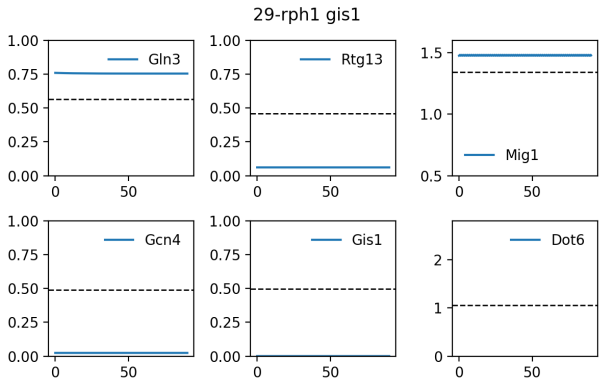


Figure D.29: Qualitative state comparison - *rph1 gis1* in SD glucose -ura.

Description: H. Ronne et al, 1999 studied a *rph1 gis1* in SD glucose -ura.

Representation:

Preshift Parameters		*Postshift Parameters*	
ATP	1.0	ATP	1.0
Carbon	1.0	Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Initial Conditions

Mutant definition

Parameters		*Initial conditions*
Gis1_{T}	0.0	

Model agrees with experiment.

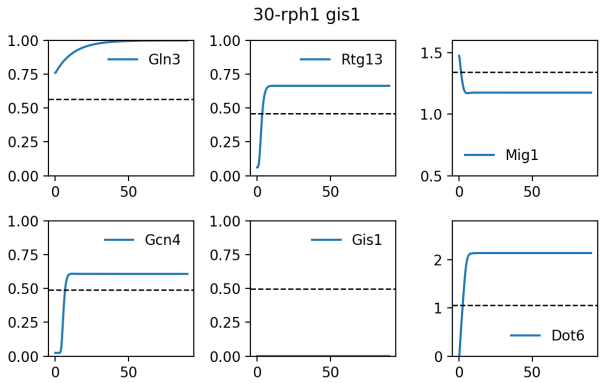


Figure D.30: Qualitative state comparison - *rph1 gis1* in YP ethanol.

D.30 30-rph1 gis1

TF	Interpreted	Simulated	Simulation
Gis1	Off	Off	0.000
Mig1	Off	Off	1.168
Dot6	-	-	2.078
Gcn4	-	-	0.608
Rtg13	-	-	0.664
Gln3	-	-	1.000

Description: H. Ronne et al, 1999 studied a *rph1 gis1* in YP ethanol.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	0.1
Carbon	0.1
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Initial Conditions

Mutant definition

Parameters

Gis1_{T}	0.0

Initial conditions

Model agrees with experiment.

D.31 31-mig1 snf1 pde2

TF	Interpreted	Simulated	Simulation
Gis1	Off	Off	0.000
Mig1	-	-	-45.173
Dot6	-	-	-0.256
Gcn4	-	-	0.024
Rtg13	-	-	0.058
Gln3	-	-	0.661

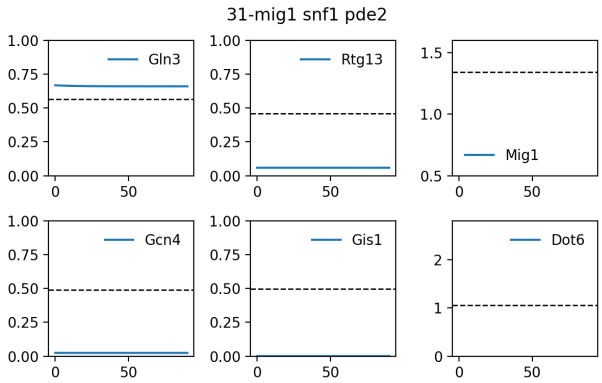


Figure D.31: Qualitative state comparison - *mig1 snf1 pde2* in SD glucose.

Description: H. Ronne et al, 1999 studied a *mig1 snf1 pde2* in SD glucose.
Representation:

Preshift Parameters		*Postshift Parameters*		*Postshift Initial Conditions*
ATP	1.0	ATP	1.0	
Carbon	1.0	Carbon	1.0	
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0	

Mutant definition

Parameters		*Initial conditions*
Mig1_{T}	0.0	
PDE_{T}	0.2	
Snf1_{T}	0.0	

Model agrees with experiment.

D.32 32-mig1 snf1 pde2

TF	Interpreted	Simulated	Simulation
Gis1	Off	Off	0.000
Mig1	-	-	-45.182
Dot6	-	-	-0.145
Gcn4	-	-	0.024
Rtg13	-	-	0.058
Gln3	-	-	0.661

Description: H. Ronne et al, 1999 studied a *mig1 snf1 pde2* in SD raffinose.
Representation:

Preshift Parameters		*Postshift Parameters*		*Postshift Initial Conditions*
ATP	1.0	ATP	0.7	
Carbon	1.0	Carbon	0.7	
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0	

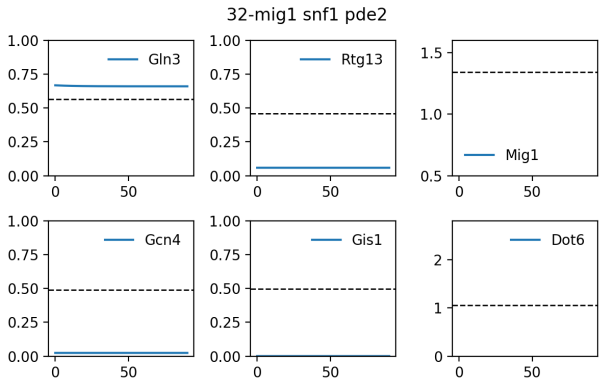


Figure D.32: Qualitative state comparison - *mig1 snf1 pde2* in SD raffinose.

Mutant definition

Parameters

$Mig1_T$	0.0
PDE_T	0.2
$Snf1_T$	0.0

Initial conditions

Model agrees with experiment.

D.33 33-mig1 snf1 pde2

TF	Interpreted	Simulated	Simulation
Gis1	Off	Off	0.000
Mig1	-	-	-45.159
Dot6	-	-	-0.254
Gcn4	-	-	0.024
Rtg13	-	-	0.058
Gln3	-	-	0.661

Description: H. Ronne et al, 1999 studied a *mig1 snf1 pde2* in SD galactose.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$Glutamine_{ext}$	1.0

Postshift Parameters

ATP	0.8
Carbon	0.8
$Glutamine_{ext}$	1.0

Postshift Initial Conditions

Mutant definition

Parameters

$Mig1_T$	0.0
PDE_T	0.2
$Snf1_T$	0.0

Initial conditions

Model agrees with experiment.

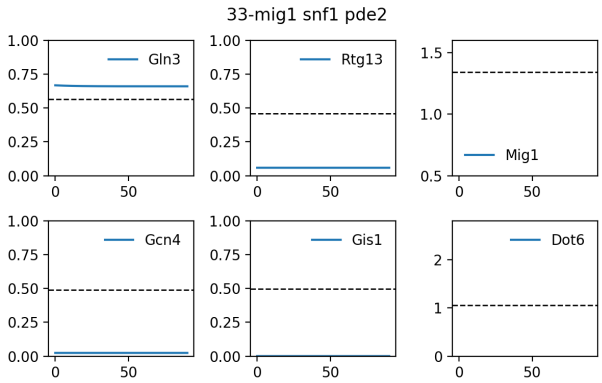


Figure D.33: Qualitative state comparison - *mig1 snf1 pde2* in SD galactose.

D.34 34-snf1

Readout used is Gln3 Gcn4

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.504
Dot6	-	-	1.108
Gcn4	Off	On	0.955
Rtg13	-	-	0.866
Gln3	Off	On	1.000

Description: Bertram et al, 2002 studied a *snf1* in YPD + rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Snf1_{T}	0.0

Initial conditions

Model does not agree with experiment.

D.35 35-reg1

Readout used is Gln3 Gcn4

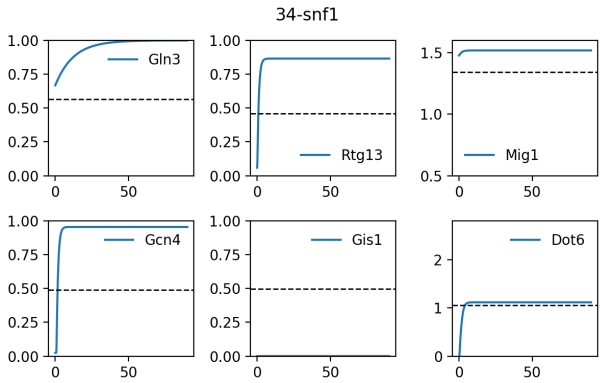


Figure D.34: Qualitative state comparison - *snf1* in YPD + rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.478
Dot6	-	-	1.108
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

Description: Bertram et al, 2002 studied a *reg1* in YPD + rapamycin.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{snfglc}	0.0

Initial conditions

Model agrees with experiment.

D.36 36-ure2

Readout used is Gln3 Gcn4

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	On	On	0.955
Rtg13	-	-	0.866
Gln3	On	On	1.000

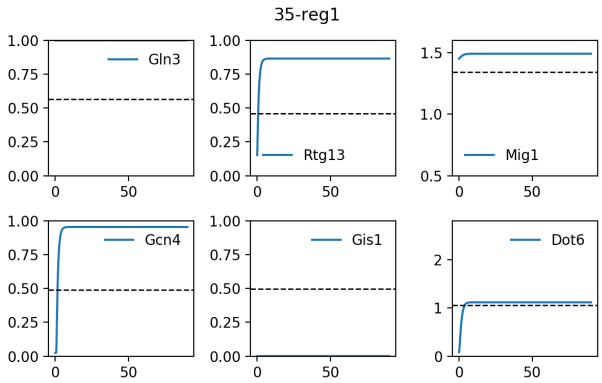


Figure D.35: Qualitative state comparison - *reg1* in YPD + rapamycin.

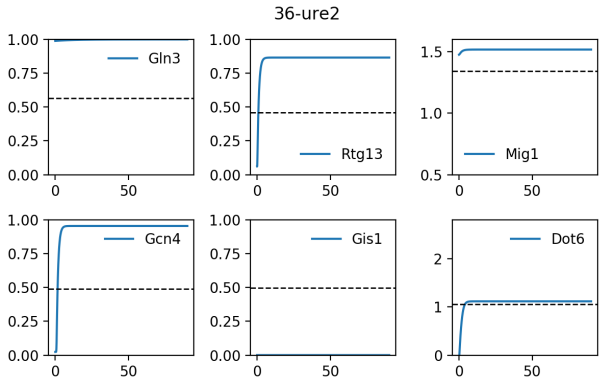


Figure D.36: Qualitative state comparison - *ure2* in YPD + rapamycin.

Description: Bertram et al, 2002 studied a *ure2* in YPD + rapamycin.
Representation:

Preshift Parameters	
ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters	
ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions	
TORC1	0

Mutant definition

Parameters	
w_{gln3}	0.0

Initial conditions

Model agrees with experiment.

D.37 37-tap42-11

Readout used is Gln3 Gcn4

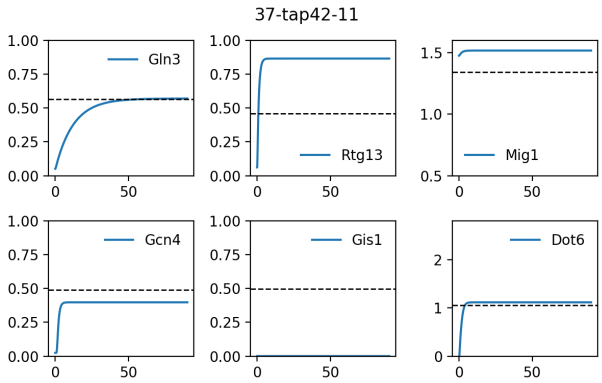


Figure D.37: Qualitative state comparison - *tap42-11* in YPD + rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.503
Dot6	-	-	1.108
Gcn4	Off	Off	0.398
Rtg13	-	-	0.866
Gln3	Off	On	0.570

Description: Huber et al, 2009 studied a *tap42-11* in YPD + rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

w_{gcn}	0.0
w_{glnsit}	0.0

Initial conditions

Model does not agree with experiment.

D.38 38-SCH9$^{\text{DE}}$

Readout used is Gln3 Gcn4

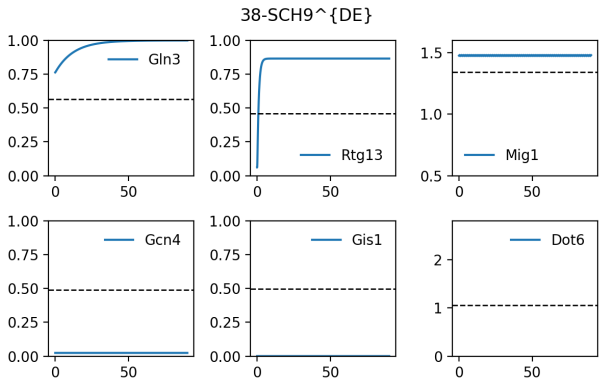


Figure D.38: Qualitative state comparison - *SCH9^{DE}* in YPD + rapamycin.

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.467
Dot6	-	-	-0.103
Gcn4	Off	Off	0.024
Rtg13	-	-	0.866
Gln3	Off	On	1.000

Description: Huber et al, 2009 studied a *SCH9^{DE}* in YPD + rapamycin.
Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Sch9_{T}	1.0
w_{sch9}	-10

Initial conditions

Sch9	1.0

Model does not agree with experiment.

D.39 39-SCH9^{DE} tap42

TF	Interpreted	Simulated	Simulation
Gis1	-	-	0.000
Mig1	-	-	1.467
Dot6	-	-	-0.103
Gcn4	Off	Off	0.024
Rtg13	-	-	0.866
Gln3	Off	On	0.596

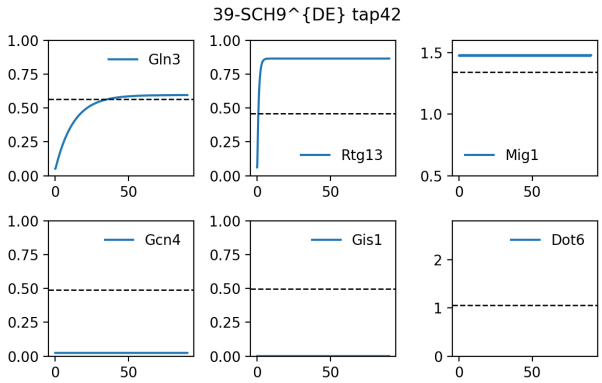


Figure D.39: Qualitative state comparison - *SCH9^{DE} tap42* in YPD + rapamycin.

Description: Huber et al, 2009 studied a *SCH9^{DE} tap42* in YPD + rapamycin.

Representation:

Preshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0

Postshift Parameters

ATP	1.0
Carbon	1.0
$\text{Glutamine}_{\text{ext}}$	1.0
TORC1_{T}	0.0

Postshift Initial Conditions

TORC1	0

Mutant definition

Parameters

Sch9_{T}	1.0
w_{sch9}	-10
w_{gcn}	0.0
w_{glnsit}	0.0

Initial conditions

Sch9	1.0

Model does not agree with experiment.

D.40 40-sch9

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	On	Off	0.046
Mig1	-	-	1.264
Dot6	-	-	1.825
Gcn4	-	-	0.963
Rtg13	-	-	0.577
Gln3	-	-	1.000

Description: Roosen et al, 2005 studied a *sch9* in YP + glycerol.

Representation:

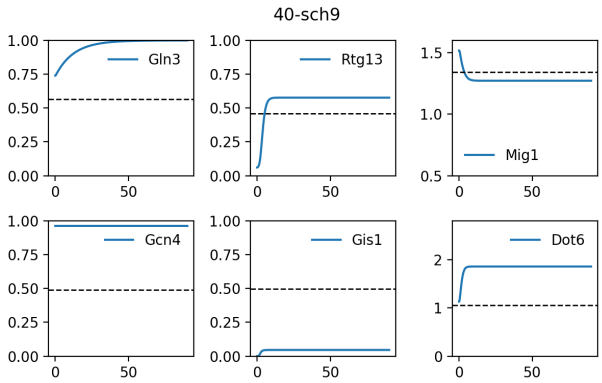


Figure D.40: Qualitative state comparison - *sch9* in YP + glycerol.

Preshift Parameters		*Postshift Parameters*		*Postshift Initial Conditions*
ATP	1.0	ATP	0.01	
Carbon	1.0	Carbon	0.01	
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0	

Mutant definition

Parameters		*Initial conditions*	
Sch9_{T}	0.0	Sch9	0.0

Model does not agree with experiment.

D.41 41-sch9 gis1

TF	**Interpreted**	**Simulated**	**Simulation**
Gis1	Off	Off	0.000
Mig1	-	-	1.264
Dot6	-	-	1.825
Gcn4	-	-	0.963
Rtg13	-	-	0.577
Gln3	-	-	1.000

Description: Roosen et al, 2005 studied a *sch9 gis1* in YP + glycerol.

Representation:

Preshift Parameters		*Postshift Parameters*		*Postshift Initial Conditions*
ATP	1.0	ATP	0.01	
Carbon	1.0	Carbon	0.01	
$\text{Glutamine}_{\text{ext}}$	1.0	$\text{Glutamine}_{\text{ext}}$	1.0	

Mutant definition

Parameters		*Initial conditions*	
Gis1_{T}	0.0	Gis1	0.0
Sch9_{T}	0.0	Sch9	0.0

Model agrees with experiment.

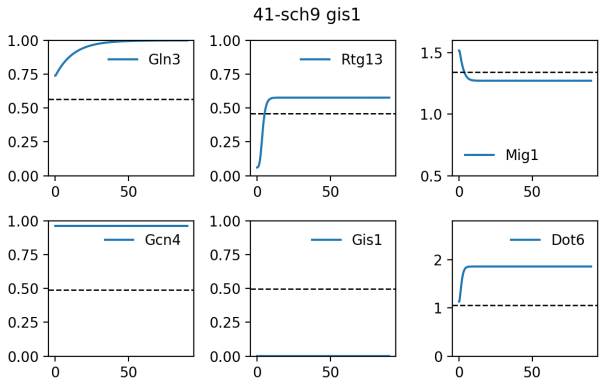


Figure D.41: Qualitative state comparison - *sch9 gis1* in YP + glycerol.

Bibliography

[1] Suckjoon Jun and Sattar Taheri-Araghi. Cell-size maintenance: Universal strategy revealed. *Trends in Microbiology*, 23(1):4–6, 2015.

[2] Samuel Marguerat and Jürg Bähler. Coordinating genome expression with cell size. *Trends in Genetics*, 28(11):560–565, 2012.

[3] Moshe Kafri, Eyal Metzl-Raz, Felix Jonas, and Naama Barkai. Rethinking cell growth models. *FEMS Yeast Research*, 16(7):fow081, 2016.

[4] John J. Tyson and Kenneth B. Hannsgen. Cell growth and division: a deterministic/probabilistic model of the cell cycle. *Journal of Mathematical Biology*, 23(2):231–246, 1986.

[5] FM Williams. A model of cell growth dynamics. *Journal of Theoretical Biology*, 15(2):190–207, 1967.

[6] M. Schaechter, O. MaalOe, and N. O. Kjeldgaard. Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium. *Journal of General Microbiology*, 19(3):592–606, 1958.

[7] M. Schaechter, J. P. Williamson, J. R. Hood, and A. L. Koch. Growth, cell and nuclear divisions in some bacteria. *Journal of General Microbiology*, 29(3):421–434, 1962.

[8] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, 2010.

[9] MICHAEL CASHEL and JONATHAN GALLANT. Two compounds implicated in the function of the rc gene of escherichia coli. *Nature*, 221(5183):838–841, 1969.

[10] Anjana Srivatsan and Jue D Wang. Control of bacterial transcription, translation and replication by (p)ppgpp. *Current Opinion in Microbiology*, 11(2):100–105, 2008.

[11] J. J. Lemke, P. Sanchez-Vazquez, H. L. Burgos, G. Hedberg, W. Ross, and R. L. Gourse. Direct regulation of escherichia coli ribosomal protein promoters by the transcription factors ppgpp and dksa. *Proceedings of the National Academy of Sciences*, 108(14):5712–5717, 2011.

[12] Masayasu Nomura. Regulation of ribosome biosynthesis in escherichia coli and saccharomyces cerevisiae: diversity and common principles. *Journal of bacteriology*, 181(22):6857–6864, 1999.

[13] Arthur Harden. *Alcoholic fermentation*. Feb 2020. [Online; accessed 11. Feb. 2020].

[14] P Nurse and A Wiemken. Amino acid pools and metabolism during the cell division cycle of arginine-grown candida utilis. *Journal of bacteriology*, 117(3):1108–1116, 1974.

[15] H. Halvorson. A study of the properties of the free amino acid pool and enzyme synthesis in yeast. *The Journal of General Physiology*, 38(4):549–573, 1955.

[16] W E Courchesne and B Magasanik. Regulation of nitrogen assimilation in saccharomyces cerevisiae: Roles of the ure2 and gln3 genes. *Journal of Bacteriology*, 170(2):708–713, 1988.

[17] Boris Magasanik and Chris A Kaiser. Nitrogen regulation in saccharomycescerevisiae. *Gene*, 290(1):1–18, 2002.

[18] Marian Carlson, Barbara C Osmond, and David Botstein. Mutants of yeast defective in sucrose utilization. *Genetics*, 98(1):25–40, 1981.

[19] A. Woods, M. R. Munday, J. Scott, X. Yang, M. Carlson, and D. Carling. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo. *J. Biol. Chem.*, 269(30):19509–19515, Jul 1994.

[20] K Mbonyi, L van Aelst, J C Argüelles, A W Jans, and J M Thevelein. Glucose-induced hyperaccumulation of cyclic amp and defective glucose repression in yeast strains with reduced activity of cyclic amp-dependent protein kinase. *Molecular and Cellular Biology*, 10(9):4518–4523, 1990.

[21] J. M. Thevelein. Fermentable sugars and intracellular acidification as specific activators of the ras-adenylate cyclase signalling pathway in yeast: the relationship to nutrient-induced cell cycle control. *Molecular Microbiology*, 5(6):1301–1307, 1991.

[22] J. Heitman, N. Movva, and M. Hall. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science*, 253(5022):905–909, 1991.

[23] Michaela Conrad, Joep Schothorst, Harish Nag Kankipati, Griet Van Zeebroeck, Marta Rubio-Texeira, and Johan M Thevelein. Nutrient sensing and signaling in the yeast saccharomyces cerevisiae. *FEMS microbiology reviews*, 38(2):254–299, 2014.

[24] W MAGER and M SIDERIUS. Novel insights into the osmotic stress response of yeast. *FEMS Yeast Research*, 2(3):251–257, 2002.

[25] Ariel Stanhill, Naomi Schick, and David Engelberg. The yeast ras/cyclic amp pathway induces invasive growth by suppressing the cellular stress response. *Molecular and Cellular Biology*, 19(11):7529–7538, 1999.

[26] Isabel Mayordomo, Francisco Estruch, and Pascual Sanz. Convergence of the target of rapamycin and the snf1 protein kinase pathways in the regulation of the subcellular localization of msn2, a transcriptional activator of stre (stress response element)-regulated genes. *Journal of Biological Chemistry*, 277(38):35650–35656, 2002.

[27] Marta Moreno-Torres, Malika Jaquenoud, and Claudio De Virgilio. TORC1 controls G1–S cell cycle transition in yeast via Mpk1 and the greatwall kinase pathway. *Nat. Commun.*, 6(8256):1–10, Sep 2015.

[28] Kazuki Suda, Atsuki Kaneko, Mitsugu Shimobayashi, Akio Nakashima, Maeda Tatsuya, Michael N. Hall, and Takashi Ushimaru. Torc1 regulates autophagy induction in response to proteotoxic stress in yeast and human cells. *Biochemical and Biophysical Research Communications*, nil(nil):nil, 2019.

[29] Shintaro Kira, Keisuke Tabata, Kanae Shirahama-Noda, Akiko Nozoe, Tamotsu Yoshimori, and Takeshi Noda. Reciprocal conversion of gtr1 and gtr2 nucleotide-binding states by npr2-npr3 inactivates torc1 and induces autophagy. *Autophagy*, 10(9):1565–1578, 2014.

[30] Marie-Anne Deprez, Elja Eskes, Tobias Wilms, Paula Ludovico, and Joris Winderickx. Ph homeostasis links the nutrient sensing pka/torc1/sch9 ménage-à-trois to stress tolerance and longevity. *Microbial Cell*, 5(3):119–136, 2018.

[31] T. Yorimitsu, S. Zaman, J. R. Broach, and D. J. Klionsky. Protein kinase a and sch9 cooperatively regulate induction of autophagy in saccharomyces cerevisiae. *Molecular Biology of the Cell*, 18(10):4180–4189, 2007.

[32] Johnny Roosen, Kristof Engelen, Kathleen Marchal, Janick Mathys, Gerard Griffioen, Elisabetta Cameroni, Johan M Thevelein, Claudio De Virgilio, Bart De Moor, and Joris Winderickx. Pka and sch9 control a molecular switch important for the proper adaptation to nutrient availability. *Molecular microbiology*, 55(3):862–880, 2005.

[33] A. Trott, L. Shaner, and K. A. Morano. The molecular chaperone sse1 and the growth control protein kinase sch9 collaborate to regulate protein kinase a activity in saccharomyces cerevisiae. *Genetics*, 170(3):1009–1021, 2005.

[34] Asier González and Michael N Hall. Nutrient sensing and tor signaling in yeast and mammals. *The EMBO Journal*, 36(4):397–408, 2017.

[35] Lynne Chantranupong, Rachel L. Wolfson, and David M. Sabatini. Nutrient-sensing mechanisms across evolution. *Cell*, 161(1):67–83, 2015.

[36] Robert A. Saxton and David M. Sabatini. Mtor signaling in growth, metabolism, and disease. *Cell*, 169(2):361–371, 2017.

[37] Rachel L. Wolfson and David M. Sabatini. The dawn of the age of amino acid sensors for the mtorc1 pathway. *Cell Metabolism*, 26(2):301–309, 2017.

[38] J. E. Hughes Hallett, X. Luo, and A. P. Capaldi. State transitions in the torc1 signaling pathway and information processing in saccharomyces cerevisiae. *Genetics*, 198(2):773–786, 2014.

[39] Kevin Gonzales, Ömür Kayikçi, David G Schaeffer, and Paul M Magwene. Modeling mutant phenotypes and oscillatory dynamics in the saccharomyces cerevisiae camp-pka pathway. *BMC Systems Biology*, 7(1):40, 2013.

[40] J. Ross and A. P. Arkin. Complex systems: From chemistry to systems biology. *Proceedings of the National Academy of Sciences*, 106(16):6433–6434, 2009.

[41] Joel Keizer. Nonequilibrium thermodynamics and the stability of states far from equilibrium. *Accounts of Chemical Research*, 12(7):243–249, 1979.

[42] Joel Keizer. Dissipation and fluctuations in nonequilibrium thermodynamics. *The Journal of Chemical Physics*, 64(4):1679–1687, 1976.

[43] Joel Keizer. Fluctuations, stability, and generalized state functions at nonequilibrium steady states. *The Journal of Chemical Physics*, 65(11):4431–4444, 1976.

[44] H Strogatz Steven. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 1994.

[45] John J. Tyson. Bringing cartoons to life. *Nature*, 445(7130):823–823, 2007.

[46] Pavel Kraikivski, Katherine C Chen, Teeraphan Laomettachit, TM Murali, and John J Tyson. From start to finish: computational analysis of cell cycle control in budding yeast. *npj Systems Biology and Applications*, 1:15016, 2015.

[47] John J. Tyson, Christian I. Hong, C. Dennis Thron, and Bela Novak. A simple model of circadian rhythms based on dimerization and proteolysis of per and tim. *Biophysical Journal*, 77(5):2411–2417, 1999.

[48] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, Jan 1999.

[49] Raymond Cheong, Alexander Hoffmann, and Andre Levchenko. Understanding nf-κb signaling via mathematical modeling. *Molecular Systems Biology*, 4(1):192, 2008.

[50] Palakkad Krishnan Unni Vinod and Kareenhalli Viswanath Venkatesh. Quantification of the effect of amino acids on an integrated mtor and insulin signaling pathway. *Molecular BioSystems*, 5(10):1163, 2009.

[51] Kevin J. Verstrepen, Dirk Iserentant, Philippe Malcorps, Guy Derdelinckx, Patrick Van Dijck, Joris Winderickx, Isak S. Pretorius, Johan M. Thevelein, and Freddy R. Delvaux. Glucose and sucrose: Hazardous fast-food for industrial yeast? *Trends in Biotechnology*, 22(10):531–537, 2004.

[52] S. Colombo. Involvement of distinct g-proteins, gpa2 and ras, in glucose-and intracellular acidification-induced camp signalling in the yeast saccharomyces cerevisiae. *The EMBO Journal*, 17(12):3326–3341, 1998.

[53] K Mbonyi, M Beullens, K Detremerie, L Geerts, and J M Thevelein. Requirement of one functional ras gene and inability of an oncogenic ras variant to mediate the glucose-induced cyclic amp signal in the yeast saccharomyces cerevisiae. *Molecular and Cellular Biology*, 8(8):3051–3057, 1988.

[54] P. Ma, S. Wera, P. Van Dijck, and J. M. Thevelein. The pde1-encoded low-affinity phosphodiesterase in the yeast saccharomyces cerevisiae has a specific function in controlling agonist-induced camp signaling. *Molecular Biology of the Cell*, 10(1):91–104, 1999.

[55] Eitan Gross, Doron Goldberg, and Alexander Levitzki. Phosphorylation of the s. cerevisiae cdc25 in response to glucose results in its dissociation from ras. *Nature*, 360(6406):762–765, 1992.

[56] I Pedruzzi, N Burckert, P Egger, and C. De Virgilio. Saccharomyces cerevisiae ras/-camp pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein gis1. *The EMBO Journal*, 19(11):2569–2579, 2000.

[57] A. Smith. Yeast pka represses msn2p/msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. *The EMBO Journal*, 17(13):3556–3564, 1998.

[58] C Klein and K Struhl. Protein kinase a mediates growth-regulated expression of yeast ribosomal protein genes by modulating rap1 transcriptional activity. *Molecular and Cellular Biology*, 14(3):1920–1928, 1994.

[59] Shadia Zaman, Soyeon I Lippman, Lisa Schneper, Noam Slonim, and James R Broach. Glucose regulates transcription in yeast through a network of signaling pathways. *Molecular Systems Biology*, 5(nil):nil, 2009.

[60] Dakshayini G Chandrashekarappa, Rhonda R McCartney, and Martin C Schmidt. Ligand binding to the amp-activated protein kinase active site mediates protection of the activation loop from dephosphorylation. *Journal of Biological Chemistry*, 288(1):89–98, 2013.

[61] Marcin Maziarz, Aishwarya Shevade, LaKisha Barrett, and Sergei Kuchin. Springing into action: Reg2 negatively regulates snf1 protein kinase and facilitates recovery from prolonged glucose starvation in saccharomyces cerevisiae. *Applied and Environmental Microbiology*, 82(13):3875–3885, 2016.

[62] Karin Elbing, Rhonda R. McCartney, and Martin C. Schmidt. Purification and characterization of the three snf1-activating kinases ofsaccharomyces cerevisiae. *Biochemical Journal*, 393(3):797–805, 2006.

[63] S.-P. Hong, F. C. Leiper, A. Woods, D. Carling, and M. Carlson. Activation of yeast snf1 and mammalian amp-activated protein kinase by upstream kinases. *Proceedings of the National Academy of Sciences*, 100(15):8839–8843, 2003.

[64] A Schöler and H J Schüller. A carbon source-responsive promoter element necessary for activation of the isocitrate lyase gene icl1 is common to genes of the gluconeogenic pathway in the yeast saccharomyces cerevisiae. *Molecular and Cellular Biology*, 14(6):3613–3622, 1994.

[65] V. A. Cherkasova, H. Qiu, and A. G. Hinnebusch. Snf1 promotes phosphorylation of the subunit of eukaryotic translation initiation factor 2 by activating gcn2 and inhibiting phosphatases glc7 and sit4. *Molecular and Cellular Biology*, 30(12):2862–2873, 2010.

[66] Cecilia Garmendia-Torres, Albert Goldbeter, and Michel Jacquet. Nucleocytoplasmic oscillations of the yeast transcription factor msn2: Evidence for periodic pka activation. *Current Biology*, 17(12):1044–1049, 2007.

[67] Riko Hatakeyama and Claudio De Virgilio. Unsolved mysteries of rag gtpase signaling in yeast. *Small GTPases*, 7(4):239–246, 2016.

[68] Jörg Urban, Alexandre Soulard, Alexandre Huber, Soyeon Lippman, Debdyuti Mukhopadhyay, Olivier Deloche, Valeria Wanke, Dorothea Anrather, Gustav Ammerer, Howard Riezman, et al. Sch9 is a major target of torc1 in saccharomyces cerevisiae. *Molecular cell*, 26(5):663–674, 2007.

[69] C J Di Como and K T Arndt. Nutrients, via the tor proteins, stimulate the association of tap42 with type 2a phosphatases. *Genes & Development*, 10(15):1904–1916, 1996.

[70] J. L. Crespo, T. Powers, B. Fowler, and M. N. Hall. The tor-controlled transcription activators gln3, rtg1, and rtg3 are regulated in response to intracellular levels of glutamine. *Proceedings of the National Academy of Sciences*, 99(10):6784–6789, 2002.

[71] J. J. Tate, I. Georis, A. Feller, E. Dubois, and T. G. Cooper. Rapamycin-induced gln3 dephosphorylation is insufficient for nuclear localization: Sit4 and pp2a phosphatases are regulated and function differently. *Journal of Biological Chemistry*, 284(4):2522–2534, 2008.

[72] R. Rai, J. J. Tate, K. Shanmuganatham, M. M. Howe, D. Nelson, and T. G. Cooper. Nuclear gln3 import is regulated by nitrogen catabolite repression whereas export is specifically regulated by glutamine. *Genetics*, 201(3):989–1016, 2015.

[73] M. Crauwels, M. C. V. Donaton, M. B. Pernambuco, J. Winderickx, J. H. de Winde, and J. M. Thevelein. The sch9 protein kinase in the yeast saccharomyces cerevisiae controls capk activity and is required for nitrogen activation of the fermentable-growth-medium-induced (fgm) pathway. *Microbiology*, 143(8):2627–2637, 1997.

[74] Aili Zhang and Wenxuan Gao. Mechanisms of protein kinase sch9 regulating bcy1 in saccharomyces cerevisiae. *FEMS microbiology letters*, 331(1):10–16, 2012.

[75] Aili Zhang, Yubao Shen, Wenxuan Gao, and Jian Dong. Role of sch9 in regulating ras-camp signal pathway in saccharomyces cerevisiae. *FEBS letters*, 585(19):3026–3032, 2011.

[76] T Toda, S Cameron, P Sass, and M Wigler. Sch9, a gene of saccharomyces cerevisiae that encodes a protein distinct from, but functionally and structurally related to, camp-dependent protein kinase catalytic subunits. *Genes & Development*, 2(5):517–527, 1988.

[77] A. Reinders, N. Burckert, T. Boller, A. Wiemken, and C. De Virgilio. Saccharomyces cerevisiae camp-dependent protein kinase controls entry into stationary phase through the rim15p protein kinase. *Genes & Development*, 12(18):2943–2955, 1998.

[78] Ivo Pedruzzi, Frédérique Dubouloz, Elisabetta Cameroni, Valeria Wanke, Johnny Roosen, Joris Winderickx, and Claudio De Virgilio. Tor and pka signaling pathways converge on the protein kinase rim15 to control entry into g 0. *Molecular cell*, 12(6):1607–1613, 2003.

[79] Erwin Swinnen, Valeria Wanke, Johnny Roosen, Bart Smets, Frédérique Dubouloz, Ivo Pedruzzi, Elisabetta Cameroni, Claudio De Virgilio, and Joris Winderickx. Rim15 and the crossroads of nutrient signalling pathways in Saccharomyces cerevisiae. *Cell Division*, 1(1):3, Dec 2006.

[80] James R Broach. Nutritional control of growth and development in yeast. *Genetics*, 192(1):73–105, 2012.

[81] P. Jorgensen. A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes & Development*, 18(20):2491–2505, 2004.

[82] Harri Lempiäinen, Aino Uotila, Jörg Urban, Ilse Dohnal, Gustav Ammerer, Robbie Loewith, and David Shore. Sfp1 interaction with torc1 and mrs6 reveals feedback regulation on tor signaling. *Molecular Cell*, 33(6):704–716, 2009.

[83] S. A. Zurita-Martinez and M. E. Cardenas. Tor and cyclic amp-protein kinase a: Two parallel pathways regulating expression of genes required for cell growth. *Eukaryotic Cell*, 4(1):63–71, 2005.

[84] A. Huber, B. Bodenmiller, A. Uotila, M. Stahl, S. Wanka, B. Gerrits, R. Aebersold, and R. Loewith. Characterization of the rapamycin-sensitive phosphoproteome reveals that sch9 is a central coordinator of protein synthesis. *Genes & Development*, 23(16):1929–1943, 2009.

[85] D. DeMille, B. D. Badal, J. B. Evans, A. D. Mathis, J. F. Anderson, and J. H. Grose. Pas kinase is activated by direct snf1-dependent phosphorylation and mediates inhibition of torc1 through the phosphorylation and activation of pbp1. *Molecular Biology of the Cell*, 26(3):569–582, 2014.

[86] J. Zhang, S. Vaga, P. Chumnanpuen, R. Kumar, G. N. Vemuri, R. Aebersold, and J. Nielsen. Mapping the interaction of snf1 with torc1 in saccharomyces cerevisiae. *Molecular Systems Biology*, 7(1):545–545, 2011.

[87] James E Hughes Hallett, Xiangxia Luo, and Andrew P Capaldi. Snf1/ampk promotes the formation of kog1/raptor-bodies to increase the activation threshold of torc1 in budding yeast. *eLife*, 4(nil):nil, 2015.

[88] Manoël Prouteau, Ambroise Desfosses, Christian Sieben, Clélia Bourgoint, Nour Lydia Mozaffari, Davide Demurtas, Alok K. Mitra, Paul Guichard, Suliana Manley, and Robbie Loewith. Torc1 organized in inhibited domains (toroids) regulate torc1 activity. *Nature*, 550(7675):265–269, 2017.

[89] P. G. Bertram, J. H. Choi, J. Carvalho, T.-F. Chan, W. Ai, and X. F. S. Zheng. Convergence of tor-nitrogen and snf1-glucose signaling pathways onto gln3. *Molecular and Cellular Biology*, 22(4):1246–1252, 2002.

[90] Jennifer C Ewald. How yeast coordinates metabolism, growth and division. *Current Opinion in Microbiology*, 45(nil):1–7, 2018.

[91] Raffaele Nicastro, Farida Tripodi, Marco Gaggini, Andrea Castoldi, Veronica Reghellin, Simona Nonnis, Gabriella Tedeschi, and Paola Coccetti. Snf1 phosphorylates adenylate cyclase and negatively regulates protein kinase a-dependent transcription in-saccharomyces cerevisiae. *Journal of Biological Chemistry*, 290(41):24715–24726, 2015.

[92] L. Barrett, M. Orlova, M. Maziarz, and S. Kuchin. Protein kinase a contributes to the negative control of snf1 protein kinase in saccharomyces cerevisiae. *Eukaryotic Cell*, 11(2):119–128, 2011.

[93] D. Gonze, M. Jacquet, and A. Goldbeter. Stochastic modelling of nucleocytoplasmic oscillations of the transcription factor msn2 in yeast. *Journal of The Royal Society Interface*, 5(Suppl 1):S95–S109, 2008.

[94] Paolo Cazzaniga, Dario Pescini, Daniela Besozzi, Giancarlo Mauri, Sonia Colombo, and Enzo Martegani. Modeling and stochastic simulation of the ras/camp/pka pathway in the yeast saccharomyces cerevisiae evidences a key regulatory function for intracellular guanine nucleotides pools. *Journal of Biotechnology*, 133(3):377–385, 2008.

[95] Dario Pescini, Paolo Cazzaniga, Daniela Besozzi, Giancarlo Mauri, Loredana Amigoni, Sonia Colombo, and Enzo Martegani. Simulation of the ras/camp/pka pathway in budding yeast highlights the establishment of stable oscillatory states. *Biotechnology Advances*, 30(1):99–107, 2012.

[96] Jacob Stewart-Ornstein, Susan Chen, Rajat Bhatnagar, Jonathan S. Weissman, and Hana El-Samad. Model-guided optogenetic study of pka signaling in budding yeast. *Molecular Biology of the Cell*, 28(1):221–227, 2017.

[97] Thomas Williamson, Jean-Marc Schwartz, Douglas B Kell, and Lubomira Stateva. Deterministic mathematical models of the camp pathway in saccharomyces cerevisiae. *BMC Systems Biology*, 3(1):70, 2009.

[98] Raúl García-Salcedo, Timo Lubitz, Gemma Beltran, Karin Elbing, Ye Tian, Simone Frey, Olaf Wolkenhauer, Marcus Krantz, Edda Klipp, and Stefan Hohmann. Glucose de-repression by yeast amp-activated protein kinase snf1 is controlled via at least two independent steps. *FEBS Journal*, 281(7):1901–1917, 2014.

[99] Niek Welkenhuysen, Johannes Borgqvist, Mattias Backman, Loubna Bendrioua, Mattias Goksör, Caroline B Adiels, Marija Cvijovic, and Stefan Hohmann. Single-cell study links metabolism with nutrient signaling and reveals sources of variability. *BMC Systems Biology*, 11(1):59, 2017.

[100] Annika G. Sonntag, Piero Dalle Pezze, Daryl P. Shanley, and Kathrin Thedieck. A modelling-experimental approach reveals insulin receptor substrate (irs)-dependent regulation of adenosine monosphosphate-dependent kinase (ampk) by insulin. *FEBS Journal*, 279(18):3314–3328, 2012.

[101] Piero Dalle Pezze, Stefanie Ruf, Annika G. Sonntag, Miriam Langelaar-Makkinje, Philip Hall, Alexander M. Heberle, Patricia Razquin Navas, Karen van Eunen, Regine C. Tölle, Jennifer J. Schwarz, Heike Wiese, Bettina Warscheid, Jana Deitersen, Björn Stork, Erik Fäßler, Sascha Schäuble, Udo Hahn, Peter Horvatovich, Daryl P. Shanley, and Kathrin Thedieck. A systems study reveals concurrent activation of AMPK and mTOR by amino acids. *Nature Communications*, 7(1):13254, 2016.

[102] Kirsten Thobe, Christine Sers, and Heike Siebert. Unraveling the regulation of mtorc2 using logical modeling. *Cell Communication and Signaling*, 15(1):6, 2017.

[103] Neelanjan Sengupta, P.K. Vinod, and K.V. Venkatesh. Crosstalk between camp-pka and map kinase pathways is a key regulatory design necessary to regulate flo11 expression. *Biophysical Chemistry*, 125(1):59–71, 2007.

[104] Niek Welkenhuysen, Barbara Schnitzer, Linnea Österberg, and Marija Cvijovic. Robustness of nutrient signaling is maintained by interconnectivity between signal transduction pathways. *Frontiers in Physiology*, 9(nil):nil, 2019.

[105] Teeraphan Laomettachit, Katherine C. Chen, William T. Baumann, and John J. Tyson. A model of yeast cell-cycle regulation based on a standard component modeling strategy for protein regulatory networks. *PLOS ONE*, 11(5):e0153738, 2016.

[106] Dong Jian, Zhang Aili, Bai Xiaojia, Zhao Huansheng, and Hu Yun. Feedback Regulation of Ras2 Guanine Nucleotide Exchange Factor (Ras2-GEF) Activity of Cdc25p By Cdc25p Phosphorylation in the Yeast Saccharomyces Cerevisiae. *FEBS Letters*, 584(23):4745–4750, 2010.

[107] Ken Peeters, Frederik Van Leemputte, Baptiste Fischer, Beatriz M. Bonini, Hector Quezada, Maksym Tsytlonok, Dorien Haesen, Ward Vanthienen, Nuno Bernardes, Carmen Bravo Gonzalez-Blas, Veerle Janssens, Peter Tompa, Wim Versées, and Johan M. Thevelein. Fructose-1,6-bisphosphate couples glycolytic flux to activation of ras. *Nature Communications*, 8(1):922, 2017.

[108] Arash Komeili, Karen P. Wedaman, Erin K. O'Shea, and Ted Powers. Mechanism of metabolic control. *The Journal of Cell Biology*, 151(4):863–878, 2000.

[109] Thomas Beck and Michael N. Hall. The tor signalling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature*, 402(6762):689–692, 1999.

[110] Andres Wiemken and Mathias Durr. Characterization of amino acid pools in the vacuolar compartment of saccharomyces cerevisiae. *Archives of Microbiology*, 101(1):45–57, 1974.

[111] Francine Messenguy and Dominique Colin and Jean-Pierre Ten Have. Regulation of compartmentation of amino acid pools in Saccharomyces cerevisiae and its effects on metabolic control. *European Journal of Biochemistry*, 108(2):439–447, 1980.

[112] Sheena Claire Li and Patricia M. Kane. The yeast lysosome-like vacuole: Endpoint and crossroads. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(4):650–663, 2009.

[113] D. Stracka, S. Jozefczuk, F. Rudroff, U. Sauer, and M. N. Hall. Nitrogen source activates tor (target of rapamycin) complex 1 via glutamine and independently of gtr/rag proteins. *Journal of Biological Chemistry*, 289(36):25010–25020, 2014.

[114] Veli C. Özalp, Tina R. Pedersen, Lise J. Nielsen, and Lars F. Olsen. Time-Resolved Measurements of Intracellular ATP in the Yeast *Saccharomyces cerevisiae* using a New Type of Nanobiosensor. *Journal of Biological Chemistry*, 285(48):37579–37588, 2010.

[115] K. Hirimburegama, P. Durnez, J. Keleman, E. Oris, R. Vergauwen, H. Mergelsberg, and J. M. Thevelein. Nutrient-induced activation of trehalase in nutrient-starved cells of the yeast saccharomyces cerevisiae: Camp is not involved as second messenger. *Journal of General Microbiology*, 138(10):2035–2043, 1992.

[116] K. Matsumoto, I. Uno, Y. Oshima, and T. Ishikawa. Isolation and characterization of yeast mutants deficient in adenylate cyclase and camp-dependent protein kinase. *Proceedings of the National Academy of Sciences*, 79(7):2355–2359, 1982.

[117] S. Colombo, D. Ronchetti, J. M. Thevelein, J. Winderickx, and E. Martegani. Activation state of the ras2 protein and glucose-induced signaling in saccharomyces cerevisiae. *Journal of Biological Chemistry*, 279(45):46715–46722, 2004.

[118] Ankit Rohatgi. Webplotdigitizer, 2011.

[119] ImageJ1, Jan 2019. [Online; accessed 2. Apr. 2019].

[120] Wayne A. Wilson, Simon A. Hawley, and D.Grahame Hardie. Glucose repression/derepression in budding yeast: Snf1 protein kinase is activated by phosphorylation under derepressing conditions, and this correlates with a high amp:atp ratio. *Current Biology*, 6(11):1426–1434, 1996.

[121] Ryan Nicholas Gutenkunst, Joshua Waterfall, Fergal Casey, Kevin Brown, Christopher R. Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, preprint(2007):e189, 2005.

[122] I Tavassoly, J Parmar, AN Shajahan-Haq, R Clarke, WT Baumann, and JJ Tyson. Dynamic modeling of the interaction between autophagy and apoptosis in mammalian cells. *CPT: Pharmacometrics & Systems Pharmacology*, 4(4):263–272, 2015.

[123] Jan R. Magnus and H. Neudecker. The commutation matrix: Some properties and applications. *The Annals of Statistics*, 7(2):381–394, 1979.

[124] Ker-Chau Li. On principal hessian directions for data visualization and dimension reduction: Another application of stein's lemma. *Journal of the American Statistical Association*, 87(420):1025–1039, 1992.

[125] L. Valenzuela, C. Aranda, and A. Gonzalez. Tor modulates gcn4-dependent expression of genes turned on by nitrogen limitation. *Journal of Bacteriology*, 183(7):2331–2334, 2001.

[126] Kirk A. Staschke, Souvik Dey, John M. Zaborske, Lakshmi Reddy Palam, Jeanette N. McClintick, Tao Pan, Howard J. Edenberg, and Ronald C. Wek. Integration of general amino acid control and target of rapamycin (tor) regulatory pathways in nitrogen assimilation in yeast. *Journal of Biological Chemistry*, 285(22):16893–16911, 2010.

[127] Zhengchang Liu and Ronald A. Butow. A transcriptional switch in the expression of yeast tricarboxylic acid cycle genes in response to a reduction or loss of respiratory function. *Molecular and Cellular Biology*, 19(10):6720–6728, 1999.

[128] Zhengchang Liu, Takayuki Sekito, Mário Špı'rek, Janet Thornton, and Ronald A Butow. Retrograde signaling is regulated by the dynamic interaction between rtg2p and mks1p. *Molecular cell*, 12(2):401–411, 2003.

[129] H. Ronne D. Balciunas. Yeast genes gis1-4 : Multicopy suppressors of the gal - phenotype of snf1 mig1 srb8/10/11 cells. *Molecular and General Genetics MGG*, 262(4-5):589–599, 1999.

[130] N. Gasmi, P.-E. Jacques, N. Klimova, X. Guo, A. Ricciardi, F. Robert, and B. Turcotte. The switch from fermentation to respiration in saccharomyces cerevisiae is regulated by the ert1 transcriptional activator/repressor. *Genetics*, 198(2):547–560, 2014.

[131] N. C. Barbet, U. Schneider, S. B. Helliwell, I. Stansfield, M. F. Tuite, and M. N. Hall. Tor controls translation initiation and early g1 progression in yeast. *Molecular Biology of the Cell*, 7(1):25–42, 1996.

[132] Ronald A Butow and Narayan G Avadhani. Mitochondrial signaling: the retrograde response. *Molecular cell*, 14(1):1–15, 2004.

[133] M. J. Brauer, C. Huttenhower, E. M. Airoldi, R. Rosenstein, J. C. Matese, D. Gresham, V. M. Boer, O. G. Troyanskaya, and D. Botstein. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Molecular Biology of the Cell*, 19(1):352–367, 2007.

[134] Alejandro A. Granados, Julian M. J. Pietsch, Sarah A. Cepeda-Humerez, Iseabail L. Farquhar, Gašper Tkačik, and Peter S. Swain. Distributed and dynamic intracellular organization of extracellular information. *Proc. Natl. Acad. Sci. U.S.A.*, 115(23):6088–6093, Jun 2018.

[135] A. T. Specht and J. Li. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics*, 33(5):764–766, 03 2017.

[136] Thalia E Chan, Michael PH Stumpf, and Ann C Babtie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell systems*, 5(3):251–267, 2017.

[137] Nan Papili Gao, SM Minhaz Ud-Dean, Olivier Gandrillon, and Rudiyanto Gunawan. Sincerities: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics*, 34(2):258–266, 2018.

[138] Steven Woodhouse, Nir Piterman, Christoph M Wintersteiger, Berthold Göttgens, and Jasmin Fisher. Scns: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC systems biology*, 12(1):59, 2018.

[139] M. Sanchez-Castillo, D. Blanco, I. M. Tienda-Luna, M. C. Carrion, and Y. Huang. A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics*, 34(6):964–970, 03 2018.

[140] Xiaojie Qiu, Arman Rahimzamani, Li Wang, Qi Mao, Timothy Durham, José L McFaline-Figueroa, Lauren Saunders, Cole Trapnell, and Sreeram Kannan. Towards inferring causal gene regulatory networks from single cell expression measurements. *bioRxiv*, 2018.

[141] Thomas Moerman, Sara Aibar Santos, Carmen Bravo GonzÃąlez-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 11 2018.

[142] Pierre-Cyril Aubin-Frankowski and Jean-Philippe Vert. Gene regulation inference from single-cell rna-seq data with linear differential equations and velocity inference. *bioRxiv*, 2018.

[143] Atul Deshpande, Li-Fang Chu, Ron Stewart, and Anthony Gitter. Network inference with granger causality ensembles on single-cell transcriptomic data. *bioRxiv*, 2019.

[144] Joseph Swift and Gloria M. Coruzzi. A matter of time - how transient transcription factor interactions create dynamic gene regulatory networks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1):75–83, 2017.

[145] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. 2009.

[146] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[147] Al Mead. Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 41(1):27–39, 1992.

[148] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, Feb 2018.

[149] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.*, page 1, Apr 2019.

[150] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*, 107(14):6286–6291, 2010.

[151] Luke Zappia, Belinda Phipson, and Alicia Oshlack. Splatter: Simulation of single-cell rna sequencing data. *Genome Biology*, 18(1):174, 2017.

[152] Robrecht Cannoodt, Wouter Saelens, Louise Deconinck, and Yvan Saeys. dyngen: a multi-modal simulator for spearheading new single-cell omics analyses. *bioRxiv*, page 2020.02.06.936971, Feb 2020.

[153] Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12):1482–1492, 2019.

[154] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[155] Payam Dibaeinia and Saurabh Sinha. A single-cell expression simulator guided by gene regulatory networks. 2019.

[156] G. K. Ackers, A. D. Johnson, and M. A. Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133, 1982.

[157] Thomas Schaffter, Daniel Marbach, and Dario Floreano. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, 2011.

[158] Herbert Sizek, Andrew Hamel, Dávid Deritei, Sarah Campbell, and Erzsébet Ravasz Regan. Boolean model of growth signaling, cell cycle and apoptosis predicts the molecular mechanism of aberrant cell cycle progression driven by hyperactive pi3k. *PLOS Computational Biology*, 15(3):e1006402, 2019.

[159] Ryan ODonnell. *Boolean Functions and the Fourier Expansion*. Analysis of Boolean Functions. Cambridge University Press, nil.

[160] Clare E. Giacomantonio and Geoffrey J. Goodhill. A boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Computational Biology*, 6(9):e1000936, 2010.

[161] Anna Lovrics, Yu Gao, Bianka Juhász, István Bock, Helen M. Byrne, András Dinnyés, and Krisztián A. Kovács. Boolean modelling reveals new regulatory connections between transcription factors orchestrating the development of the ventral spinal cord. *PLoS ONE*, 9(11):e111430, 2014.

[162] Jan Krumsiek, Carsten Marr, Timm Schroeder, and Fabian J. Theis. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS ONE*, 6(8):e22649, 2011.

[163] Osiris Ríos, Sara Frias, Alfredo Rodríguez, Susana Kofman, Horacio Merchant, Leda Torres, and Luis Mendoza. A boolean network model of human gonadal sex determination. *Theoretical Biology and Medical Modelling*, 12(1):26, 2015.

[164] T. Powers and P. Walter. Regulation of ribosome biogenesis by the rapamycin-sensitive tor-signaling pathway in saccharomyces cerevisiae. *Molecular Biology of the Cell*, 10(4):987–1000, 1999.

[165] Alan G Hinnebusch. Translational regulation of gcn4 and the general amino acid control of yeast*. *Annu. Rev. Microbiol.*, 59:407–450, 2005.

[166] A. G. Hinnebusch and G. R. Fink. Positive regulation in the general amino acid control of saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 80(17):5374–5378, 1983.

[167] Xinsheng Liao and Ronald A. Butow. Rtg1 and rtg2: Two yeast genes required for a novel path of communication from mitochondria to the nucleus. *Cell*, 72(1):61–71, 1993.

[168] Marie-Pierre Péli-Gulli, Alessandro Sardu, Nicolas Panchaud, Serena Raucci, and Claudio De Virgilio. Amino acids stimulate torc1 through lst4-lst7, a gtpase-activating protein complex for the rag family gtpase gtr2. *Cell Reports*, 13(1):1–7, 2015.