UC Merced
UC Merced Electronic Theses and Dissertations

Title
Understanding the diauxic shift in Saccharomyces cerevisiae and its evolutionarily conserved mechanisms of regulation

Permalink
https://escholarship.org/uc/item/5z61r1tf

Author
Gibb, Edwin Hutcheon

Publication Date
2018

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Understanding the diauxic shift in *Saccharomyces cerevisiae* and its evolutionarily conserved mechanisms of regulation

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Quantitative and Systems Biology

by

Edwin Hutcheon Gibb

Committee in charge:
      Professor David Ardell, Chair
      Professor Patricia LiWang
      Professor Nestor Oviedo

2018

The Dissertation of Edwin Hutcheon Gibb is approved, and it is acceptable
in quality and form for publication on microfilm and electronically:

_____

Dr. Patricia LiWang

_____

Dr. Nestor Oviedo

_____

Dr. David Ardell, Chair

University of California, Merced
2018

DEDICATION

To God, Mom, Dad, and Favor for loving me

Table of contents

List of figures

List of tables

List of abbreviations

^{13}C: carbon-13; 1,5PP2-IP4; 1,5bis-diphosphoinositol tetrakisphosphate; 5PP-IP5: 5-diphosphoinositol pentakisphosphate; ANOVA: analysis of variance; _Ex: model with modifications to reaction pruning order; _g: model with modifications to gene-protein-reaction associations; _lb: model with modifications to lower boundary constraints; _m_n_c: model with modifications to cofactors, nucleoside phosphate moieties, and carbon balancing, also referred to as network reduction; ADH1: alcohol dehydrogenase 1; AU: absorbance units; C: model analyzed by TASMANIAN DEVIL; C_mod: TASMANIAN DEVIL model analyzed by EXAMO; C_mod_eps: TASMANIAN DEVIL model analyzed by EXAMO with adjusted minimum flux threshold; C_orig: original model from EXAMO analyzed by EXAMO; C_orig_eps: original model from EXAMO analyzed by EXAMO with adjusted minimum flux threshold; CDD: Conserved Domain Database; COBRA: COnstraint Based Reconstruction Analysis; EMT: epithelial to mesenchymal transition; ENO: enolase reaction; eps: minimum flux threshold; ETOHt: ethanol transport; EX_etoh(e): ethanol exchange reaction; EX_glc(e): D-glucose exchange reaction; EX_urea(e): urea exchange reaction; EXAMO: EXploration of Alternative Metabolic Optima; FBA: flux balance analysis; fn: false negatives; fp: false positives; FPKM: fragments per kilobase of exon per million fragments; FTHFLm: formate-tetrahydrofolate ligase; FUMm: fumarase mitochondrial reaction; FVA: flux variability analysis; GC-MS: gas chromatography-mass spectrometry; GCR1A: GCR1 is active, GCR1A B: GCR1A Baker; GCR1A J: GCR1A Johnson; wild-type; GCR1D: deletion of GCR1; GEMs: genome-scale metabolic models; GLYCLm: glycine cleavage complex; GIMME: Gene Inactivity Moderated by Metabolism and Expression; GPR: gene-protein-reaction; *H. sapiens*: *Homo sapiens*; HFRs: high frequency reactions; iMAT: Integrated Metabolic Analysis Tool; iMM904: metabolic reconstruction for *Saccharomyces cerevisiae*; kDa: kilodaltons; K_m: Michaelis constant; MBA: model-building algorithm; MFA: metabolic flux analysis; mmol/gDW/h: millimoles per gram dry weight per hour; NAD: nicotinamide adenine dinucleotide; NMR: nuclear magnetic resonance; OD: optical density; PC: pyruvate carboxylase reaction; PPCK: phosphoenolpyruvate carboxykinase reaction; Recon2: metabolic reconstruction for *Homo sapiens*; rH: reactions highly expressed; rL: reactions lowly expressed; rU: reactions with undefined expression; *S. cerevisiae*: *Saccharomyces cerevisiae*; SBML: Systems Biology Markup Language; TASMANIAN DEVIL: The Algorithm for Simplified Metabolic ANalysIs by Altering Networks and Deducing flux Estimates for VIsuaLization; TCA: tricarboxylic acid; TNP: N2-(m-Trifluorobenzyl), N6-(p-nitrobenzyl)purine; tp: true positives; WT: wild-type; YPD: yeast extract peptone dextrose; YPEtoh: yeast extract peptone ethanol; ZFRs: zero frequency reactions

*Note: Metabolite and reaction abbreviations in figures for yeast metabolism follow the naming conventions of the iMM904 model, described on the project's home page: http://bigg.ucsd.edu/models/iMM904.

Acknowledgements

There are many people I need to thank for helping me. First of all, I would like to thank Dr. Fabian Filipp who brought me to Merced and inspired me to think outside of the box. It is because of him that I looked to yeast for insights about how homeostasis mechanisms become deregulated in cancer cells. He also allowed me the freedom to transition from being a benchtop scientist to learning how to be a software developer and data analyst. Perhaps most importantly though, he helped me to discover my passion and calling in life, which is teaching and mentoring students. Without Dr. Filipp, none of this would have been possible. Secondly, I would like to thank Travis Lawrence for helping me design the TASMANIAN DEVIL software. Without Travis' collaboration, the software would not be widely distributable and it would be only a fragment of what it currently is. I learned immensely about coding in Python and developing software from him. I would also like to thank my labmates Rohit Gupta, Lauren Edwards, Simar Singh, Stephen Wilson, Keedrian Olmstead, Asa Gardner, and Theo Crouch for providing insights to my projects. I would also like to thank Theresa Li, who I had the privilege to mentor. Teaching her how to code and analyze data in Python and R were some of my most rewarding memories of graduate school, and I am thrilled that she is continuing her education to become a doctor. I would also like to thank Dr. Tracy Johnson and Dr. Henry Baker for giving me access to transcriptomic datasets for experiments studying GCR1, which were foundational to the dissertation.

I would also like to thank my committee members and advisor, Dr. David Ardell, Dr. Patricia LiWang, Dr. Nestor Oviedo, and Dr. Miriam Barlow. They provided tremendous support during my PhD, especially as I wrapped up my studies. I learned a lot from Dr. Ardell's bioinformatics class about principles of coding, statistics, and reproducible research. He provided me valuable feedback to enhance and validate TASMANIAN DEVIL. I also learned a great deal from the journal clubs with Dr. Patricia LiWang and Dr. Andy LiWang about how to read journal articles critically and from a wide spectrum of fields to better understand a biological system. In particular, Dr. Roger Tseng from these journal clubs helped me think through my projects. Dr. Oviedo taught me to how write research proposals and think critically about my research. Many of the concepts from his cancer genetics and tumor biology course helped me to understand mechanisms whereby yeast homeostasis became deregulated in cancer cells. Dr. Barlow helped me finish my dissertation, providing me advice that saved hundreds of hours. She believed in me and motivated me in so many ways, and I will be forever grateful for her mentorship.

I would also like to thank many people within Graduate Division and the School of Natural Sciences. I would like to thank the Dean of Graduate Education, Dr. Marjorie Zatz, for keeping me on track to finish my dissertation. The Dissertation Boot Camp sponsored by Graduate Division helped me to maintain goals. In particular, I would like to thank Jen Quiralte, Cassie Gunter, and Angela Winek for their help facilitating this productive writing time. I would also like to thank Joy Sanchez-Bell and Paul Roberts, Graduate Student Coordinator and Director of Graduate Programs, respectively, for their encouragement and guidance as I finished my dissertation. They both gave me inspiration about how to start my teaching career, and they provided me the opportunities to build my résumé and ensure that I had a good job immediately upon finishing my PhD.

Finally, I would like to thank my family and friends for their support during the six years I was in graduate school. My parents, Bernadette and Hutch Gibb, celebrated with me and consoled me during the best and the most difficult of times. They have always believed in me, and I am grateful for their continued support. My love, Bendu Favor Ndama, encourages me to be the better man God has called me to be. I look forward to a lifetime of learning and growing with her. I would like to thank all of my friends and spiritual mentors in Chi Alpha, the Wesley Campus Ministry, and Calvary Chapel Merced. In particular, I want to thank the Boyd family,

Vita

EDUCATION

University of California, San Diego   2012
Bachelor of Science, Bioengineering: Biotechnology

University of California, Merced   2018
Doctor of Philosophy, Quantitative and Systems Biology

EMPLOYMENT HISTORY

University of California, Merced
Teaching Assistant
Introduction to Scientific Data Analysis   Fall 2013, 2014 and Spring 2014, 2017
Nutrition   Fall 2015, 2016 and Spring 2016
Preparatory Chemistry   Fall 2017
Concepts and Issues in Biology Today   Spring 2018
Contemporary Biology Laboratory   Spring 2018

Research Assistant
Cancer Metabolism and Systems Biology   Fall 2012 and Spring 2013, 2014
Advisor: Dr. Fabian Filipp

Genomics Institute of the Novartis Research Foundation
Intern
Regenerative Medicine   Jun - Sep 2011, Mar - Aug 2012
Advisor: Dr. Jimmy Elliott

Abstract of the dissertation
Understanding the diauxic shift in *Saccharomyces cerevisiae* and its evolutionarily conserved mechanisms of regulation

by

Edwin Hutcheon Gibb

Doctor of Philosophy in Quantitative and Systems Biology
University of California, Merced, 2018
Chair: Dr. David Ardell

The diauxic shift is a change in metabolism in *Saccharomyces cerevisiae* whereby glucose consumption fuels glycolytic fermentation but then shifts to respiration by ethanol import upon glucose exhaustion. In comparison, cancer cells have increased aerobic glycolysis and lactic acid excretion compared to noncancerous cells. Cancer cells undergo glycolysis upon glucose induction, much like yeast cells do. One protein that regulates the diauxic shift is the glucose response transcription factor, GCR1. After a high growth state and as glucose becomes limited, Gcr1 is deactivated as the cell switches to respiration. By studying the mechanisms and metabolism of the diauxic shift, much can be learned about diseases in other organisms that have deregulated glycolytic metabolism, such as cancer in humans.

First, to model metabolism and to be able to characterize the phenotype of a biological system based off of transcriptomic, proteomic, or genomic datasets, condition-specific metabolic modeling software was developed. The Algorithm for Simplified Metabolic ANalysIs by Altering Networks and Deducing flux Estimates for VIsuaLization (TASMANIAN DEVIL) is comprised of four independently functional modules: gene activity determination, genome-scale metabolic model importation and simplification to reduce network complexity, robust heuristic model building and metabolic flux prediction using steady-state flux balance analysis, and flux visualization from a reference network topology. Publically available transcriptomic and measured flux datasets for yeast were used to validate the software and assumptions used for simplifying networks. TASMANIAN DEVIL is easily installable and can be utilized by researchers in many fields.

After TASMANIAN DEVIL was developed, it was used to character the metabolism of GCR1 mutant transcriptomic studies before and after the diauxic shift. It was modeled that prior to the diauxic shift, there is a high rate of metabolic flux through glycolysis, the pentose phosphate pathway, and biosynthesis, in part generated by anaplerosis. Upon ethanol import, there is a decrease in these pathways and an increase in the citric acid cycle, oxidative phosphorylation, and arginine metabolism. The modeling predicted that the upregulation of the glycine cleavage complex (GCV1, GCV2, and GCV3 in the mitochondria) by GCR1 creates an

efficient way to generate one carbon building blocks for biosynthesis for adenine production along with other purines. Gcr1 is deactivated by inositol pyrophosphorylation from 5-diphosphoinositol pentakisphosphate (5PP-IP5), which is the addition of a phosphate group to a prephosphorylated serine residue through a nonenzymatic cleavage. The process is highly endergonic, requiring a near physiological level Michaelis constant for ATP. This predicts that GCR1 is therefore a regulator of its own deactivation, and it becomes downregulated after a previous high growth state.

Finally, it was investigated how pyrophosphorylation may be conserved in humans to regulate homeostasis. A conserved multi-domain was identified over the residues where pyrophosphorylation occurs on GCR1, which may enable pyrophosphorylation to take place. In humans, several inositol polyphosphatases contain this domain near their phosphatases, perhaps indicating a positive feedback loop for the continued formation of inositol pyrophosphates. GCR1 and its transcriptional partners were also found to be cyclically expressed. The levels of 5PP-IP5 are periodically modulated with the cell cycle in mammals, influencing the activities of the proteins it pyrophosphorylates and binds to, thus governing cell cycle checkpoints. It has been shown that 5PP-IP5 regulates p53-mediated apoptosis upon DNA damage and that it inactivates AKT to downregulate glycolytic metabolism. In cancer cells, there are several ways that the inositol pyrophosphate pathway can become deregulated. However, focusing on reactivating these mechanisms of regulation may provide an effective way to target cancer cells.

Chapter 1   Introduction and overview

1.1   Regulation of the diauxic shift in *Saccharomyces cerevisiae*

The diauxic shift in *Saccharomyces cerevisiae (S. cerevisiae)* is a change in metabolism from fermentation to respiration, resulting in a decreased growth rate. Under its fermentative state, there is a high glycolytic flux fueled by the uptake of glucose, resulting in the efflux of ethanol. This is regardless of the presence of oxygen. This is known as the Crabtree effect [1]. The cells grow rapidly, indicating active biosynthetic pathways to produce the biomass (DNA, RNA, amino acids, carbohydrates, proteins, and lipids). When glucose becomes depleted, there is a switch from glucose to ethanol utilization, whereby the cells begin to respire, as glycolytic flux decreases and the tricarboxylic acid (TCA) cycle and oxidative phosphorylation increases [2]. Upon the diauxic shift, proteins related to glyoxylate metabolism, gluconeogenesis, alcohol import, and oxidative phosphorylation metabolism increase. In addition, the transcription factor Rap1 decreases [3]. Rap1 associates with the transcription factors Gcr1 and Gcr2 to upregulate ribosomal and glycolytic gene transcription [4]. Gcr1's protein levels also rapidly decrease at the beginning of the diauxic shift, as glycolytic transcription is turned off [5, 6]. A better understanding into the mechanism regulating the deactivation of this protein complex as well as the system-wide effects that it has on metabolism will better characterize how the diauxic shift occurs.

1.2   Significance for cancer metabolism

Similar to yeast growth, cancer cells proliferate rapidly and require a metabolic reprogramming to allow for increased biosynthesis. In 1924, Otto Warburg discovered many tumors have increased aerobic glycolysis and lactic acid production and impaired mitochondrial respiration compared to noncancerous cells [7]. This alteration in metabolism has been named the Warburg effect. It was later characterized by Herbert Crabtree that cancer cells undergo aerobic glycolysis upon glucose induction, regardless of the concentration of glucose [8]. However, simply downregulating glycolysis does not usually kill cancer cells. Although cancer cells are traditionally thought to rely on glycolysis for energy, this may differ among cells with a tumor. Some epithelial cancers can respire by the utilization of lactate through several scenarios involving the lactate transporters Mct1 and Mct4 [9]. One of these scenarios is known as metabolic symbiosis, whereby hypoxic cancer cells provide lactate to oxidative cancer cells on the periphery of the tumor [9–11]. Mutations in tumors in certain tissues may also make this transition more favorable. B-Raf mutated melanomas treated with the B-Raf inhibitor, vemurafenib, can be metabolically rescued by oxidative phosphorylation. This may occur in part due to genomic amplifications of genes related to oxidative phosphorylation in some V600E BRAF mutants [12]. Oxidative phosphorylation can occur in response to lactate import in these rescued cells that are characterized by slow growth in G1/0 [11, 13]. Inducing oxidative stress in these vemurafenib-resistant cells by pharmaceuticals results in apoptosis [13, 14]. Since overall metabolism is evolutionarily conserved, a better understanding of the regulation of central carbon metabolism and the diauxic shift in relation to the cell cycle in *S. cerevisiae* can provide insights for human metabolism and cancer, which may have inactivated these homeostasis mechanisms. This may provide personalized targeted pharmaceutical or multidrug approaches for physicians based on a patient's mutations and chromosomal amplifications.

1.3   Mathematical approaches to metabolism

Metabolism is the result of all the upstream components of a system (genes, transcripts, proteins, metabolite uptake and efflux) working together to sustain life. Since metabolism is the end product of a system, the phenotype is therefore largely determined by metabolism. While the most accurate indication of metabolism would come from metabolomics data, this may not be available for every lab or system. Therefore, it is important to be able to predict the phenotype of

a system from various sources of upstream omics data. Mathematically modeling *in silico* flux has widespread applications, ranging from developing a better method of identifying pharmaceutical targets and predicting a pharmaceutical's effect in different tissue types, to optimizing bacterial and yeast strains for biomass production, to characterizing general metabolic phenotypes [15–18]. One approach to model metabolic fluxes is metabolic flux analysis (MFA). MFA is a deterministic manner of modeling metabolism that determines the best fit for the change of modeled metabolite concentrations over time using a system of nonlinear differential equations. Metabolite concentrations are measured by using labeled substrates with gas chromatography-mass spectrometry (GC-MS) or nuclear magnetic resonance (NMR). Fluxes for nutrients into and out of the system also need to be measured. A solution is determined at isotopic steady-state when there is no change in growth rate. This approach is limited due to cost and time. It is also unavailable for many systems, such as for most applications in human cell lines, due to needing to achieve isotopic steady-steady state of labeled substrates. Most applications in MFA do not model full genome-scale models due to too many unknown variables compared to the number of degrees of freedom of the system [19–21]. Genome-scale modelling allows the contextualization of the entire metabolic system. Recently, a similar deterministic approach has been developed which does not require isotopic steady-state, known as unsteady flux balance analysis (FBA). However, this method does not allow for genome-scale modeling either [22]. Another approach is dynamic FBA, which creates time course-dependent kinetic metabolic models using a system of differential equations to model the change of metabolite concentrations over time. This method relies on the measured initial concentrations of metabolites and the kinetic rate constants for reactions, but it is not currently widely used [23]. This method is unavailable for many genome-scale problems due to the underlying biochemical data needed. It also cannot be used as a general analysis for most omics data sets due to relying on the initial concentrations of metabolites. Therefore, a simpler method for predicting metabolism has been utilized, which is called steady-state FBA. This is a stochastic method for modeling metabolism; it uses a linear optimization problem to predict the flux state. This approach eliminates the consideration of time and approximates a solution when there is no change in the concentrations of metabolites [24, 25]. It allows for the use of genomics, transcriptomics, or proteomics data to predict the metabolic phenotype, providing systematic downstream contextualization. **Figure 1.6.1** demonstrates key differences in approaches to genome-scale metabolic modeling.

1.4   Approaches to accounting for metabolite loops

One mathematical problem that arises from a steady-state flux distribution that can distort the accuracy of flux results is the presence of metabolite loops, whereby a metabolite cycles back to a previous node in a reaction pathway. Many of these fluxes are biologically infeasible due to the incorrectly mapped reversibility of reactions or fluxes hitting their extreme upper or lower boundary constraints. Although a reaction may theoretically be reversible, often kinetic rate constants in one direction of the reaction are much larger in the real system. When fluxes hit boundary constraints, this would require a lot of energy and would not be favorable for growth. The first methods to account for thermodynamically infeasible loops were based off of Kirchhoff's second law of electrical circuits, whereby a resistor within a loop cannot have a flux larger than the voltage entering the pathway [26–28]. Similarly, in a biological system, a flux entering a node cannot have a flux within a loop back to that node larger than the flux that entered the pathway. More complex methods have been developed to add additional constraints to the reactions in the loop to prevent their formation, while simpler methods prevent the flux from returning to a node that has already been traversed through changing the boundary constraints and then reoptimizing [28–31]. Another approach minimizes the sum of fluxes within a system to reduce large flux loops [32]. It is imperative for predicting accurate flux results for FBA that loops be reduced.

1.5   Dissertation aims and organization

Biology is increasingly becoming a big data field with the emergence of high-throughput sequencing and proteomics. With the increased amount of data available to computational biologists, the questions researchers can ask and the significance of the results of experiments can be difficult to contextualize without a mathematical model and network-based analysis approach. The opportunity is ripe for molecular biology discovery. However, a better methodology of connecting enriched biological pathways to help predict mechanisms of regulation is necessary. This type of methodology could assist researchers in reverse-engineering metabolic regulation by the extrapolation of the metabolome from genomic, transcriptomic, or proteomic data. In this dissertation, the goals are first to develop a computational platform to model condition-specific metabolism. Next, this software will be used to help elucidate a mechanism controlling the diauxic shift in *S. cerevisiae*, leading to insights of this mechanism's likely deregulation in cancer cells.

The following is a summary of the chapters of the dissertation:

Chapter 1 introduces the central themes for the dissertation.

Chapter 2 presents TASMANIAN DEVIL, a software package to classify gene expression, simplify genome-scale metabolic models to reduce flux loops, model condition-specific metabolism using steady-state FBA, and visualize the flux of nutrients through important pathways.

Chapter 3 uses TASMANIAN DEVIL to model metabolism in *S. cerevisiae* after the deletion of GCR1 in glucose-rich and glucose-exhausted conditions to better understand the regulation of the diauxic shift.

Chapter 4 summarizes the dissertation and provides insights for the importance of inositol pyrophosphorylation, which regulates Gcr1 deactivation. It also identifies a new protein domain on GCR1 that may regulate its dimeric state. While GCR1 may not be conserved in humans, it describes how the deregulation of inositol pyrophosphorylation can cause aberrant growth in cancer cells.

1.5 Figures and tables

Figure 1.5.1   Approaches to genome-scale metabolic modeling

A representation of two approaches to model a simple two reaction metabolic network. The network can be represented by a stoichiometric network of metabolites and fluxes. For dynamic FBA, a system of differential equations is utilized. The change in concentrations of each metabolite is described in terms of the rate the metabolite is formed and used. The fluxes can be separated into their forward and backwards directions, if they are reversible, by using kinetic rate constants. Dynamic FBA relies on measured biochemical data to calculate kinetic rate constants. It also requires experimenters to measure the initial concentrations of metabolites. A solution is approximated when there is no change in metabolite concentrations over time. In comparison, steady-state FBA employs linear optimization to model when the change in metabolite concentrations over time (d$\mathbf{x}$/dt), or in other words the stoichiometric matrix of metabolites and reactions (**S**) multiplied by the flux vector (**v**), is equal to 0. A set of upper and lower boundary constraint vectors (**lb** and **ub** respectively) are used to define the parameters for possible fluxes. An objective function is set (Z), whereby a weighted vector, **c**, is multiplied by **v**. Possible objective functions include maximizing biomass production or minimizing or maximizing the sum of all fluxes. The solution space for the two vectors is shown with the objection function solution for the maximization of the sum of fluxes pinpointed.

	Network	Dynamic FBA	Steady-State FBA
Mathematical Representation	$S = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{cc} \bar{v}_1 & \bar{v}_2 \\ \left[\begin{array}{cc} -1 & 0 \\ 1 & -1 \\ 1 & -1 \\ 0 & 1 \end{array}\right] \end{array}$	$\dfrac{dA}{dt} = -k_1 A + k_{-1}BC \qquad \dfrac{dC}{dt} = k_1 A - k_{-1}BC - k_2 BC$ $\dfrac{dB}{dt} = k_1 A - k_{-1}BC - k_2 BC \qquad \dfrac{dD}{dt} = k_2 BC$ $[A_0] = 0.833M \quad [B_0] = 0.025M \quad [C_0] = [D_0] = 0M$ $k_1 = 0.001Ms^{-1} \quad k_{-1} = 0.000001Ms^{-1} \quad k_2 = 0.0001Ms^{-1}$	$\dfrac{dx}{dt} = S \cdot v = 0$ $lb \le v \le ub$ $\text{min or max } Z = \sum_{i=1}^{N} c_i v_i$
Graphical Representation		solution when $\dfrac{d\bar{x}}{dt} = 0$	optimal solution for max Z

4

Chapter 2   TASMANIAN DEVIL: a software package for classifying gene activity from omics data sets, simplifying metabolic networks, and visualizing the estimated phenotypic fluxes of nutrients

2.1   Abstract

2.1.1   Background

The integration of experimental data to predict phenotypic changes in genome-scale metabolic reconstructions is an active field of research in systems biology. It is imperative to develop software that can generate condition-specific models across a wide range of organisms and data types.

2.1.2   Results

We have developed The Algorithm for Simplified Metabolic ANalysIs by Altering Networks and Deducing flux Estimates for VIsuaLization (TASMANIAN DEVIL). There are four separate modules that comprise this software package: gene activity determination, genome-scale metabolic model importation and simplification to reduce network complexity, robust heuristic model building and metabolic flux prediction, and flux visualization from a reference network topology. All modules can be utilized independently or in conjunction with one another. The methods for simplifying models to reduce cofactors and nucleoside phosphate moieties to enhance carbon-centered fluxes are novel. To test the validity of making these changes, publically available experimentally determined flux calculations from aerobic and anaerobic cultures of *S. cerevisiae* were compared to the modeled fluxes generated by our software from gene activity rules derived from paired transcriptomic data sets for these conditions. These simplifications produced significantly more accurate flux predictions compared to fluxes from unmodified models. The software implementation also significantly outperformed a previous release of the model building algorithm in essential reaction determination based off of publically available lethal knockout experiments with paired transcriptomic data sets.

2.1.3   Conclusions

We have designed software that can import the metabolic reconstruction of any model organism and predict and visualize fluxes for a condition based on gene activity rules. We have also proposed and tested a paradigm change to metabolic modeling to decrease the inherent complexity of the models, which demonstrated to be beneficial for flux predictions. Advancements to the model building algorithm enhanced essential reaction determination and flux prediction. Predicting condition-specific *in silico* fluxes has widespread applications ranging from engineering bacterial and yeast strains for biomass production, to predicting tissue-specific models for drug discovery, to characterizing general metabolic phenotypes for an experiment. TASMANIAN DEVIL has the potential to be adopted by a wide spectrum of researchers using Linux or macOS platforms and is freely available.[1]

---

[1] Please email ehgibb1@gmail.com for more information.

2.2 Background

2.2.1 Condition-specific metabolic modeling algorithms

Several algorithms have been developed for steady-state FBA to attempt to generate more accurate constraint-based models for tissue-specific or condition-specific problems using upstream omics data sets [33–39]. Many of these algorithms differ in what they are trying to solve. The Integrated Metabolic Analysis Tool (iMAT) implements user-defined discretized gene rules to identify a core set of reactions [40]. Gene Inactivity Moderated by Metabolism and Expression (GIMME) requires thresholding of gene expression data to define which genes are predicted to be on or off, while the method in Lee et al. uses continuous expression data to weight the importance of fluxes for reactions when solving for a flux state [41, 42]. E-Flux actually incorporates gene rules or discretized expression of genes directly into the optimization by adjusting flux boundaries [43]. Some algorithms require multiple data sets or time course experiments to predict flux states [44–46]. Other methods primarily focus on building a model from a core set of reactions that are important for a metabolic function rather than on predicting a flux state; this type of approach is known as a model-building algorithm (MBA) [47–49]. EXploration of Alternative Metabolic Optima (EXAMO), developed by Rossell et al., uses an MBA to build phenotypically predictive models with only one data set from a core set of reactions using discretized gene rules [32]. However, it was difficult to use the algorithm across platforms, to use other metabolic reconstructions, and the algorithm did not always succeed in generating a flux result with a given gene rule set. Many of the other algorithms for condition-specific modeling are also difficult to implement. Some of them have unannotated code without example files of how to use the algorithms. Other methodology papers only contain the theory rather than the files and software used.

2.2.2 Reducing metabolite loops through network simplification

Not all flux loops that exist within a network are large and infeasible. Loops can also exist in much smaller proportions in order to facilitate the creation of another metabolite needed for biomass productions, serve as a layer of cellular regulation, or dissipate heat [50, 51]. Thus, to limit the formation of any loops within a network may create biologically infeasible results. The interconnectedness of metabolites in the network in terms of cofactors and nucleoside phosphate moieties creates many flux loops to account for the cycling of these metabolites. Nucleoside phosphate and cofactor moieties contribute to many of the phosphate or hydrogen transfers in a network. One result of FBA when there is such high interconnectedness of metabolites is that the predicted fluxes are greater than the physiological fluxes due to the system adjusting for the cycling of metabolites. Therefore, it may be beneficial to eliminate these metabolites from reactions that do not contribute to biosynthesis. This would focus the fluxes of the network on carbon-driven biosynthesis of metabolites in the biomass reaction. This approach is similar to the models used in MFA [52–56].

2.2.3 Inspiration for study

We have redeveloped the EXAMO software as part of The Algorithm for Simplified Metabolic ANalysIs by Altering Networks and Deducing flux Estimates for VIsuaLization (TASMANIAN DEVIL) package. It is easily installable and operable via the command. We have tested it and validated it using publically available data sets for *S. cerevisiae* to show that the algorithm outperformed the original implementation in gene inclusion and flux accuracy. By creating a simplified metabolic network and eliminating cofactors and nucleoside phosphate moieties that do not contribute to direct biosynthesis, the modeled fluxes more accurately resembled the estimated fluxes.

2.3   Implementation

2.3.1   Software architecture

TASMANIAN DEVIL is an open source software package developed in Python and executed through the command line to generate condition-specific models. It is comprised of four modules that can be used separately or in conjunction with one another: one for model importing, adaptation, and simplification, another for gene classification of activity, a third for predicting fluxes from the gene activity rules, and finally one for visualizing the fluxes through specified metabolic networks. A maintained version of the program is available on GitHub, which contains detailed protocols for each module.[2] **Figure 2.6.1** demonstrates the general workflow for the package and the potential applications.

2.3.1.1   *Model module*

Metabolic reconstructions are generally encoded in Systems Biology Markup Language (SBML) and COnstraint Based Reconstruction Analysis (COBRA) models; the model importing module is able to import these models, adapt, export, and then use these models for flux prediction [57, 58]. This module contains the capability of simplifying the model in several different ways if the user desires, such as only including carbon balanced reactions, removing metabolites without carbons, and converting all nucleoside phosphate and cofactor moieties into commonly shared metabolites. These simplifications focus fluxes on carbon-driven reactions contributing to biosynthesis. There are also options within the module to change the upper or lower boundary constraints for reactions or to alter the gene-protein-reaction (GPR) associations.

2.3.1.2   *Gene module*

Gene activity rules are often needed for condition-specific modeling programs. In the gene classification module, the user loads a numerical data set with genes and values from transcriptomic or proteomic experiments. The user chooses upper and lower percent thresholds by which to define active and inactive genes, respectively. For example, if there are 904 genes in the metabolic reconstruction, like there are for the *S. cerevisiae* metabolic reconstruction (iMM904), the program first filters for those genes in the data set [59]. It then employs the user chosen thresholds, such as the top 25% of expressed genes are active and the bottom 25% of genes are inactive, to define gene activity. The gene rules can then be used for the flux prediction module to predict reaction activity to create condition-specific models. If the user wants to classify gene activity in a more sophisticated manner by using differential abundance analysis, the user can generate their own gene rules for the flux prediction module.

2.3.1.3   *Flux module*

The flux prediction module contains a robust and updated implementation of the EXAMO software [32]. Every generated reduced model in the profile is now solvable itself. An additional user-specified option now includes selecting the number of repetitions to build functional models from reaction profiles to compare more than one final flux state. Another additional option is selecting the order to remove reactions. By first attempting to prune extracellular reactions, then extracellular transport reactions, then compartmental transport reactions, and then inner-compartmental reactions, more of an emphasis is placed on the importance of transport reaction gene activity for defining the metabolic state of the system.

2.3.1.4   *Visualization module*

Fluxes are able to be mapped onto user-generated metabolic pathways designed in CellDesigner [60]. The flux files can be used from the results of the flux prediction module, or they can be imported from another program using the same file format. In brief, the user creates a graphical representation of the reactions of interest using the CellDesigner interface. The pathway

---

[2] Please email ehgibb1@gmail.com for more information.

is then saved as an SBML file. The module is then executed to make the proper adjustments to the pathway file, like broadening the width of the reaction lines to represent fluxes, choosing a direction of flux for reversible fluxes, and dimming metabolites and reactions that are predicted to be removed from the model.

2.3.2 EXAMO algorithm and updates

The flux module contains an updated implementation of the EXAMO algorithm. Previously, the algorithm had been comprised of four Python scripts. The first script accepted a set of gene rules, and it classified reactions as being highly or lowly expressed (rH or rL, respectively) or reactions with undefined expression (rU) based off of the GPR associations. It used a mixed integer linear program using the methodology from iMAT to find the optimal reaction classification system using flux variability analysis (FVA), whereby the most rH reactions and the fewest rL reactions were identified to have a flux [40, 61]. What this equated to was a maximization of pathway length to agree with the reaction rules, so that the most favorable path was found to achieve biomass production (when the biomass reaction had a nonzero lower boundary constraint). Reactions always or never included in the set of most optimal classification solutions were classified as high frequency reactions (HFRs) or zero frequency reactions (ZFRs), respectively. These reactions were then used as the basis for inclusion or exclusion in the reduced condition-specific models using the MBA [47]. ZFRs were removed first from consideration. HFRs were preserved and were not able to be pruned from the model. Reactions were heuristically removed one-by-one from the model unless a pruned reaction forced any of the HFRs to be unable to produce a flux. Any resulting reactions that became inactive along a pathway due to the removal of a pruned reaction were also removed themselves. A profile of pruned models was thereby created for a user-defined number of repetitions the MBA was to be completed. The third script added reactions back one-by-one to the set of HFRs in the order of the frequency of the reactions in the model profile, until all of the HFRs were able to produce a flux [47]. The fourth script then found a steady-state flux solution for a biomass producing system by minimizing the sum of fluxes as the objective function to minimize metabolic feedback loops in the system [32].

In the new implementation included in this module, the four scripts are adapted into one to ensure robustness of minimizing the model and being able to determine a nonzero biomass flux solution without failing. Other adaptations to these scripts included preserving GPR parsing to account for AND and OR logic gates for proteins in complex or if there are isoforms or separate complexes available for a reaction. If a reaction had multiple gene mapping gates, if any of the proteins or protein complexes produced a reaction defined to be rH or rU, this was preferred in determining reaction activity compared to one that produced a rL, since that reaction would still theoretically have enough of the other protein present to enable its enzymatic function. The order of pruning reactions in the MBA is also randomized using the Python Cryptography Toolkit to prevent each repetition from producing similar results to one another.

2.3.3 Validation

2.3.3.1 Model importing, adaptation, and simplification

To test TASMANIAN DEVIL, the iMM904 metabolic reconstruction with adaptations for nicotinamide adenine dinucleotide (NAD) biosynthesis was used as the base model [59, 62]. One additional change, in accordance with the recent metabolic reconstruction for *Homo sapiens (H. sapiens)*, Recon2, was to make the reaction for mitochondrial formate-tetrahydrofolate ligase (FTHFLm) reversible [63]. Another change was adding ADP as a reactant to the biomass equation and modifying the stoichiometric coefficients for ADP and ATP to reflect their biological ratios to AMP [64]. We performed testing for EXAMO using the model for iMM904 with NAD adaptations included in the original software distribution; we also tested the software using the default model without additional parameters created by TASMANIAN DEVIL for the

iMM904 model with NAD adaptations, FTHFLm modification, and adenosine nucleotide modifications.

A description of all testing parameters for TASMANIAN DEVIL from user-defined options is described in **Table 2.6.1**. A list of the GPR associations changed for extracellular metabolite uptake is described in **Table 2.6.2**, in accordance with only direct experimental evidence from the Saccharomyces Genome Database [65]. If one of these genes was mapped to many transport reactions, it was generally not included in the mapping. By default, the lower boundary constraints were left open as in the original EXAMO study. For each condition, the composition of the extracellular media was used to determine the maximum exchange rate. A description of the lower boundary constraint changes for extracellular metabolite uptake for each condition tested is described in **Table 2.6.3**. For validation of the flux module, if a set of parameters was used for TASMANIAN DEVIL that did not change the lower boundary constraints to define metabolite uptake, the lower boundary constraint for the minimum biomass production was set to be the maximal biomass production for the default unmodified iMM904 model (0.2879 millimoles per gram dry weight per hour (mmol/gDW/h)). We used this same constraint for biomass production when testing EXAMO as well. This guaranteed nonzero fluxes through the system when the objective function was set to minimize the sum of fluxes. The cofactor and nucleoside phosphate moiety simplifications are described in **Table 2.6.4**. Examples of simplifications and reduced model files for iMM904 with adaptations for NAD biosynthesis, FTHFLm, and adenosine nucleotide modifications, as well as simplifications for the metabolic reconstruction for *H. sapiens* (Recon2), can be found in the test data at the project home page.

2.3.3.2   Gene specificity and precision

We tested how well TASMANIAN DEVIL removed reactions from the GEMs for the flux module to create biologically meaningful condition-specific models. Gene rules were defined from transcriptomic data sets from Gasch et al. for *S. cerevisiae* cultures grown in glucose and ethanol and compared to essential genes based off of gene-deletion mutant studies by Giaever et al. and Snitkin et al. in glucose and ethanol, respectively [66–68]. In brief, the microarray data sets were downloaded using GEOquery, and the probes were normalized across genes using affyPLM in R [69, 70]. The genes in the iMM904 model with expression in the highest and lowest 25% of genes were classified as active and inactive, respectively, using the gene module, following similar expression thresholds in other studies [34, 42]. These results were compared against a negative control, whereby 25% of the genes were randomly assigned active and inactive rules.

The performance of TASMANIAN DEVIL was also compared to EXAMO. Combinations of the **Table 2.6.1** parameters were used to test the TASMANIAN DEVIL package. The detailed concentrations of yeast extract peptone dextrose (YPD) containing glucose and yeast extract peptone ethanol (YPEtoh), with the glucose replaced by ethanol, and with calculated amino acid and vitamin composition according to the analysis by the manufacturer were used as alternative lower boundaries for testing. The nucleotide concentrations in the extracts based on a compositional study were also used as alternative lower boundaries [71]. In terms of parameters changed for EXAMO, the algorithm was run using the original minimum flux threshold (eps) of 1E-3 mmol/gDW/h, as well as with an eps of 1E-10 mmol/gDW/h for the final script when optimizing the pruned model. 50 iterations of pruning the models for each reaction profile were performed, and the process was repeated 5 times.

For the gene-deletion mutant studies used as a reference to characterize sensitivity and precision, essential genes were classified as homozygous mutants producing slow to no growth. True positives (*tp*) refer to included reactions that were deemed essential and were experimentally, false positives (fp) were genes that were deemed essential but were not experimentally, and false negatives (fn) were genes that were removed from the model but were

actually essential. Sensitivity was calculated to be *tp/(tp+fn)*, and precision was determined from *tp/(tp+fp)*. Similar to other studies, genes essential in YPD were not considered for testing when determining gene essentiality for YPEtoh, as YPD essential genes were considered to be essential in all conditions in the original mutant study due to culture media history [32, 68].

2.3.3.3 *Flux accuracy*

To analyze the accuracy of TASMANIAN DEVIL predicted fluxes, fluxes calculated from aerobic and anaerobic transcriptomic data sets were compared to fluxes calculated from MFA experiments in the same conditions. Gene rules were defined from microarray experiments from Rintala et al. at 20.9% and 0% oxygen concentrations in glucose-limited cultures [72]. The microarray data sets were downloaded using GEOquery in R, and the probes across replicates and genes were averaged [69]. The 25% expression threshold was again used to define gene activity, and these were compared to a negative control which randomly assigned gene activity and inactivity to 25% of the genes.

To compare the flux accuracy of TASMANIAN DEVIL against EXAMO, the same experimental groups were utilized as those for the gene sensitivity and precision validation. The parameters described in **Table 2.6.1** were again used to test flux accuracy for TASMANIAN DEVIL. The detailed concentrations of components in the media were used for determining the alternative lower boundary constraints for extracellular metabolite uptake when formatting the model [73]. 5 repetitions with 50 iterations were executed.

Fluxes from EXAMO and TASMANIAN DEVIL were compared to fluxes fitted by the MFA program METAFoR for a simplified metabolic network using carbon-13 (^{13}C) labeled 2D NMR tracer experiments for the same strains in the same conditions as in Rintala et al.'s study [52, 72, 74]. The total absolute differences in fluxes were calculated for the overlapping reactions that appeared in both the iMM904 model and the MFA model to determine accuracy. The flux per reaction per pathway was determined by taking the average of the fluxes of the reactions for each pathway in which the reaction was not pruned. Since MFA modeling uses ^{13}C labeled metabolite data and uptake rates for reactions from extracellular media, the condition-specific modeling algorithm with the results most similar to these experimentally fitted models produced the most accurate flux results.

2.4 Results

2.4.1 TASMANIAN DEVIL is robust and achieves high sensitivity for reaction inclusion

We compared gene sensitivity and precision of TASMANIAN DEVIL to EXAMO to analyze how well each algorithm removed reactions mapped to genes according to experimentally validated essential genes. **Figure 2.6.2a** and **Figure 2.6.2b** demonstrate the sensitivity and precision of the algorithms and their parameters. TASMANIAN DEVIL was more sensitive than EXAMO when an unmodified model was used. The model analyzed in TASMANIAN DEVIL (C) without modifications outperformed EXAMO for sensitivity compared to the original model (C_orig), increasing the sensitivity by 16.4% ± 1.2% for glucose and by 25.2% ± 8.1% for ethanol. The precisions were comparable for these two specifications. The same unmodified model used for TASMANIAN DEVIL was not solvable for most of the conditions in EXAMO (C_mod and C_mod_eps), as shown by **Figure 2.6.2c**. All TASMANIAN DEVIL models were able to be pruned and were solvable. In contrast, most models from the original implementation of EXAMO were able to be pruned, but not all repetitions were solvable.

When we manipulated the parameters described in **Table 2.6.1**, sensitivity and precision did not increase for many of the parameters that were tested for TASMANIAN DEVIL, although there was slight improvement in sensitivity after changing the order the reactions were removed from the model (C_Ex and C_g_Ex). Otherwise, for the purpose of predicting essential genes, the

unmodified model (C) performed better than most other parameters, having some of the largest differences between a condition and the negative control for that condition for both sensitivity and precision. C, C_Ex, C_g, and C_g_Ex had greater sensitivities and precision for the gene rules created from glucose and ethanol transcriptomic data compared to the negative controls. When the lower boundary was adjusted as one of the parameters changed (C_lb, C_lb_Ex, C_lb_g, and C_lb_g_Ex), the sensitivities and precisions were comparable between the conditions and the negative controls, indicating that the gene rules for each condition were not able to create more predictive models compared to randomly assigned genes. When the model was reduced to account for nucleoside phosphate moieties, cofactors, and carbon balancing (for the sets of parameters including _m_n_c, hereafter referenced as network reduction), there was a large drop in sensitivity and precision for the ethanol related models. The results demonstrated that while precision was comparable across the packages, TASMANIAN DEVIL was always more robust; it was also more sensitive compared to EXAMO when using an unmodified input model.

2.4.2   TASMANIAN DEVIL produced more accurate flux prediction

Fluxes produced by TASMANIAN DEVIL and EXAMO were compared to MFA modeled fluxes for models created from gene rules from aerobic and anaerobic conditions in *S. cerevisiae*. **Figure 2.6.3** shows the flux profiles for EXAMO and TASMANIAN DEVIL for the different parameters tested in comparison to the MFA fluxes. TASMANIAN DEVIL produced more accurate flux states than EXAMO when comparing the total absolute flux difference. The aerobic condition with the network reduction parameter selected produced the most accurate flux results compared to the MFA simulated fluxes. These modeled gene rules also had more accurate flux results than their negative control counterparts. The anaerobic condition had higher total absolute flux differences, but the modeled gene rules produced more accurate flux results than the negative control as well. **Figure 2.6.4a** demonstrates that the total absolute flux difference between gene rules for the aerobic and anaerobic conditions compared to the negative controls were significant for EXAMO when using an eps of 1E-10 mmol/gDW/h and for TASMANIAN DEVIL for network reduction models ($p < 0.001$, using analysis of variance tests (ANOVAs) with replicates, but the results were insignificant for TASMANIAN DEVIL when network reduction was not adjusted compared to the negative controls. **Figure 2.6.4b** highlights that the total absolute flux differences from TASMANIAN DEVIL were more accurate compared to EXAMO, and the difference between the two algorithms was highly significant for both the aerobic and anaerobic conditions ($p < 0.001$).

The network reduction models in TASMANIAN DEVIL created the most phenotypically predictive models. As can be gleaned from **Figure 2.6.3**, the set of parameters with the lowest combined absolute flux difference between TASMANIAN DEVIL and the MFA implementation for the aerobic and anaerobic conditions was C_m_n_c_lb_g. **Figure 2.6.4c** shows that the anaerobic condition had larger glycolytic and extracellular fluxes for uptake and excretion, while the citric acid cycle and oxidative phosphorylation pathways had greater fluxes for the aerobic conditions. These trends matched the fitted MFA results for the two conditions [52]. **Figure 2.6.5** shows a visual representation of the flux network created by CellDesigner using the visualization module of TASMANIAN DEVIL for glycolytic-related pathways, the pentose phosphate pathway, glycine/serine metabolism, and alanine biosynthesis. Methods were also developed to visualize GPR rule determinations and FVA/MBA results in relation to a network, as demonstrated in **Figure 2.6.6** and **Figure 2.6.7,** respectively. TASMANIAN DEVIL predicted fluxes more accurately than EXAMO; the results also indicated that the network reduction models created the most phenotypically predictive fluxes.

2.5 Discussion

2.5.1 Performance comparison to previous software and recommendations for parameter usage

TASMANIAN DEVIL was more robust, more sensitive, and had more accurate flux predictions than EXAMO. When unmodified models were used, TASMANIAN DEVIL had increased sensitivity in predicting essential genes. If the user is solely focused on predicting essential genes, we would recommend using the unmodified model or just changing the order of reaction pruning as the only parameter adjusted. The network reduction modifications had lower sensitivity and precision, sometimes even resulting in the randomly assigned negative control gene rules having better gene essentiality predictions than the gene rules created from the data set of interest. TASMANIAN DEVIL also produced more accurate flux predictions than EXAMO from the condition-specific models it generated. For the purpose of predicting fluxes, we would recommend utilizing simplified models that reduce the complexity of the interconnectedness of the network and focus on carbon-centric fluxes, as these simplified models produced the most accurate flux results in relation to MFA fitted fluxes. If information is readily available for the composition of extracellular media, we would advise making changes to the lower boundary constraints as well when creating the model. Finally, because the most accurate flux results were generated from the C_m_n_c_lb_g parameter, we would advise checking gene rules for extracellular transport reactions as well, as this may influence the results. The flux results phenotypically resembled the conditions being modeled.

2.5.2 Analysis of testing limitations

We acknowledge that the increased robusticity of TASMANIAN DEVIL was in part due to changing eps, but it was also caused by modifying the approach to the MBA. When eps was maintained at its original value of 1E-3 mmol/gDW/h, EXAMO was not able to optimize for any condition when using the model exported from TASMANIAN DEVIL (C_mod). When eps was lowered to 1E-10 mmol/gDW/h, it was able to find a solution for all of the conditions except for the gene rules from glucose (C_mod_eps). This may in part be a result of not providing EXAMO a degapped model as it states it needs, whereby all dead-end metabolites and reactions have been removed, since the degapped model provided with the algorithm was able to produce more solutions (C_orig) than C_mod. However, it was also likely that the lower eps threshold played a key role in solvability, as the solvable repetitions increased from the higher eps to the lower eps for both the original and TASMANIAN DEVIL exported models. Other factors likely affecting the increased solvability from EXAMO to TASMANIAN DEVIL were the randomization of attempting to remove reactions and checking that models were still solvable after every reaction was removed for every iteration of the MBA process. TASMANIAN DEVIL was always able to find a solution for every iteration of every repetition; whereas for EXAMO, not every iteration created a pruned model and not every repetition was solvable.

Additional limitations of the validation of the modeling approaches were that we only tested TASMANIAN DEVIL with *S. cerevisiae* data and did not comprehensively compare it against algorithms that reduce inherent flux loops. Although we demonstrated that the software generally characterized the correct phenotype for yeast, we have not demonstrated this for other organisms. The study was in part conducted in yeast because of the availability of gene essentiality results and publically available MFA fluxes under various conditions. Validating the software in *S. cerevisiae* helped contrast the findings with EXAMO, which was also validated with the same model organism. *S. cerevisiae* have a higher correlation of mRNA abundance to protein abundance than more complex organisms do, making transcriptional condition-specific studies more relevant for phenotypic prediction [75]. For organisms with more transcriptional regulation such as *H. sapiens*, we would recommend using protein abundance data if available. Another limitation is that we did not compare the flux results of the network reduction

modifications with algorithms which attempt to remove flux loops completely. These other algorithms reduce the overall fluxes in the system as well through their methodology, which often make the modeled fluxes closer to the physiological fluxes.

2.5.3   User interface and intended uses

TASMANIAN DEVIL is easily installable and operational via the command line, increasing the number of potential users. The user manual provides documentation about how to use each module of the software with the different parameters offered. The package on GitHub also provides examples of usage. **Figure 2.6.1** highlights some general intended uses. Since each module can function independently, the inputs and outputs show what is needed and produced for TASMANIAN DEVIL. The package can be used for other modeling purposes, preliminary data analysis, and visualization. The model module exports the model as a COBRA file for the capability of modeling in other programs. The flux module exports the flux results as .csv files which can be further analyzed. The gene module can be used as a first analysis for other applications. Finally, the visualization module can illustrate fluxes for predefined pathways from flux distributions from other algorithms.

2.5   Conclusions

We have developed and tested the Python software package TASMANIAN DEVIL for condition-specific metabolic modeling. The package is divided into four separately functioning modules: importing, modifying, and simplifying metabolic models; gene activity characterization; flux estimation; and flux visualization. We have presented a few applications and validated the flux estimation software and modeling assumptions using data from *S. cerevisiae*. The approach to the simplification of metabolic networks for FBA is novel and may be applied for modeling purposes outside of the scope of the software. TASMANIAN DEVIL has the potential to be used by a wide array of researchers for metabolic modeling purposes.

2.6   Figures and tables

Figure 2.6.1   TASMANIAN DEVIL architecture

Workflow for TASMANIAN DEVIL demonstrating the interconnectedness and independence of the model, gene, flux, and visualization modules. The inputs for each model, the necessary and optional arguments from the command line, and the general outputs are shown.
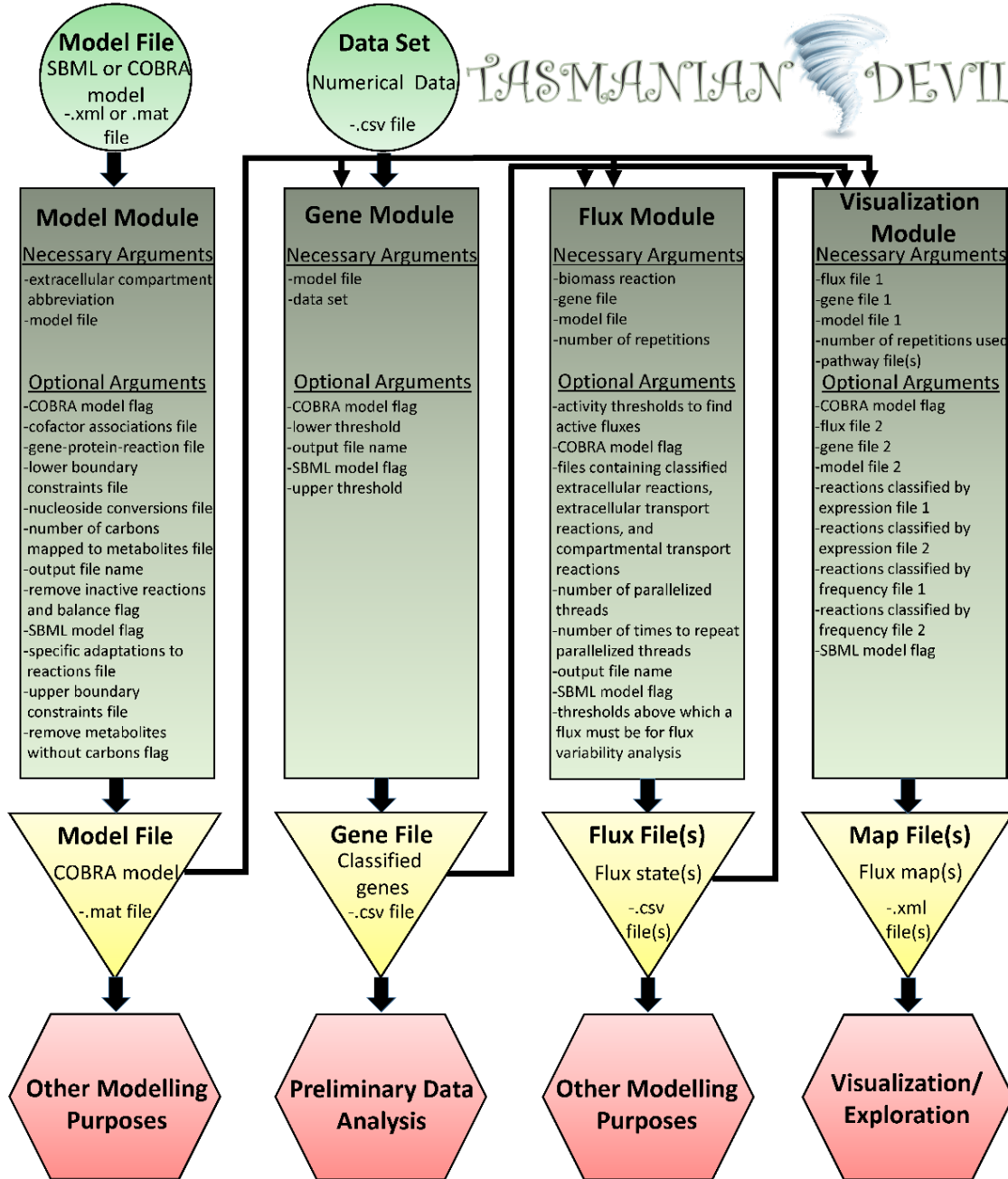
Figure 2.6.2   Sensitivity, precision, and solvability of conditions and software
Sensitivity (**a**) and precision (**b**) of reaction inclusion based off of experimental gene knockout data were analyzed for the condition-specific models generated from the gene rules for ethanol, glucose, and their respective negative controls. For the EXAMO software using the original model included in the distribution (shown in light blue) and for the EXAMO software using the same model that was tested in TASMANIAN DEVIL (shown in light orange), eps was lowered and changed to 1E-10 mmol/gDW/h when solving the reduced network (C_orig_eps and C_mod_eps, respectively). TASMANIAN DEVIL was tested using combinations of model parameters as described in **Table 2.6.1** (shown in light red). If all repetitions of a condition for a parameter were not solvable, the sensitivity and precision were 0. (**c**) A comparison of the solvability of pruned models and optimizations.
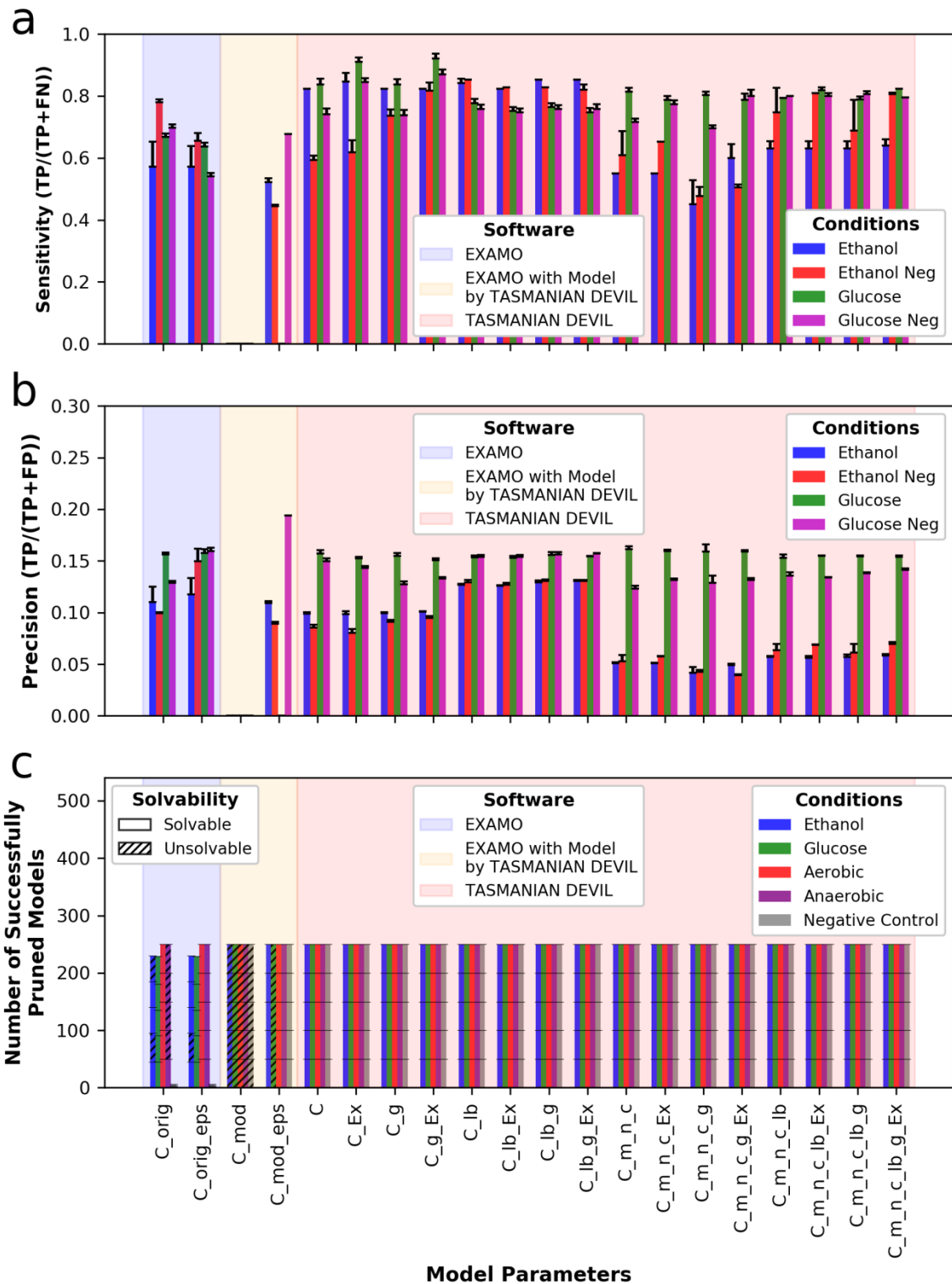
Figure 2.6.3   Comparison of accuracy of flux profiles
Simulated fluxes from condition-specific models for aerobic and anaerobic gene rules and their
negative controls were compared to experimentally modeled flux results using NMR tracer
experiments under the same conditions. This was performed for both the gene rules for the
condition and for the negative control gene rules for the condition. The absolute flux difference
was determined for each reaction between the experimental and simulated results. If a reaction
corresponded to more than one reaction in the larger metabolic reconstruction, the redundant
reactions were noted, and the fluxes were added together for those reactions. The absolute flux
difference for each reaction for the aerobic and anaerobic gene rules was then compared against
the negative controls, and the difference was plotted by metabolic pathway. Darker cells represent
measured fluxes by which the gene rules were more accurate than the negative controls. The total
absolute flux difference of all reactions for a parameter for the aerobic and anaerobic gene rules is
noted at the top of each column of the heatmaps.
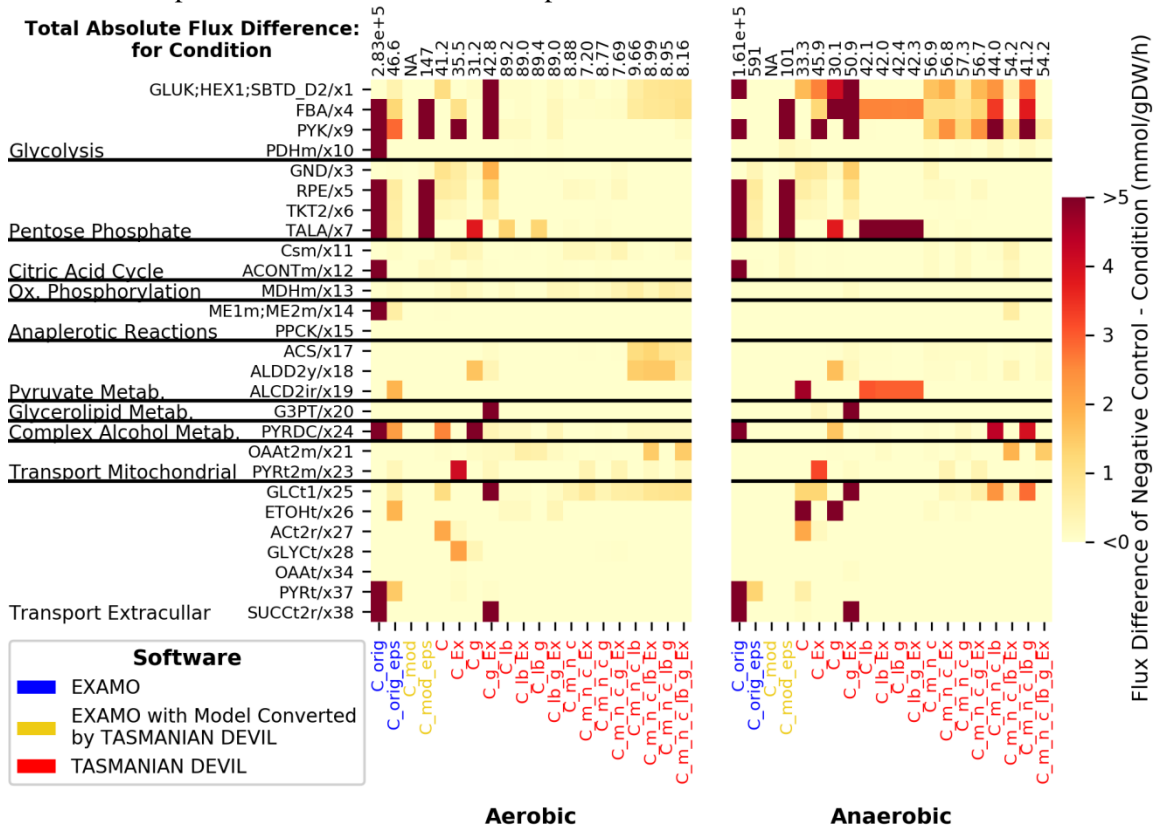
Figure 2.6.4   Comparison of flux predictions

Flux differences were compared for the aerobic and anaerobic conditions versus the negative control for each platform (**a**) and across platforms (**b**). For TASMANIAN DEVIL, the model parameters with network reduction were combined to make a C_m_n_c Cohort, and all model parameters excluding network reduction were also combined to make a C_non_m_n_c Cohort. For (**a**), a two-way ANOVA with replicates was performed on the total absolute flux difference of the five replicates for each model parameter, comparing the negative control versus each condition for a software platform and accounting for each parameter group for the cohorts of TASMANIAN DEVIL, shown in light red. A one-way ANOVA with replicates was used for testing the difference between the negative control versus each condition for the original EXAMO software with the model converted by TASMANIAN DEVIL, shown in light yellow. For (**b**), a two-way ANOVA with replicates was used to compare the total absolute flux difference of the original software with the two cohorts for TASMANIAN DEVIL while accounting for the model parameter groups. (**c**) A representation of fluxes per reaction per pathway of interest for the aerobic and anaerobic conditions for the C_m_n_c_lb_g parameter combination, which produced the least combined total absolute flux difference for the aerobic and anaerobic conditions.
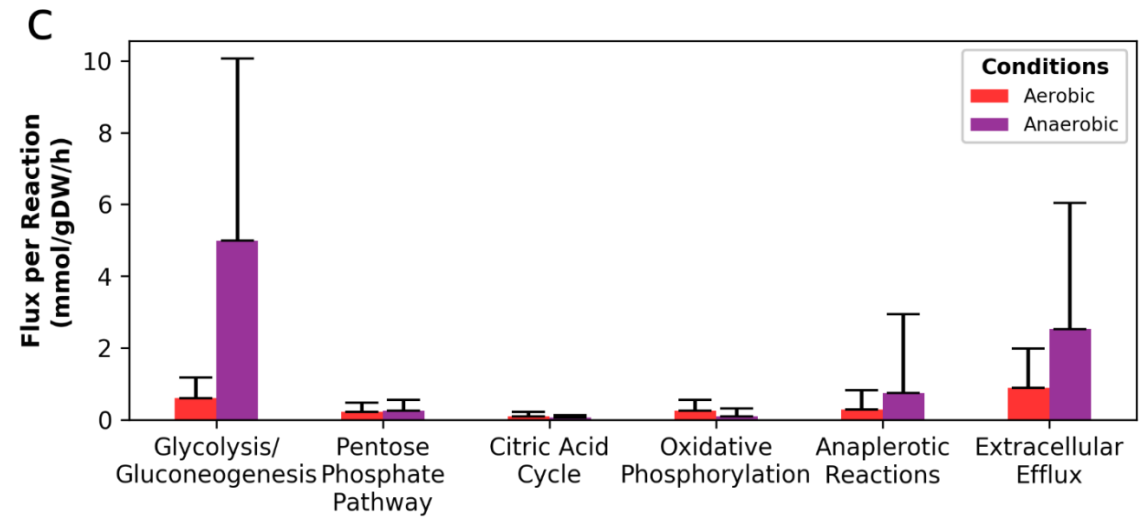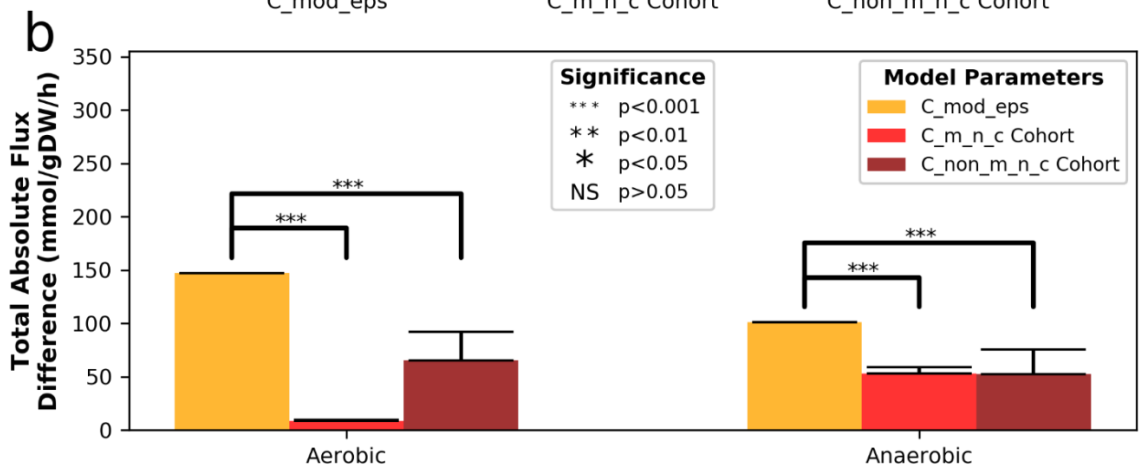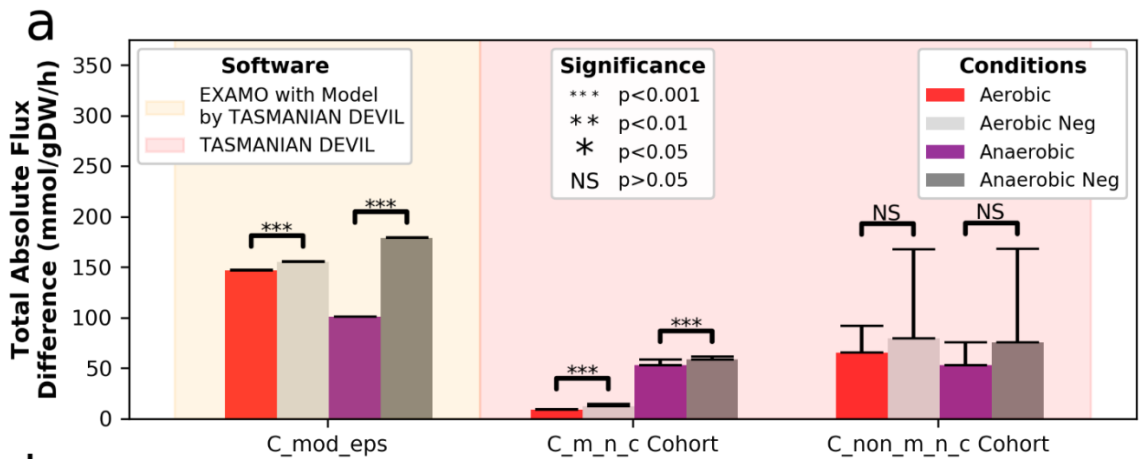
Figure 2.6.5   Flux maps of central carbon metabolism
Flux maps were designed in CellDesigner and overlaid with flux profiles for the (**a**) aerobic and
(**b**) anaerobic conditions for the C_m_n_c_lb_g parameter using the visualization module from
TASMANIAN DEVIL. If a flux or its standard deviation is followed by e_4, such as the flux
121109e_4 for ENO for the anaerobic condition, this translates to 12.1109, whereas e_4 is
scientific notation for E-4.
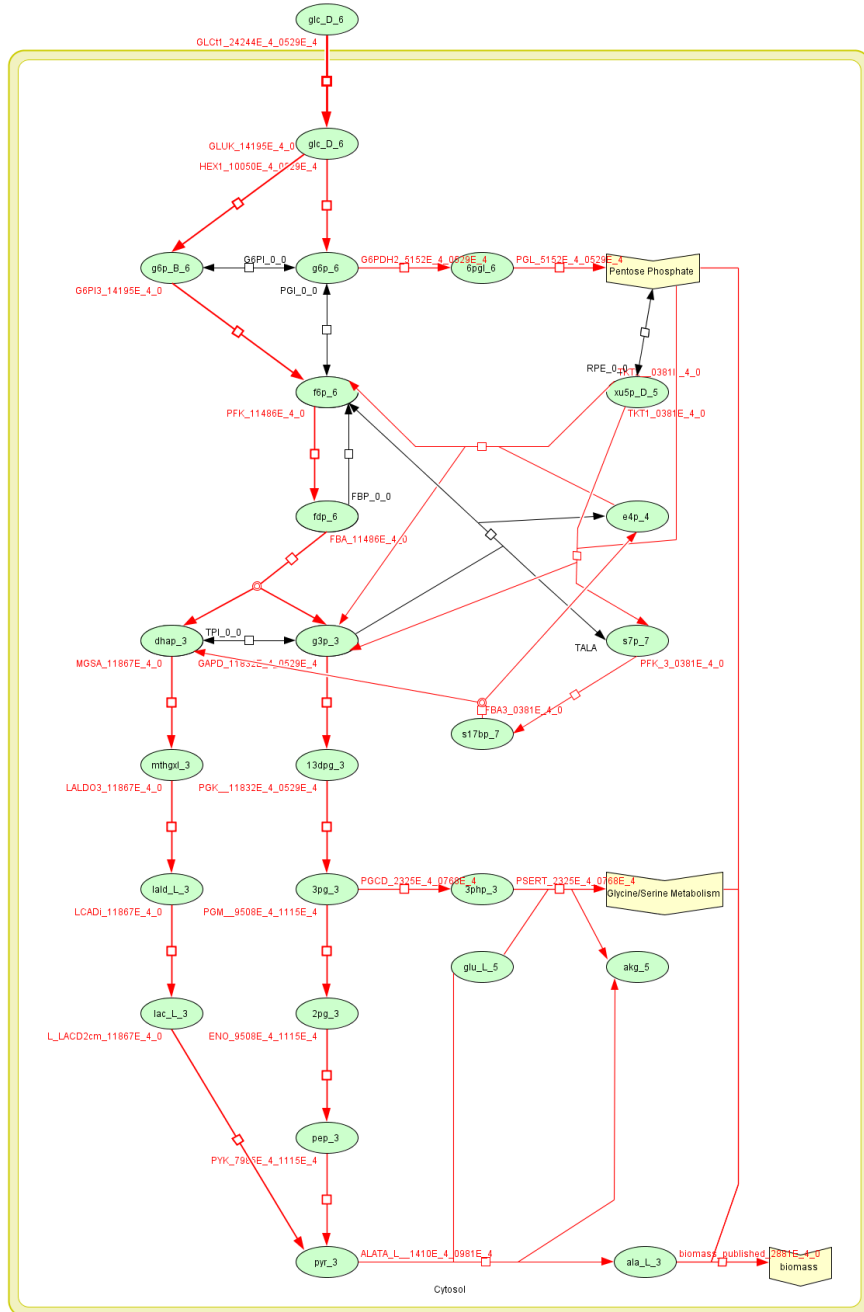
Rxn_AverageFlux_StandardDeviation
Flux>0: Width Scaled to Flux; Flux = 0: Inactive; Rxn not in Model: Pruned

b
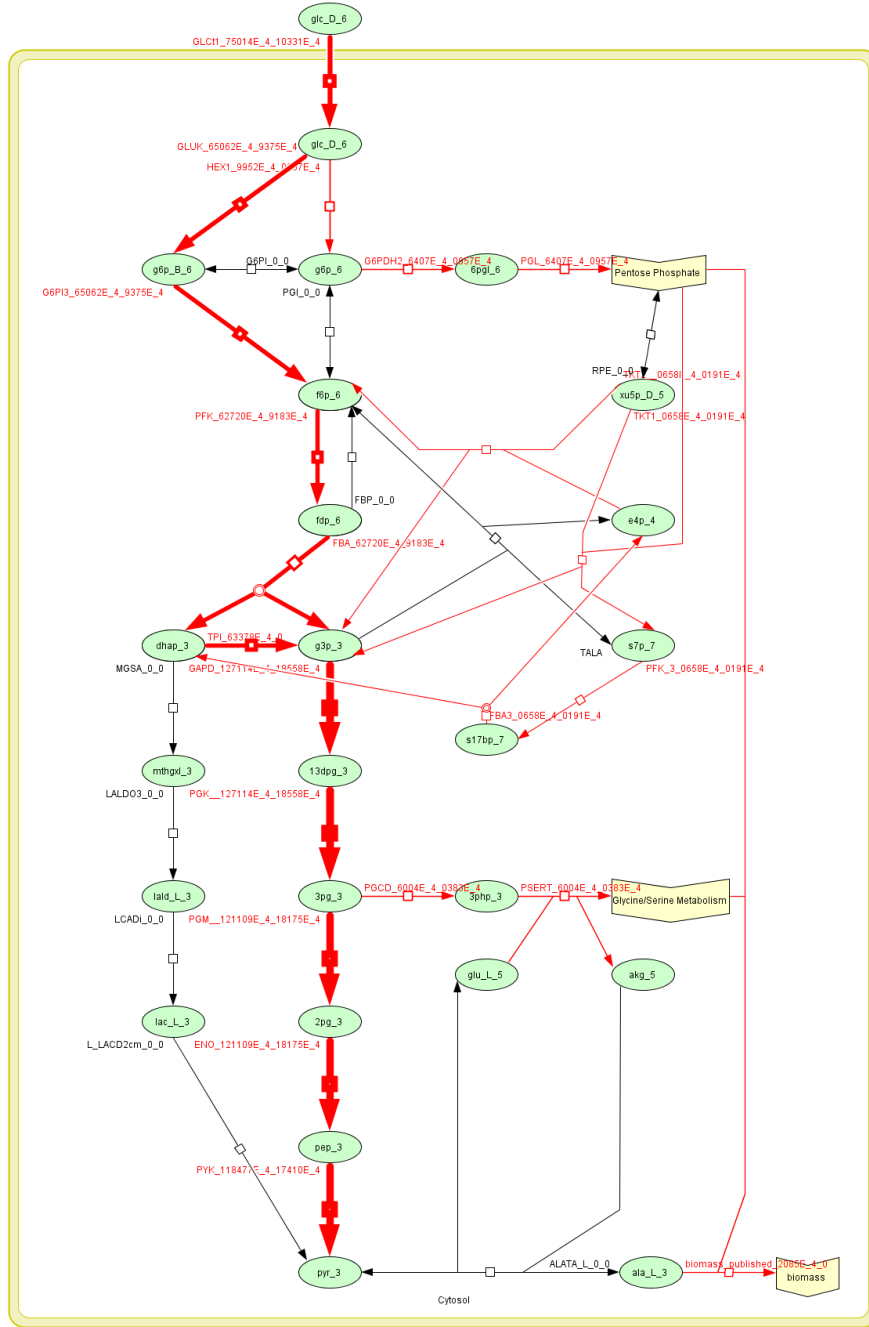
Rxn_AverageFlux_StandardDeviation
Flux>0: Width Scaled to Flux; Flux = 0: Inactive; Rxn not in Model: Pruned

21

Figure 2.6.6   Gene rule maps for central carbon metabolism
Pathway maps were designed in CellDesigner and overlaid with gene rules for the (**a**) aerobic and (**b**) anaerobic conditions for the C_m_n_c_lb_g parameter using the visualization module from TASMANIAN DEVIL. Genes are described by their Boolean logic, and the rule determination is color encoded.
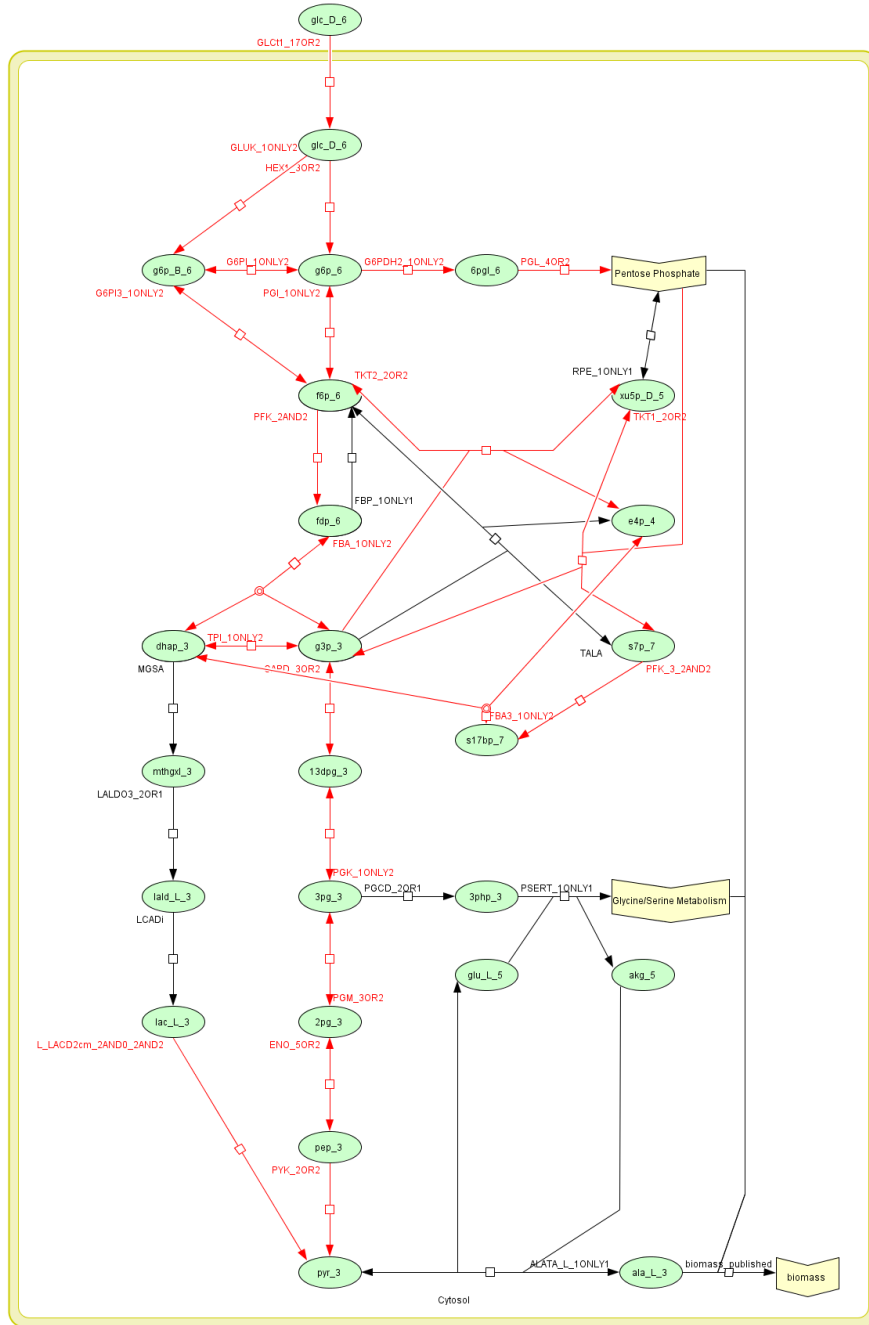
Rxn_NumberofGenes(AND/OR/ONLY STATEMENT)RuleOutcome
RuleOutcome=2: On; RuleOutcome=0: Off; RuleOutcome=1: Undefined

b
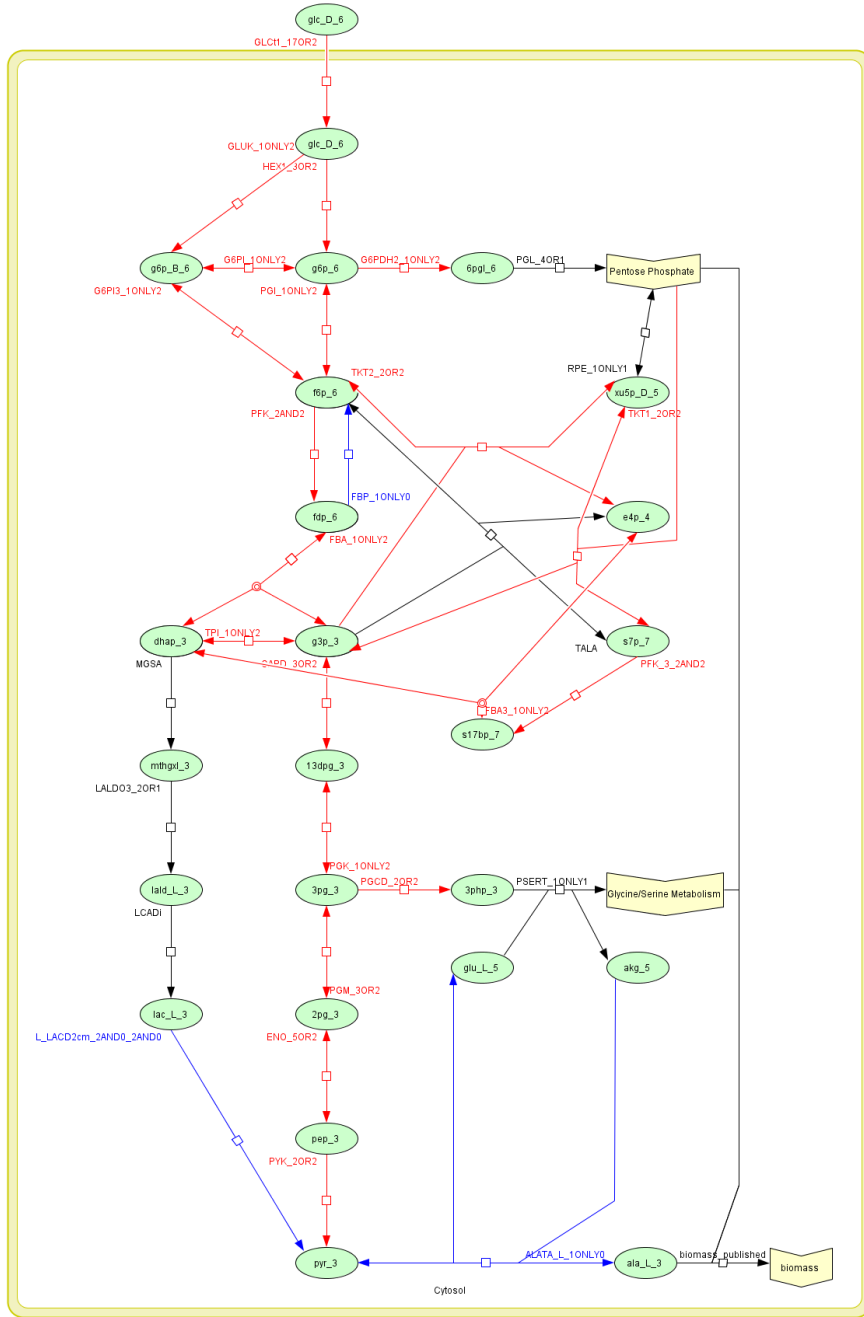
Rxn_NumberofGenes(AND/OR/ONLY STATEMENT)RuleOutcome
RuleOutcome=2: On; RuleOutcome=0: Off; RuleOutcome=1: Undefined

Figure 2.6.7   Reaction inclusion consistency maps for central carbon metabolism
Pathway maps were designed in CellDesigner and overlaid with flux variability analysis (FVA)
results and model building algorithm (MBA) profiles for the (**a**) aerobic and (**b**) anaerobic
conditions for the C_m_n_c_lb_g parameter using the visualization module from TASMANIAN
DEVIL.

b
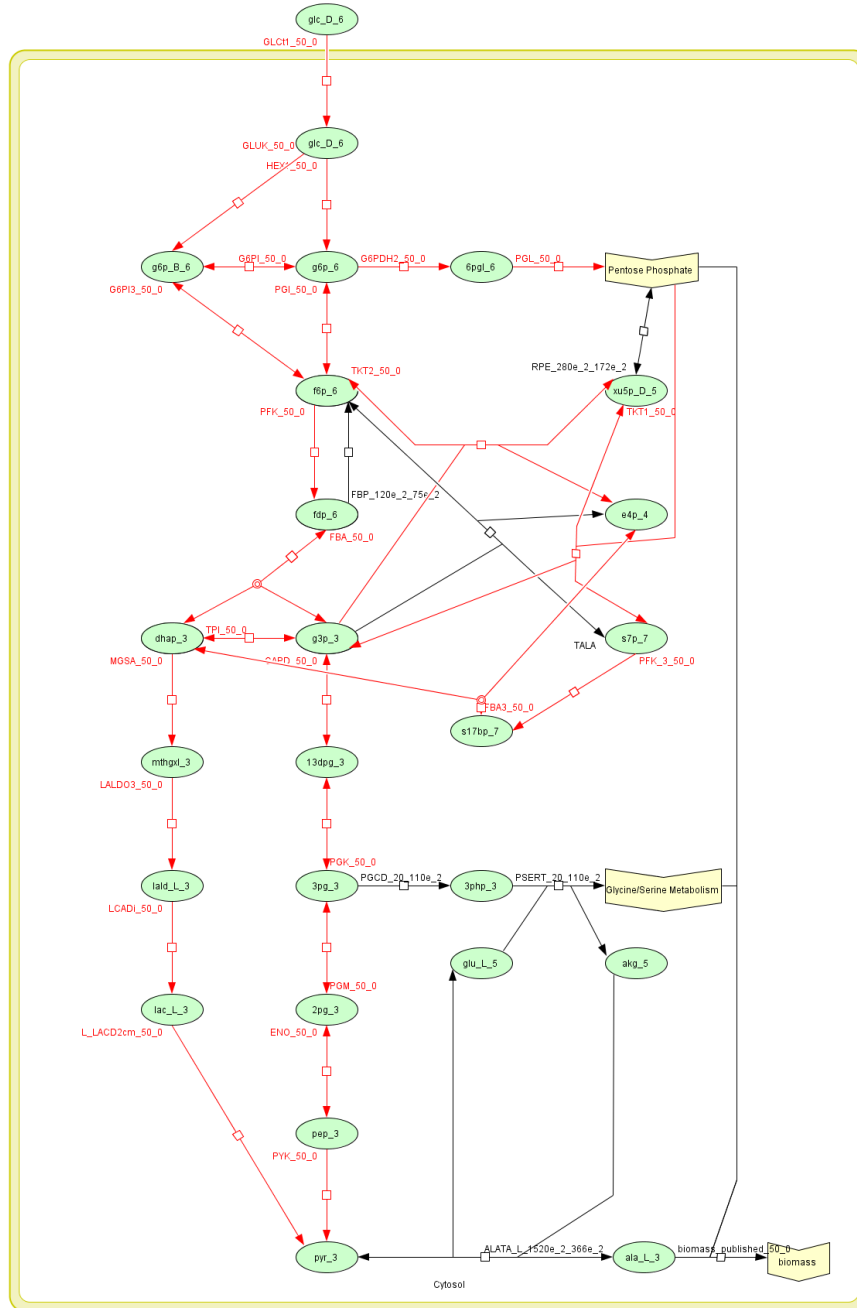
FVA Frequency for Rxn: HFR; FVA Frequency for Rxn: ZFR; FVA Frequency for Rxn: Undefined
Rxn_AverageNumberofTimesinMBAProfile_StandardDeviation

Table 2.6.1   Testing parameters for validation
The parameters used for testing TASMANIAN DEVIL, as well as the module in which the
optional parameter is available, the function of the parameter, and the logic behind the change.

Condition	Module	Function	Logic
_g	Model	Change the GPR association	Account for additional gene associations from literature
_lb	Model	Adjust the lower boundary constraints	Adjust model to account for extracellular media/uptake rates
_m_n_c	Model	Account for cofactors and nucleoside phosphate moieties, and make the model carbon balanced	Make flux model more carbon-centered
_Ex	Flux	Prune extracellular reactions first, then extracellular transport reactions, then compartmental transport reactions, then inner-compartmental transport reactions	More of an emphasis is placed on transport reaction gene activity

Table 2.6.2   Gene-protein-reaction associations for iMM904 modifications
Gene-protein-reaction changes that were made to the iMM904 metabolic model for validating
TASMANIAN DEVIL when the option to alter them was selected.

Reaction	Original Gene Mapping	Adapted Gene Mapping
ALAt2r	(YKR039W or YPL265W or YCL025C or YOL020W or YOR348C)	
ARGt2r	(YKR039W or YEL063C or YNL270C)	(YEL063C or YNL270C)
ASNt2r	(YKR039W or YCL025C or YDR508C or YPL265W)	YCL025C
ASPt2r	(YFL055W or YKR039W or YPL265W)	YPL265W
CYSt2r	(YKR039W or YDR508C or YBR068C or YDR046C or YBR069C or YOL020W)	YDR046C
GLNt2r	(YKR039W or YCL025C or YDR508C or YPL265W)	(YCL025C or YDR508C)
GLUt2r	(YFL055W or YDR536W or YKR039W or YCL025C or YPL265W)	YPL265W
GLYt2r	(YKR039W or YOL020W or YPL265W or YOR348C)	
HISt2r	(YGR191W or YKR039W or YCL025C or YBR069C)	YGR191W
ILEt2r	(YBR069C or YKR039W or YCL025C or YBR068C or YDR046C)	(YBR069C or YBR068C or YDR046C)
LEUt2r	(YBR069C or YKR039W or YCL025C or YBR068C or YDR046C or YDR508C)	(YBR069C or YBR068C or YDR046C)
LYSt2r	(YNL268W or YKR039W)	YNL268W
METt2r	(YKR039W or YCL025C or YDR508C or YBR068C or YDR046C or YGR055W or YHL036W)	(YGR055W or YHL036W)
PHEt2r	(YKR039W or YCL025C or YOL020W or YBR068C or YDR046C)	
PROt2r	(YKR039W or YOR348C)	YOR348C
SERt2r	(YFL055W or YCL025C or YDR508C or YKR039W or YPL265W)	
THRt2r	(YKR039W or YBR069C or YCL025C or YDR508C)	
TRPt2r	(YBR069C or YKR039W or YOL020W or YBR068C or YDR046C)	YOL020W
TYRt2r	(YBR069C or YKR039W or YCL025C or YBR068C or YOL020W or YDR046C or YCL025C)	(YBR069C or YOL020W)
VALt2r	(YKR039W or YCL025C or YDR046C or YBR069C or YBR068C)	(YDR046C or YBR069C or YBR068C)

Table 2.6.3   Lower boundary constraint modifications for iMM904 conditions
Lower boundary constraint alterations for extracellular uptake reactions that were modified for each condition.

Reaction	Original lb	YPD (Glucose) lb	YPEtoh (Ethanol) lb	Aerobic lb	Anaerobic lb
EX_4abz(e)	0	0	0	-0.0002	-0.0002
EX_ade(e)	0	-0.00767027	-0.00767027	0	0
EX_ala_L(e)	0	-0.00073	-0.00073	0	0
EX_arg_L(e)	0	-0.00041	-0.00041	0	0
EX_asp_L(e)	0	-0.00023	-0.00023	0	0
EX_chol(e)	0	-0.015	-0.015	0	0
EX_csn(e)	0	-0.002222703	-0.002222703	0	0
EX_ergst(e)	0	0	0	-0.01	-0.01
EX_etoh(e)	0	0	-20	0	0
EX_glc(e)	-10	-20	0	-10	-10
EX_glu_L(e)	0	-0.00056	-0.00056	0	0
EX_gly(e)	0	-0.0004	-0.0004	0	0
EX_gsn(e)	0	-0.032575676	-0.032575676	0	0
EX_his_L(e)	0	-0.00064	-0.00064	0	0
EX_ile_L(e)	0	-0.00074	-0.00074	0	0
EX_inost(e)	0	0	0	-0.025	-0.025
EX_leu_L(e)	0	-0.00102	-0.00102	0	0
EX_lys_L(e)	0	-0.0003	-0.0003	0	0
EX_met_L(e)	0	-0.00089	-0.00089	0	0
EX_nac(e)	0	-0.006	-0.006	0	0
EX_o2(e)	-2	-2	-2	-2	-0.02
EX_phe_L(e)	0	-0.00081	-0.00081	0	0
EX_pnto_R(e)	0	-0.00105	-0.00105	-0.001	-0.001
EX_pro_L(e)	0	-0.00039	-0.00039	0	0
EX_ribflv(e)	0	-0.00125	-0.00125	0	0
EX_ser_L(e)	0	-0.00079	-0.00079	0	0
EX_thm(e)	0	-0.00021	-0.00021	-0.001	-0.001
EX_thr_L(e)	0	-0.00062	-0.00062	0	0
EX_trp_L(e)	0	-0.01667	-0.01667	0	0
EX_tyr_L(e)	0	-0.00027	0.00027	0	0
EX_uri(e)	0	-0.005883784	-0.005883784	0	0
EX_val_L(e)	0	-0.00076	-0.00076	0	0
EX_xtsn(e)	0	-0.002222703	-0.002222703	0	0

Table 2.6.4   Simplifications for cofactors and nucleoside conversions for iMM904
The listed cofactors and their partners were removed from reactions if a reaction did not lead to
the cofactor's direct biosynthesis. In other words, if there were another product and reactant in the
reaction, the cofactors were removed from that particular reaction. The nucleosides listed in lists
were converted to their monophosphate versions.

Simplification	Mapping
Cofactors	fad_m ⟺ fadh2_m, nad_c ⟺ nadh_c, nad_m ⟺ nadh_m, nad_x ⟺ nadh_x, nadp_c ⟺ nadph_c, nadp_m ⟺ nadph_m, nadp_r ⟺ nadph_r
Nucleoside Conversions	amp_c = ['adp_c', 'atp_c', 'damp_c', 'dadp_c', 'datp_c'], amp_m = ['adp_m', 'atp_m'], amp_x = ['adp_x', 'atp_x'], gmp_c = ['gdp_c', 'gtp_c', 'dgmp_c', 'dgdp_c', 'dgtp_c'], cmp_c = ['cdp_c', 'ctp_c', 'dcmp_c', 'dcdp_c', 'dctp_c'], cmp_m= ['cdp_m', 'ctp_m'], dtmp_c= ['dtdp_c', 'dttp_c'], ump_c= ['udp_c', 'utp_c', 'dump_c', 'dudp_c', 'dutp_c']

Chapter 3 Condition-specific metabolic modeling to identify mechanisms of regulation: GCR1 autoregulation through the transcriptional activation of the glycine cleavage complex to initiate the diauxic shift, a case study

3.1 Abstract

3.1.1 Background

One goal in studying gene deletion experiments is to identify biological processes that are upregulated. By only examining the processes associated with differentially expressed genes through enrichment analysis, systematic mechanisms that activate these processes may not be revealed. Creating a condition-specific metabolic model based off of transcriptomic data can provide insights for better understanding the system of interest.

3.1.2 Results

Gene expression studies were contextualized using previously developed condition-specific metabolic modeling software to investigate the mechanisms by which GCR1, a transcriptional activator of glycolytic genes in *S. cerevisiae*, regulates global metabolism in glucose-rich and glucose-limited cultures. The modeling demonstrated that under glucose-rich conditions, Gcr1 promotes the transcription of enzymes controlling fermentation. When glucose becomes limited after a previous high growth rate, transcript levels for GCR1 and its transcriptional targets decrease and respiration begins. Gcr1 is deactivated by pyrophosphorylation, which is the addition of a phosphate group to a prephosphorylated serine residue through a nonenzymatic cleavage. The process is highly endergonic, requiring a near physiological level Michaelis constant for ATP. The pyrophosphorylation of Gcr1 causes it to disassociate from its transcription factor binding partner, Gcr1, causing Gcr1 to lose its glycolytic transcriptional activation. The modeling predicted that the upregulation of the glycine cleavage complex (GCV1, GCV2, and GCV3 in the mitochondria) by Gcr1 creates an efficient way to generate one carbon building blocks for biosynthesis for adenine production along with other purines. This predicts that Gcr1 is therefore a regulator of its own deactivation.

3.1.3 Conclusions

The modeling predicted that Gcr1 is an energy sensor and regulates its own activity based off of glucose availability and previous growth rate, thus initiating the diauxic shift through its own deactivation after a previous state of high growth. The approach used in this study can be used by other experimenters to provide insights into regulatory mechanisms in other organisms.

3.2   Background

3.2.1   Metabolic control by Gcr1

Gcr1 is a glucose response transcription factor in *S. cerevisiae* that has global effects on transcription, affecting a wide variety of genes [76]. Upon deletion of GCR1 (GCR1D) with glucose in the media, metabolism changes, whereby there is a decrease in glycolytic fermentation and an increase in the TCA cycle and oxidative phosphorylation [77, 78]. There is also an increase in carbon storage, such as trehalose. As glucose becomes exhausted and before the wild-type (WT) cells enters quiescence, they enter the diauxic shift, whereby they transition to a state similar to GCR1D [2, 6]. Previously excreted ethanol is imported back into the system, and the cells undergo respiration and gluconeogenesis [2, 78, 79].

3.2.2   Gcr1 isoforms

Gcr1 has two protein isoforms. One is expressed under glucose induction in the exponential growth phase (785 amino acids and 89 kilodaltons (kDa)), and the other is continually expressed except during quiescence, even during the diauxic, albeit to a lesser degree than during exponential growth (847 amino acids and 95 kDa) [6, 79]. Upon glucose exhaustion, the first isoform's RNA is targeted for nonsense-mediated decay by Xrn1 in P-bodies and stress granules [79, 80]. The first isoform is formed by splicing of the first and second exon, with the first exon being very small [79]. There is a large intron between the first and second exon (739 nucleotides), which is inefficiently spliced [81]. The second isoform's start site is encoded within the annotated intron, producing a protein that is 62 amino acids longer than the canonical Gcr1 isoform [6].

3.2.3   Transcriptional coactivators and regulation of Gcr1

The two isoforms of Gcr1 associate with different transcription factors to promote different pathways. The canonical isoform of Gcr1 can create a heterodimer with Gcr2 [4, 82]. It also associates with Rap1 to upregulate ribosomal and glycolytic gene transcription. When target genes have both Gcr1 and Rap1 consensus sequences in their promotes, transcription of these genes are highly upregulated when Gcr1 is active [83]. Gcr1 can be pyrophosphorylated, and this causes the canonical Gcr1 isoform to disassociate from Gcr2 [82]. The expression of the canonical Gcr1 isoform is dependent on the Gcr1-Gcr2 heterodimer [84]. Pyrophosphorylation is the process by which a phosphate group is added to a prephosphorylated serine residue through a nonenzymatic cleavage of phosphate from 5-diphosphoinositol pentakisphosphate (5PP-IP5) [85]. The formation of 5PP-IP5 by the inositol hexakisphosphate kinase, KCS1, is a highly endergonic reaction, requiring a near physiological level Michaelis constant for ATP [86]. This indicates that pyrophosphorylation represses canonical GCR1 transcription, thereby downregulating glycolytic flux. Hexakisphosphate kinase (KCS1) mutants, which are incapable of generating pyrophosphates, are unable to make the diauxic shift from aerobic glycolysis to oxidative phosphorylation, thereby producing more ethanol [82].

In comparison, the isoform produced from the intron has not been studied in-depth. However, it has been predicted to be associated with Skn7, Adr1, Sko1, Hot1, and Rtg2 based on transcriptional profiles [79]. The region of the protein encoded by intronic nucleotides contains a coiled-coil motif, which contains binding sites for transcription factor partners. These are in addition to the coiled-coil motif present within the second exon which present the typical binding site for Rap1 [83, 87]. Finally, it has been shown that both isoforms of Gcr1 can form homodimers with themselves and heterodimers with each other, creating a diverse capability of Gcr1 to correspond to cellular needs [6].

3.2.4   Inspiration for study

Most metabolic studies of GCR1 have focused on the canonical GCR1 isoform in exponential growth phase with excess glucose. There has not been an in-depth study of the

systematic changes to metabolism upon the initiation of the diauxic shift. Previous metabolic studies on GCR1 have also not emphasized side pathways that are affected, such as serine, glycine, and one carbon metabolism, or how these pathways work in tandem to promote growth and regulate proliferation. This study will model the metabolic difference between exponential growth with glucose, driven by the canonical Gcr1 isoform, contrasted with the beginning of the diauxic shift, when the canonical Gcr1 isoform is downregulated and the intronic version is still expressed. The TASMANIAN DEVIL software package, which has been previously validated in *S. cerevisiae,* will be utilized to predict the condition-specific metabolism.

3.3   Methods

3.3.1   Culture conditions and transcriptomic analyses

To determine the primary metabolic roles of Gcr1 in *S. cerevisiae,* transcriptomic data was compared during the exponential growth phase with excess glucose and during the diauxic shift. For exponential growth, the laboratory of Henry Baker cultured S150-2B (*MATa leu2-3,112 his3Δ trp1-289 ura3-52*) and isogenic GCR1 and GCR2 null mutants in YP media with 2% lactate and 2% glycerol. 2% glucose was added, and the cells were harvested 4 hours later in the log-growth phase at an optical density (OD) of 600nm equal to 1. RNA was isolated and reverse transcribed, as previously described [77]. Microarray analyses were performed on Affymetrix GeneChips and processed using MAS5 normalization for biological triplicates. MAS5 scales data globally by fitting a linear regression line to a trimmed data set for the top and bottom 1% of expressed genes. The algorithm then subtracts each value by the intercept of the regression line and then divides that value by the slope [88]. The methods were identical to a study published by Baker, but the samples were run on a newer microarray chip [77]. The laboratory of Tracy Johnson cultured BY4742 (*MATa his3Δ1 lue2Δ0 lys2Δ0 ura3Δ0)* and the isogenic GCR1 null mutant in YP media with 2% glucose. RNA was isolated post log-growth phase around the diauxic shift at an OD of 600nm near 2.8 [6]. Samples were run on the SOLiD platform at the Sanford Consortium for Regenerative Medicine. Samples were aligned using BFAST and transcripts were analyzed using Cufflinks as previously described [89–91].

3.3.2   Differential expression analyses and gene rule classification

Statistically differentially expressed genes were used to define the gene activity rules for TASMANIAN DEVIL. Bioconductor in R was used to determine differential expression and pathway enrichment [92]. The analysis of Baker's microarray data employed Limma, which is an Empirical Bayes approximation for normally distributed data [93]. edgeR was used to determine differentially expressed genes using a Fisher's exact test for the negative binomial distributed RNA-Seq samples from Johnson's study [94, 95]. The Benjamini-Hochberg correction was used for both methods to adjust for false discovery rates [96]. GOstats was used to test Gene Ontology pathway enrichment for functional inference and as a way to check the results of the metabolic modeling [97, 98]. For gene rules, genes that were very lowly expressed (less than 10 fragments of kilobase of exon per million fragments (FPKM) when all samples were added together for both conditions), were assigned to be inactive, regardless of differential expression. Genes that were differentially expressed for a condition were defined to be active. Genes that were differentially expressed in the opposing condition for an experimenter and were not in the top 20% of highly expressed genes were defined to be inactive. All other genes were undefined in the gene rules.

3.3.3   Metabolic modeling

3.3.3.1   Adjustments to the modeling parameters

The same starting model was used for *S. cerevisiae* which produced the lowest combined error compared to the actual fluxes in the TASMANIAN DEVIL study. GPR associations for extracellular uptake were changed, and modifications were made to account for cofactors, nucleoside phosphate moieties, and making the model carbon balanced. The lower boundary

constraints were adjusted according to **Table 3.6.1**, which were adapted from the YPD lower boundary constraints used in the TASMANIAN DEVIL study. When glucose was induced in the media, glycerol and lactate import were repressed [77]. Ethanol was allowed to be imported for Johnson since it was beginning the diauxic shift. One other change that was implemented to the GPR associations was that the gene for alcohol dehydrogenase 1 (ADH1) was also assigned to ethanol transport (ETOHt). There is no gene directly responsible for the import/export of ethanol. This change was made such that ethanol could be exported and the transport reaction would not be pruned if ADH1 was active. For Baker, the upper boundary constraints of the extracellular transport reactions were limited to 0.2 mmol/gDW/h for every reaction except ethanol export to ensure that ethanol export was used to carbon balance the model instead of other excreted metabolites.

3.3.3.2 *Adjustments to the gene rules*

Some minor modifications to the gene rules were necessary to insure that metabolic changes were observed for pathways with the greatest enrichment. Since the gene rules for the glycine cleavage complex consists of 4 genes and the model required that all 4 genes must be active, and the same 3 genes were differentially expressed in both Baker and Johnson's data for when GCR1 was active (GCR1A), the fourth gene that was not differentially expressed (LPD1) was forced to be active. LPD1 is a lipoamide reductase that also functions as a disulfide oxidoreductase, priming the glycine cleavage system to add a carbon from glycine to tetrahydrofolate to form methylenetetrahydrofolate [99]. This change allowed for flux to go through the glycine cleavage system. The other modification to the gene rules was for the minor arginine transporter, ANP1. It was modified from being defined as active to undefined for Johnson for GCR1A. This was because upon testing, it was identified that if it were active, this would mean the flux through arginine biosynthesis and the urea cycle would not occur and arginine could instead be imported to reduce the sum of fluxes in the network. Arginine biosynthesis and the urea cycle were two of the most differentially expressed pathways for GCR1A Johnson, as shown in **Table 3.6.2**.

3.3.3.1 *Additional TASMANIAN DEVIL flux module options and adjustments to the growth rate*

50 iterations were run with 5 replicates for each condition for each experimenter in TASMANIAN DEVIL to account for variability between pruned models. The biomass production rate for GCR1A Baker was set to be equal to the rate measured by Zampar et al. for *S. cerevisiae* cultured with excess glucose (approximately 0.305 mmol/gDw/h) [2]. For GCR1A Johnson, glucose was modeled as being exhausted and ethanol was the primary carbon source. Biomass production was set to be equal to the rate measured by Zampar et al. under similar conditions at the beginning of the diauxic shift (0.06 mmol/gDW/h) [2].

3.4 Results

3.4.1 Gcr1 promotes glycolysis and biosynthesis prior to switching to oxidative phosphorylation after the diauxic shift

First, it was investigated what the differentially expressed genes indicated based off of pathway enrichment using Gene Ontology pathway analysis. GCR1A Baker showed enrichment for alcohol metabolism and glycolytic fermentation, while GCR1A Johnson was enriched for glycolytic fermentation, glycolysis, as well as gluconeogenesis, as highlighted in **Table 3.6.2**. Both GCR1A Baker and GCR1A Johnson showed increased serine, glycine, and one carbon biosynthesis gene expression compared to the GCR1 null mutants. GCR1A Johnson was enriched for the regulation of cellular respiration as well as mitochondrial respiratory chain complexes II-IV. GCR1A Johnson was also upregulated for arginine and urea metabolism, related to the nitrogen cycle. Finally, GCR1A Baker was enriched for phosphatidylinositol biosynthesis, while GCR1A Johnson was enriched for inositol biosynthesis.

Next, it was determined whether the modeled fluxes from the differentially expressed gene rules resembled the statistically enriched metabolic pathways. It was predicted that in general, GCR1A Baker had more flux through glycolysis and the pentose phosphate pathway than for GCR1A Johnson, whereas GCR1A Johnson had more flux through the citric acid cycle and oxidative phosphorylation (**Figures 3.6.1a**). Since it was predicted that GCR1A Baker did not have as much oxidative phosphorylation occurring, this was contrasted by an increase in anaplerosis to fuel biosynthesis. GCR1A Baker also had increased biosynthesis in general, which was demonstrated by a greater flux through glycine and serine metabolism and purine and pyrimidine biosynthesis. GCR1A Johnson also had increased arginine and proline metabolic flux, which matched the enriched pathway analysis.

Individual fluxes were analyzed to determine other key differences between the models. Glucose was the primary carbon source for GCR1A Baker and ethanol was the primary metabolite excreted to manage carbon balancing (**Figure 3.6.1b**). In contrast, GCR1A Johnson consumed ethanol, and it was predicted that urea was used as a carbon sink to balance the model through arginine metabolism. GCR1A Baker utilized glycolysis to advance biosynthesis, as observed by the enolase flux, and pyruvate carboxylase was used to convert pyruvate into oxaloacetate, diverting flux into the citric acid cycle for anaplerosis. GCR1A Johnson relied on gluconeogenesis, as demonstrated by the reversed enolase flux. This was predicted to be produced primarily from oxaloacetate being converted into phosphoenolpyruvate by phosphoenolpyruvate carboxykinase. GCR1A Jonson had greater mitochondrial oxidative phosphorylation, as shown by the mitochondrial fumarase flux. **Figure 3.6.1b** also demonstrates the differences in the modeled biomass flux assumptions, as described in the methods.

3.4.2   The glycine cleavage complex promotes one carbon biosynthesis

Next, it was analyzed how Gcr1 upregulated biosynthesis. The glycine cleavage complex genes (GCV1, GCV2, and GCV3) were all differentially expressed for GCR1A Baker and GCR1A Johnson compared to GCR1D (**Figure 3.6.2a** and **Figure 3.6.2b**). When GCR2 was deleted (GCR2D) for Baker, the reductions in expression of GCV1, GCV2, and GCV3 were almost identical to GCR1D Baker compared to when GCR2 was active (GCR2A), as shown in **Figure 3.6.2c**. The glycine cleavage complex (GLYCLm) transfers a carbon from glycine to 5,10-methylenetetrahydrofolate in the mitochondria. **Figure 3.6.2d** demonstrates how GLYCLm functions as a system of four genes: GCV1, GCV2, GCV3, and LPD1. A carbon is cleaved off glycine by Gcv2 through decarboxylation, producing carbon dioxide as a product. The decarboxylated moiety of glycine becomes attached to the lipoate attached to Gcv3. Gcv2 then degrades the moiety further, and a methylene is donated to tetrahydrofolate to generate 5,10-methylenetetrahydrafolate. Lpd1 then reoxidizes the lipoate attached to Gcv3 by nicotinamide adenine dinucleotide [100, 101].

The glycine cleavage complex provides an efficient way to maximize one carbon production from glycine. For general serine metabolism, two of the carbons of serine are converted into glycine, while a methylene is donated to tetrahydrofolate to generate 5,10-methylenetetrahydrafolate. This methylene can eventually go into one carbon biosynthesis for the generation of purines. Glycine itself is also used as a two carbon building block for the generation of purines. **Figure 3.6.2e** shows that GCR1A Baker had greater flux through these pathways than GCR1A Johnson as demonstrated in **Figure 3.6.2f**, but both relied on the glycine cleavage complex to maximize one carbon production. The modeling predicted that anaplerosis involving threonine produced most of the glycine for GCR1A Baker, whereas serine produced from gluconeogenesis generated most of the glycine for GCR1A Johnson.

3.4.3   GCR1 expression is downregulated upon the diauxic shift

GCR1 had indeed been successfully repressed in GCR1D Baker and was statistically different than GCR1A Baker ($p \leq 0.01$, using an Empirical Bayes approximation), as shown in

Figure 3.6.3a. However, GCR1 was not differentially expressed for Johnson, even though no transcripts of GCR1 were detected by RNA-Seq for GCR1D Johnson, as demonstrated in **Figure 3.6.3b**. However, this is only for the canonical annotated isoform of Gcr1, and not the isoform with the start site beginning from the intron that remains expressed during the diauxic shift [6, 79]. This suggested that the canonical GCR1's transcription was being turned off in GCR1A Johnson as the diauxic shift began.

3.5 Discussion

3.5.1 Analysis of modeling results and limitations

The modeled fluxes embodied the phenotypes portrayed by the statistical enrichments of the differentially expressed genes. GCR1A Baker was upregulated for glycolytic fermentation and one-carbon metabolism, and GCR1A Johnson was upregulated for gluconeogenesis, one carbon metabolism, oxidative phosphorylation, and arginine and urea metabolism. This is in general agreement with the pathways upregulated before and after the diauxic shift [2]. GCR1A Baker had increased biosynthesis fueled by glycolysis to promote the greater biosynthetic state. GCR1A Johnson exhibited oxidative phosphorylation and gluconeogenesis from ethanol import, which previously had been exported during exponential growth. Ethanol was predicted to be the carbon sink for glucose for GCR1A Baker, and urea was predicted to be the carbon sink for ethanol for GCR1A Johnson. Follow up experiments should be conducted to confirm if urea is indeed produced from arginine during the diauxic shift, or if perhaps another TCA cycle byproduct is excreted due to ethanol import.

While the studies by Baker and Johnson were carried out upon glucose induction before and after the diauxic shift, the direct effect on metabolism by each isoform of GCR1 remains unclear. Mutants in both studies were lacking both isoforms. While the canonical Gcr1 protein expression has been shown to upregulate glycolysis and fermentation, it is not clear what the role of the second isoform is. While putative binding partners include Skn7, Adr1, Rtg2, Sko1, and Hot1, it has not been verified which if any of these precipitate with the second isoform of Gcr1 [79]. However, there is evidence that these partners would promote respiration fueled by ethanol import. Adr1 promotes the transcription of ADH2, which converts ethanol to acetaldehyde [102]. Skn7 is responsible for the transcriptional activation of oxidative response genes, while Rtg2 is a sensor of mitochondrial dysfunction [103, 104]. More experiments should be conducted to determine the specific role of the second isoform of Gcr2.

3.5.2 The regulation of GCR1 expression influences its metabolic activity

The protein levels of Gcr1 have been shown to decrease rapidly during the diauxic shift around an OD of 2.8, which was when the samples were collected for GCR1A Johnson [5, 6]. It had been shown by Szigyarto et al. that Gcr1 can be pyrophosphorylated, causing Gcr1 to disassociate from it transcriptional binding partner, GCR2 [82]. Sasaki et al. proposed that the transcriptional activation of GCR1 depends on the Gcr1-Gcr2 heterodimer, given that GCR2 null mutants repressed GCR1 expression [84]. Evidence for the upregulation of pyrophosphorylation would indicate that the canonical isoform of GCR1 is being downregulated.

Pyrophosphates are produced by Kcs1, and KCS1 mutants are unable to make the diauxic shift from glycolysis to oxidative phosphorylation [82]. The rate-limiting enzyme of inositol phosphate biosynthesis is Ino1, the first reaction in inositol metabolism branching from glucose-6-phosphate from glycolysis to produce myo-inositol-1-phosphate [105]. INO1 expression is almost completely eliminated in a KCS1 null mutant, and it had been proposed that inositol pyrophosphates themselves activate expression of INO1. Therefore, a significant overexpression of INO1 signifies a state in which inositol pyrophosphates have been produced [106]. Another mechanism by which INO1 is regulated is from the amount of inositol present within the cell. When there is an excess of inositol, INO1 is repressed, as Opi1 localizes to the nucleus to repress

INO1's transcriptional activator, INO2. INO1 is upregulated when there is only limited inositol, as Opi1 is tethered to the nuclear membrane [107]. INO1's expression is also upregulated when the pH becomes more acidic, as the bond between Opi1 and phosphatidic acid is weakened by protons, resulting in the delocalization of Opi1 from the plasma membrane to the nucleus [108]. If there is limited glucose, such as for GCR1A Johnson, INO1 is transiently upregulated, driving inositol generation and other downstream targets in the inositol phosphate pathway. GCR1A Johnson's gene expression levels show a decrease in canonical GCR1 expression and an increase in INO1 expression, compared to GCR1A Baker, indicating that GCR1 is being down-regulated by inositol pyrophosphates, as shown in **Figure 3.6.3a** and **Figure 3.6.3b**. Inositol biosynthesis is also upregulated in GCR1A Johnson, as shown in **Table 3.6.2.** This portrays that the system is indeed in the diauxic shift and has switched from glycolysis to oxidative phosphorylation, which is in agreement with the modeling results.

Kcs1's homologues in humans have an extremely high K_m for ATP (1000-1700 μM), and they have therefore been proposed to be 'energostats', requiring a high amount of ATP to catalyze the formation of 5PP-IP5 [86]. KCS1 mutants had a much higher ATP concentration and lower ADP and AMP concentrations compared to WT [82]. If the ATP barrier for KCS1 catalysis is overcome though, the reaction would lower the ATP concentration, thereby rebalancing the ATP/ADP and ATP/AMP ratios [86]. The similar transcriptional profiles of GCV1, GCV2, and GCV3 under the deletion of GCR1 and GCR2 could indicate that Gcr1 and Gcr2 act as a heterodimer to activate the expression of these genes involved in the glycine cleavage complex, and that the loss of either GCR1 or GCR2 would constitute reduced one carbon biosynthesis and the adenosine phosphate pool through the inability to efficiently catabolize glycine. It is proposed that Ino1 is a nutrient sensor and has a direct effect on the energy sensor, Gcr1. If there is enough ATP to generate 5PP-IP5, the canonical isoform of Gcr1 is pyrophosphorylated, thereby disassociating it from Gcr2 and downregulating its transcriptional activation, thus beginning the diauxic shift. **Figure 3.6.4** demonstrates an overview of this mechanism in relation to the modeled metabolic fluxes.

3.5 Conclusions

In this study, metabolism upon glucose induction has been modeled before and after the diauxic shift in *S. cerevisiae*. Gene knockout studies of GCR1 were conducted, and differentially expressed genes were used to characterize gene rules and to determine enriched pathways. Gcr1 promoted glycolytic biosynthesis and fermentation during exponential growth. Upon glucose exhaustion, the diauxic shift began, and the canonical isoform of Gcr1 was downregulated. The glycine cleavage complex was identified as being upregulated by GCR1, and it played an important role in producing one carbon building blocks and adenosine phosphates. It is proposed that the glycine cleavage complex helps form enough ATP from efficient carbon utilization of glycine to later prime GCR1 for deactivation upon pyrophosphorylation by Kcs1. The approach in this study can be used to model other metabolic factors in *S. cerevisiae* and other systems.

3.6 Figures and tables

Figure 3.6.1 Modeled fluxes of GCR1 conditions

(**a**) Pathway fluxes per the number of unpruned reactions remaining in the pathways. (**b**) Average fluxes for key reactions. Abbreviations are as follow: EX_glc(e): D-glucose exchange, EX_etoh(e): ethanol exchange, EX_urea(e): urea exchange, ENO: enolase, PC: pyruvate carboxylase, PPCK, phosphoenolpyruvate carboxykinase, and FUMm: fumarase mitochondrial.
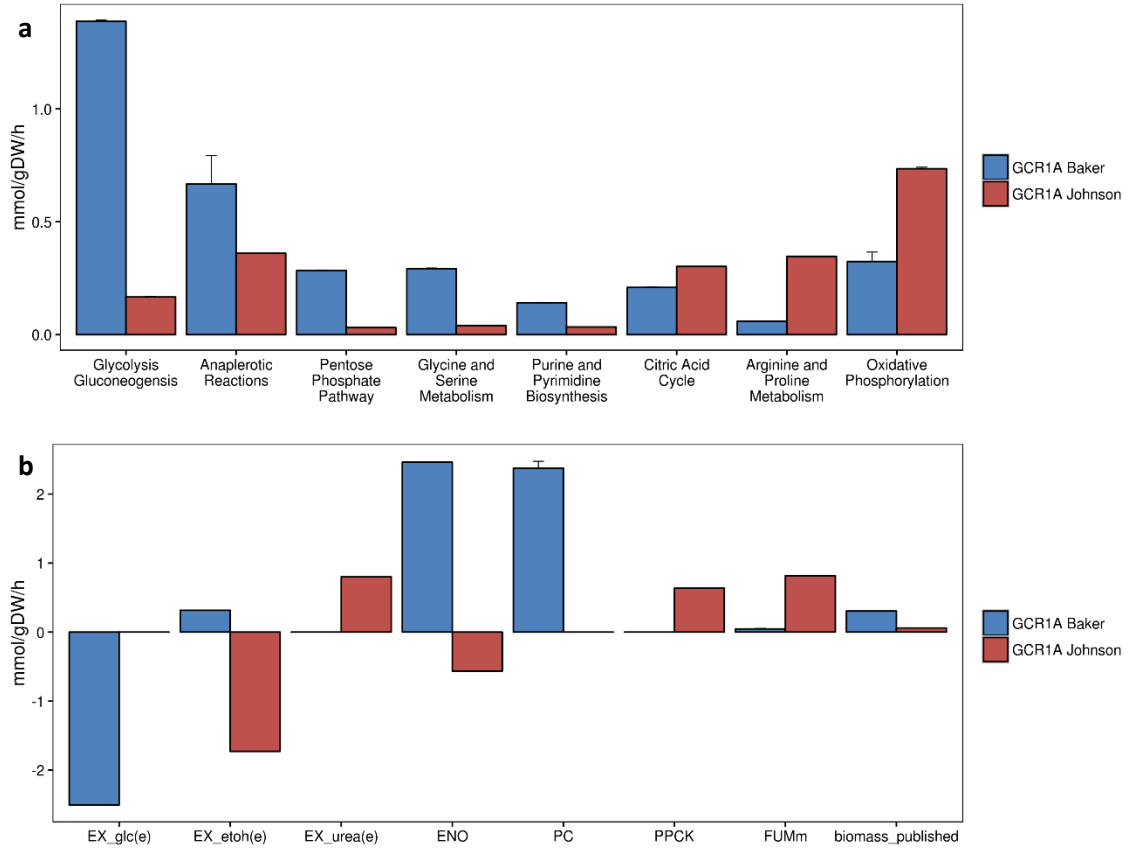
Figure 3.6.2   The role of the glycine cleavage complex
Transcriptomic profiles for genes mapped to the glycine cleavage system mitochondrial
(GLYCLm) reaction for (**a**) mutant GCR1 for Baker, (**b**) mutant GCR1 for Johnson, and mutant
GCR2 for Baker. (**d**) Schematic of the modeled glycine cleavage complex mechanism with
proteins scaled to their metabolic weights. (**e**) Glycine and serine metabolism for GCR1A Baker.
(**f**) Glycine and serine metabolism for GCR1A Johnson. The symbols * and ** denote $p \leq 0.01$
and $p \leq 0.001$, respectively. For Baker's data, an Empirical Bayes approximation was employed.
For Johnson's data, a Fisher's exact test was used.
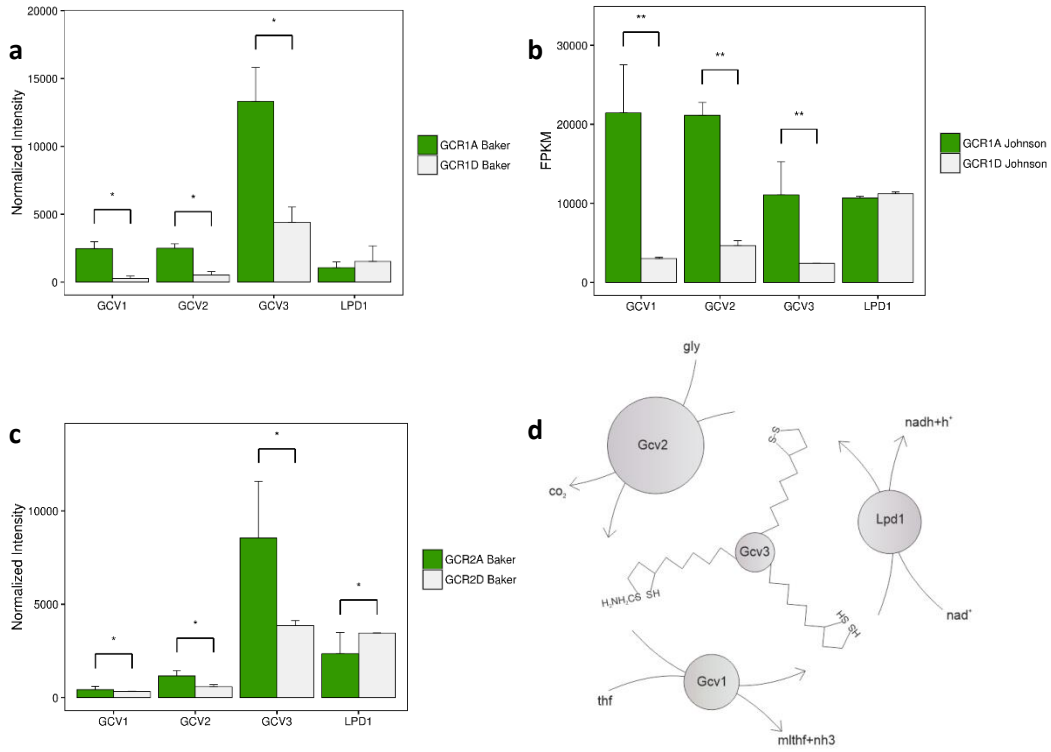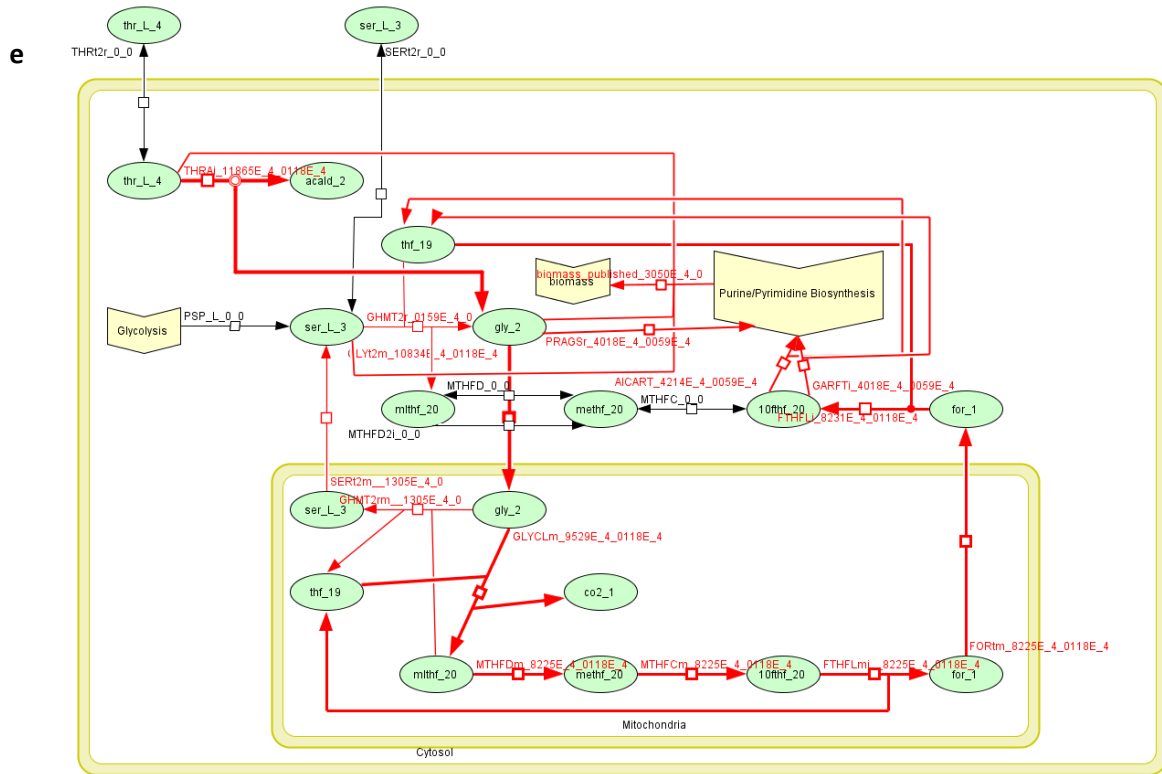
Rxn_AverageFlux_StandardDeviation
Flux>0: Width Scaled to Flux; Flux = 0: Inactive; Rxn not in Model: Pruned
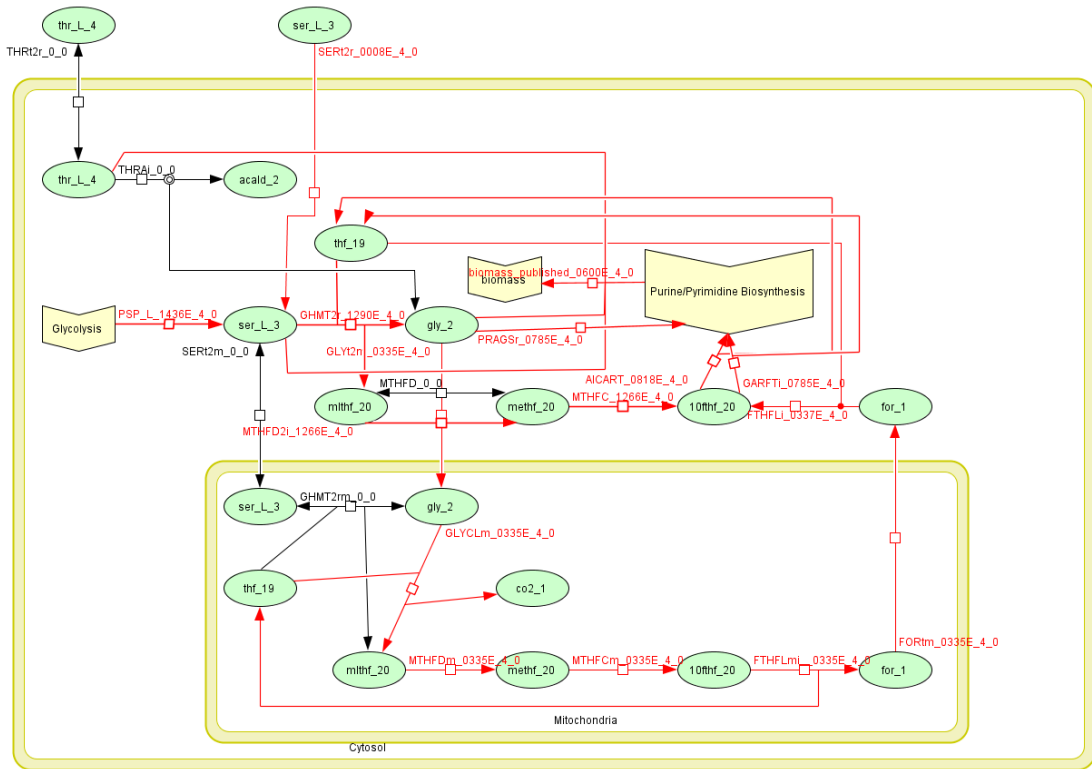
Rxn_AverageFlux_StandardDeviation
Flux>0: Width Scaled to Flux; Flux = 0: Inactive; Rxn not in Model: Pruned

Figure 3.6.3   Expression profiles for inositol regulation of GCR1
Rate limiting inositol biosynthesis enzyme and GCR1 expression for the experiments by (**a**)
Baker and (**b**) Johnson. The symbols * and ** denote $p \leq 0.01$ and $p \leq 0.001$, respectively. For
Baker's data, an Empirical Bayes approximation was employed. For Johnson's data, a Fisher's
exact test was used.
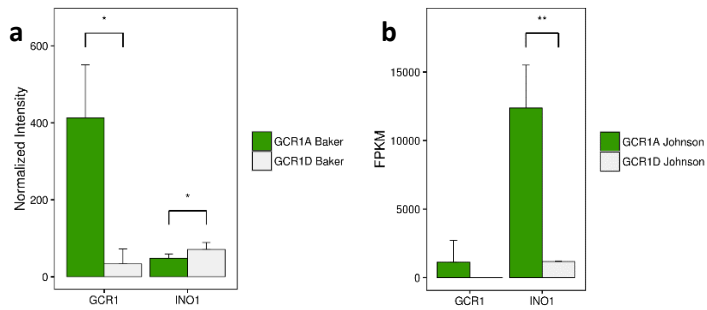
Figure 3.6.4   General overview of the regulation of the canonical isoform of Gcr1
Overview of mechanisms controlled by and controlling the canonical isoform of Gcr1 in relation
to its metabolic function (**a**) before and (**b**) after the initiation of the diauxic shift. The fluxes are
scaled to those modeled for GCR1A Baker and GCR1A Johnson. OD and AU stand for optical
density and absorbance units.
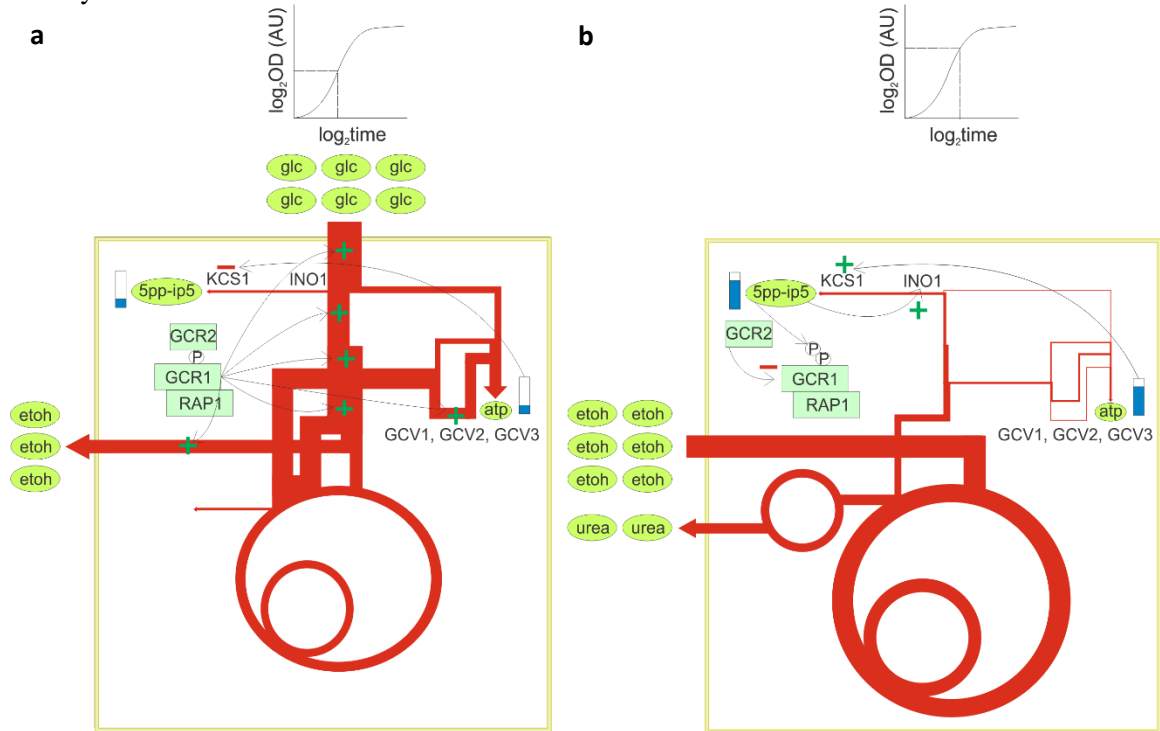
Table 3.6.1   Lower boundary constraint adjustments for modeling GCR1 conditions
Lower boundary constraint alterations for extracellular uptake reactions that were modified for
each condition.

Reaction	Original lb	YPDGL Baker lb	YPD Johnson lb
EX_4abz(e)	0	0	0
EX_ade(e)	0	-0.00767027	-0.00767027
EX_ala_L(e)	0	-0.00073	-0.00073
EX_arg_L(e)	0	-0.00041	-0.00041
EX_asp_L(e)	0	-0.00023	-0.00023
EX_chol(e)	0	-0.015	-0.015
EX_csn(e)	0	-0.002222703	-0.002222703
EX_etoh(e)	0	0	-20
EX_glc(e)	-10	-20	0
EX_glu_L(e)	0	-0.00056	-0.00056
EX_gly(e)	0	-0.0004	-0.0004
EX_glyc(e)	0	0	0
EX_gsn(e)	0	-0.032575676	-0.032575676
EX_his_L(e)	0	-0.00064	-0.00064
EX_ile_L(e)	0	-0.00074	-0.00074
EX_lac_L(e)	0	0	0
EX_leu_L(e)	0	-0.00102	-0.00102
EX_lys_L(e)	0	-0.0003	-0.0003
EX_met_L(e)	0	-0.00089	-0.00089
EX_nac(e)	0	-0.006	-0.006
EX_o2(e)	-2	-2	-2
EX_phe_L(e)	0	-0.00081	-0.00081
EX_pnto_R(e)	0	-0.00105	-0.00105
EX_pro_L(e)	0	-0.00039	-0.00039
EX_ribflv(e)	0	-0.00125	-0.00125
EX_ser_L(e)	0	-0.00079	-0.00079
EX_thm(e)	0	-0.00021	-0.00021
EX_thr_L(e)	0	-0.00062	-0.00062
EX_trp_L(e)	0	-0.01667	-0.01667
EX_tyr_L(e)	0	-0.00027	0.00027
EX_uri(e)	0	-0.005883784	-0.005883784
EX_val_L(e)	0	-0.00076	-0.00076
EX_xtsn(e)	0	-0.002222703	-0.002222703

Table 3.6.2   Enrichment of metabolic pathways

Statistical enrichment for metabolic pathways using a one-tailed Fisher's exact test. The genes from Baker's experiments were tested at a threshold of $p \leq 0.01$ and the genes from Johnson's experiments were tested at a threshold of $p \leq 0.001$. Statistically enriched genes for each condition are displayed in the columns prior to the p values. Pathways with $p < 0.05$ are shaded in gray.

Theme(s)	GO ID	Pathway	# Genes	GCR1A B # genes	GCR1A B p	GCR1A J # genes	GCR1A J p
Glycolysis, Gluconeogenesis, and Fermentation	0044107	cellular alcohol metabolic process	34	20	2.85e-04	4	0.896
	0019660	glycolytic fermentation	7	5	0.0253	4	0.025
	0006094	gluconeogenesis	32	6	0.938	15	2.11e-04
	0006096	glycolytic process	30	9	0.528	14	3.63e-04
Respiration	0097034	mitochondrial respiratory chain complex IV biogenesis	21	3	0.968	16	1.31e-06
	0017062	respiratory chain complex III assembly	7	1	0.91	5	0.00317
	0034552	respiratory chain complex II assembly	5	0	1	4	0.00489
	0043457	regulation of cellular respiration	6	3	0.24	4	0.025
Glycine, Serine, and Purine Biosynthesis	0006164	purine nucleotide biosynthetic process	55	29	1.93e-04	17	0.0243
	0006730	one-carbon metabolic process	17	12	4.8e-04	11	3.28e-05
	0035999	tetrahydrofolate interconversion	6	5	0.00952	4	0.0125
	0006544	glycine metabolic process	9	6	0.0218	6	0.00194
	0006144	purine nucleobase metabolic process	19	10	0.0265	9	0.00369
	0019464	glycine decarboxylation via glycine cleavage system	2	2	0.0849	2	0.0339
Nitrogen Metabolism	0006525	arginine metabolic process	17	4	0.776	11	3.28e-05
	0019740	nitrogen utilization	20	9	0.097	9	0.00827
	0019627	urea metabolic process	4	0	1	3	0.0215
Inositol Regulation	0006661	phosphatidylinositol biosynthetic process	35	18	0.00447	7	0.566
	1900088	regulation of inositol biosynthetic process	3	1	0.644	3	0.00623

Chapter 4   Conclusions and future directions

4.1   Conclusions about metabolic modeling

In the second chapter, a condition-specific metabolic modelling software called TASMANIAN DEVIL has been developed. It allows for importing any model organism reconstruction in either SBML or COBRA format. The software was constructed upon previously designed libraries and algorithms. Its modeling assumptions and performance were validated using publically available datasets in *S. cerevisiae*. It introduced a novel modeling approach to FBA that is similar to MFA to reduce the interconnectivity of metabolites and complexity of the network. It is easily installable and available for Python 2 and 3, which increases its potential user base and allows for continued development from the community in the future.

4.2   Conclusions about Gcr1 and the influence of inositol pyrophosphorylation

4.2.1   Pyrophosphorylation deactivates the canonical Gcr1 isoform, and this may be catalyzed by a domain within Gcr1

It has been concluded from the third chapter that Gcr1 controls the diauxic shift. The switch from aerobic glycolysis and fermentation through glucose consumption to oxidative phosphorylation by ethanol import is triggered by the pyrophosphorylation of Gcr1. Pyrophosphorylation of Gcr1 would only occur if enough ATP has been generated and the system is switching from a glucose-rich to a glucose-limited state. Pyrophosphorylation has been characterized as a nonenzymatic event that donates a phosphate as a posttranslational modification to particular proteins. A pyrophosphorylation consensus sequence across proteins does not exist, and it is unclear how phosphates are donated from inositol pyrophosphates to specific proteins. It has also been difficult to identify which proteins are pyrophosphorylated, as a pyrophosphate affinity reagent is not yet publically available [109, 110]. However, it has been shown that proteins across eukaryotes are phosphorylated by 5PP-IP5.[111]

This led to the investigation that perhaps Gcr1 is somehow able to catalyze its own pyrophosphorylation. The canonical isoform of Gcr1 is 785 amino acids long, as illustrated by **Figure 4.5.1a**. The DNA binding domain is at the C-terminus of the protein. Gcr1 is pyrophosphorylated between amino acids 509-518, where there are several serine residues and ambiguous phosphorylation patterns [82, 112]. To predict the structure of the Icp4 domain, this region of Gcr1 was modeled after the protein structure 3zxa using MODELLER, which had a similar secondary structure according to mGenTHREADER (p = 1e-4 using confidence estimation for identifying a false positive) and whose structural compatibility was checked with Verify3D [113–116]. This domain appears to be shaped like a hand with a hinge, as shown in **Figure 4.5.1b**. When Gcr1 becomes pyrophosphorylated, it is possible that the faces of the protein are no longer able to interact with other proteins, thus disrupting the Gcr1-Gcr2 heterodimer complex [82].

Over the region where Gcr1 is pyrophosphorylated is a conserved multi-domain with structural similarity to the herpes simplex virus gene, ICP4 (with an E-value = 0.04, calculated by DELTA-BLAST, which is the number of chance alignments predicted with a score as high). Icp4 has been identified to be involved in lytic gene expression [117]. This Icp4 multi-domain is conserved across eukaryotes according to DELTA-BLAST, an enhanced domain lookup method that uses a predefined set of position score matrices (PSSMs) constructed from NCBI's Conserved Domain Database (CDD) to find distant homologues of queried proteins [118]. It is possible that this Icp4 similar region of Gcr1 catalyzes Gcr1's own pyrophosphorylation, and that this domain might be able to catalyze other proteins to pyrophosphorylate themselves or bind inositol pyrophosphates. Three other genes were found to be significantly similar for this domain in *S. cerevisiae* (PPZ1, TUS1, and SEC31) [118, 119]. Pyrophosphorylation occurs on a prephosphorylated serine residue [85]. Therefore, PhosphoPep, a database of phosphorylation

sites in yeast, worms, flies, and humans, was used to see if phosphorylated serine residues occurred near the GCR1 pyrophosphorylation site in the other proteins [112]. All three proteins contained peptides with phosphorylated serine residues that were unable to be digested. The most significantly similar protein, Ppz1, contained an ambiguous phosphorylation pattern similar to Gcr1. **Figure 4.5.1c** summarizes these results.

4.2.2 The relation of Gcr1 and pyrophosphorylation to the cell cycle and nutrient limitation

GCR1 and RAP1 are highly periodically expressed, and this may be dictated by the level of inositol pyrophosphates. Based on the cell cycle transcription profiles on Cyclebase, GCR1 and RAP1 are only expressed at certain times in the cell cycle (p = 3.68E-05 and p = 9.38E-07, respectively, using a Fourier score for periodicity of expression). They have similar transcriptional profiles that become the most enriched in G2, when a high amount of biosynthesis is needed, although RAP1 peaks its expression at the end of S. GCR2's expression peaks as well in G2. Similarly, GCV1, GCV2, and GCV3 are all at their highest expression levels following GCR1's peak expression [120].

Pyrophosphorylation may also further influence the cell cycle beyond deactivating Gcr1. When the KCS1 homologue, IP6K1, was deleted in mouse embryonic fibroblasts (MEFs), there was an increased number of cells in G2 that were hypodiploid under DNA damage and fewer cells in G1. With this came an increase of cells undergoing apoptosis in G2. In comparison, under DNA damage in wild-type IP6K1 MEFs, there was about an even population of cells stalled in G1 and G2 [121]. When 5PP-IP5 levels increased in 1,5bis-diphosphoinositol tetrakisphosphate (1,5PP2-IP4) deficient mutants in colorectal cancer cells, there was also a slowed G1/S transition [122]. 5PP-IP5 has twice the relative concentration in G1 compared to S for *in vitro* rat mammary tumor cells. 5PP-IP5 peaks in the middle of G1. Then there is a gradual decrease until a spike in 5PP-IP5 concentration at the G2/M barrier, depending on which protocol is used to measure 5PP-IP5 [123]. This evidence indicates that perhaps pyrophosphorylation inhibits the G1/S transition and that it advances the G2/M transition.

Pyrophosphorylation may further regulate the cell cycle and cellular homeostasis by nutrient sensing through Ppz1. The most significantly similar protein of the Icp4 domain to that of Gcr1 in yeast is Ppz1, which is a serine/threonine phosphatase that regulates salt tolerance, making yeast more sensitive to toxicity under high salinity conditions [124]. It can do this in part by inhibiting the G1/S transition. When PPZ1 is overexpressed, cells exhibit a slow-growth phenotype and cannot make the transition from G1 to S [125]. Hal3 binds to Ppz1 and inhibits it from acting upon its targets. When the genomic region containing the Icp4 domain was deleted in PPZ1(Δ1-344), Ppz1 coprecipitated more with its phosphatase inhibitor Hal3 [126]. It was also demonstrated that when HAL3 expression was blocked in a conditional mutant upon doxycycline induction, Ppz1 inhibited potassium import through the Trk1 receptor. This resulted in potassium starvation, which mimicked the effects of phosphate starvation, resulting in the upregulation of PHO gene expression to import extracellular phosphate primarily through the proton/phosphate symporter Pho84 [127]. Pho84 is a high affinity transporter that brings both protons and phosphates into the cell [128]. PHO gene expression is mediated by inositol pyrophosphate 1,5PP2-IP4 generation by Vip1, a heptakisphosphate kinase [127]. Inositol pyrophosphate binds allosterically to the Pho85-Pho80-Pho81 kinase complex, inactivating it and allowing PHO genes to be expressed [129, 130]. Taken together with the assumption that Ppz1 is able to be pyrophosphorylated, this could indicate that inositol pyrophosphorylation disassociates Ppz1 from Hal3 similar to how Gcr1 is disassociated from Gcr2 by pyrophosphorylation, and that this upregulates PHO gene expression to import phosphates and protons. This could provide links between pyrophosphorylation and the diauxic shift to glucose, phosphate, and/or potassium exhaustion.

4.3   Relations to cancer

4.3.1   The role of inositol pyrophosphate deregulation

Inositol pyrophosphates have been identified as key regulators in cellular homeostasis and cell-cycle regulation in yeast. It was investigated whether inositol pyrophosphates became deregulated in cancer cells to promote aberrant growth. Homology for the putative Icp4 multi-domain revealed that several inositol polyphosphatases contained the domain in *H. sapiens* near their phosphatase domains [118]. If the Icp4 multi-domain is a self-phosphorylating pyrophosphate kinase, this could indicate a positive feedback mechanism that promotes the continued formation of inositol pyrophosphates and inhibits inositol pyrophosphatases from reducing the inositol pyroyphosphates to their lower phosphate versions. Also intriguing is that human brain cells had been shown previously to be particularly enriched for 5PP-IP5 [111]. Most of the genes that were not inositol pyrophosphatases with homology to this domain were characterized as being related to brain function [118].

5PP-IP5 formation in cancer cells is disrupted by mutations, deletions, and protein inhibition. In humans, 5PP-IP5 is formed by Ip6k1, Ip6k2, and Ip6k3 [131, 132]. IP6K1 knockout mice are sensitive to insulin and are resistant to becoming obese, IP6K2 knockout mice are more susceptible to developing tumors, whereas IP6K3 null mice develop neurological defects [133–137]. IP6K2 overexpression promotes apoptosis in cancer cells *in vitro*, while knockdown prevents IP6K2 from initiating cell death [138, 139]. While IP6K2 is the isoform most associated with cancer, single nucleotide polymorphisms in IP6K1, IP6K2, and IP6K3 were demonstrated to have an increased risk in developing renal cell carcinoma [140]. In addition, 8.3% of B-cell chronic leukemia patients have chromosomal deletions of IP6K2 [141]. Although only a few cancers may directly downregulate the synthesis of 5PP-IP5 at a genomic level, Ip6k2 activity is inhibited on a broader scale by Hsp90. Hsp90 binds to Ip6k2 and inhibits its kinase activity. It is estimated that 30-50% of cellular Ip6k2 protein is bound to Hsp90 in cell culture analyses of immortalized noncancerous cells [142]. Hsp90 is upregulated in many cancer cells, as it plays roles into most of the hallmarks of cancer, thereby potentially increasing its inhibition of Ip6k2 [143, 144].

While all of the isoforms of the inositol hexakisphosphate kinases in humans are able to produce 5PP-IP5, Ip6k2 may confer selectivity for cancer-related targets and signaling cascades based off of the isoform being in complex with other proteins that it either pyrophosphorylates or donates 5PP-IP5 to for binding. One example is that Ip6k2 is in complex with Tti1 and Tel2, which are targets for phosphorylation of CK2. Upon DNA damage, CK2 binds allosterically to 5PP-IP5, which then phosphorylates Tti1 and Tel2, promoting the proteins to be in complex with ATM and DNA-PKcs. ATM/DNA-PKcs phosphorylate p53, promoting apoptosis [135]. Ip6k2 is necessary for p53-mediated apoptosis [145]. P53 is a master regulator of genomic stability and the cell cycle; it determines whether the cell repairs its DNA and continues proliferating or induces apoptosis after DNA damage has occurred [146–149]. TP53 is the most mutated gene in cancer with over 25,000 point mutations, with over 75% of these mutations resulting in the loss of WT p53 function [150, 151]. Not accounting for deletions, it has been estimated that 17% of all cancers have p53 mutations [152, 153]. In many other cancers with a functional p53 protein, Mdm2 is upregulated, which inhibits the transactivation domain of p53 [154]. Ip6k2 deregulation is another mechanism of p53 inactivation. While Ip6k2 and pyrophosphates may be important to prevent tumor formation, it has also been shown that Ip6k2 is activated downstream of the WNT/β catenin pathway as part of the epithelial to mesenchymal transition (EMT) [155]. One way it promotes oncogenesis in tandem with IP7 is by sequestering the tumor suppressor Lkb1 in the nucleus so that it cannot act on its cytosolic targets [135]. It is important to determine what proteins are pyrophosphorylated and bind to particular pyrophosphates to better understand the

mechanisms by which pyrophosphates inhibit and promote tumorigenesis so that more specific pharmaceuticals targeting the inositol pyrophosphate pathway can be developed.

4.3.2 Insights into cancer metabolism

In yeast, high amounts of 5PP-IP5 trigger a switch from glycolysis to respiration, and there appears to be a similar mechanism conserved in mammals. When KCS1 is deleted in yeast, there is a 4-fold increase in ATP concentration, a 3-fold increase in ethanol production, and respiration decreases. Similarly, when IP6K1 is deleted in MEFs, there is a 3-fold increase in ATP concentration and respiration decreases [82]. 1,5PP2-IP4 has been demonstrated to be important for glycolytic metabolism as well in colorectal cancer cells. 1,5PP2-IP4 is produced by the two isoforms of PPIP5K, the homologs of Vip1, which add an additional pyrophosphate to 5PP-IP5. When the PPIP5K isoforms are deleted, there is a modest but significant increase in ATP, an increase in glycolysis and lactate production, but also an increase in oxygen consumption. Surprisingly, when PPIP5K was deleted, there was a 3-fold increase in 5PP-IP5 levels, which is 25-fold greater than the amount of 1,5PP2-IP4 lost. The authors did not investigate this mechanism of how 5PP-IP5 synthesis was upregulated, but it indicates that 1,5PP2-IP4 may play a signaling role that downregulates the production of 5PP-IP5. It was also shown in this study that 5PP-IP5 levels are directly proportional to ATP levels, signifying that 5PP-IP5 is a metabolic sensor as had been hypothesized before due to the high K_m of the Ip6ks [86, 122].

Both Ip6k1 and Ip6k2 are responsible for generating 5PP-IP5 in cancer cells, but it is unclear to what degree and how much this varies among tissues. There may even be variability within a tumor. It has been demonstrated that in the same colorectal cancer cell line, HCT119, it can become divergent depending on cell culture conditions and cell lineage history, expressing varying degrees of IP6K1 and IP6K2 [156]. It has also been demonstrated that in HCT116 cells either IP6K1 or IP6K2 deletion can almost completely decimate 5PP-IP5 levels depending on cell lineage [145, 157]. Similarly, 1,5PP2-IP4 levels vary widely depending on cell linage in HCT119, which would also influence 5PP-IP5 levels [122, 156].

IP6K1 mutants in MEFs have exhibited similar effects on metabolism to KCS1 mutants in yeast, which has implications for cancer metabolism [82]. By inhibiting IP6K1 kinase activity in MEFs by N2-(m-Trifluorobenzyl), N6-(p-nitrobenzyl)purine (TNP), an IP6K inhibitor, AKT activity is increased, thus promoting energy expenditure by glycolysis through glucose consumption [158]. 5PP-IP5 binds to AKT and inhibits Pdk1-mediated AKT phosphorylation, thereby inactivating AKT [133, 159]. AKT is a known oncogene that stimulates aerobic glycolysis in cancer cells [160]. Thus 5PP-IP5 may potentially inhibit aerobic glycolysis in cancer cells through Ip6k1 by inactivating AKT. Studies have not yet demonstrated which IP6K isoform is responsible for this metabolic phenotype in cancer cells. However, it appears that 5PP-IP5 produced by Ip6k1, which binds to and inactivate AKT, is a regulator of glycolytic homeostasis, much like 5PP-IP5 formed by Kcs1 in yeast is to pyrophosphorylate and inactivate Gcr1.

4.4 Future directions

4.4.1 Software development

There are a few software enhancements that would be desirable to add to TASMANIAN DEVIL. One of these would be introducing manually specified parallelization across computer cores for optimization to make the models more operable on a supercomputer. This would make performing optimizations on larger stoichiometric models like Recon2 more feasible. This update to parallelization from processes being forked would also allow the package to be executed on any operating system, as the forking in Python is not permitted in Windows. Another enhancement would be to add multiple linear solvers to the flux module to allow for user preference. These changes would increase the user base and increase capabilities for complex modeling.

4.4.2 Metabolic model simplification analysis compared to loop correction software

The effectiveness of the metabolic simplifications implemented in TASMANIAN DEVIL should be compared to loop correction software. The Monte Carlo hit-and-run techniques that many of these loop correction studies utilize focus on a subset of flux loops, frequently identifying and correcting for loops that are short in pathway length [28, 29]. It should be more closely investigated whether all of the large cofactor loops are accounted for in these optimizations. A comprehensive study should be conducted comparing these loop correction techniques to better portray the benefit of the reduced network reduction models. It should also be analyzed if there are additional benefits for accuracy when using a loop algorithm in addition to implementing network reduction models.

4.4.3 Flux balance analysis

With the increased use of GC-MS for metabolomics analysis, the field is more apt to utilize MFA for contextualizing systems deterministically than it had been in the past. However, this approach is not applicable for every system, such as for human metabolism, in which it is difficult to achieve isotopic steady-state. For systems like this, unsteady FBA may be more appropriate if GC-MS is available. If GC-MS is unavailable, or if the desire is to model genome-scale metabolism, researchers should employ some type of FBA method. The network reduced condition-specific modeling presented in this dissertation is an appealing approach for estimating phenotypic fluxes. An improvement that could be made in general to FBA is to adjust the biomass equation. The concentrations of biomass metabolites do not always remain in constant proportions, and the proportions vary among organisms. Determining the ratios of biomass metabolites to each other for a particular cell line would generate more accurate flux predictions.

4.4.4 Validation of the glycine complex genes in *S. cerevisiae* in relation to GCR1

The glycine cleavage complex upregulated by Gcr1 transactivation was predicted to be important for energy generation to fuel ATP generation. Prior evidence showed that the glycine cleavage system was involved in one carbon metabolism for the generation of purines and pyrimidines, and it was utilized preferentially in highly proliferative cells, such as cancer cells [100, 161–163]. However, it has not been directly validated if Gcr1 binds to and promotes GCV1, GCV2, and GCV3 transcription. ChIP-Seq experiments would be able to determine this, as well as if this interaction decreases upon the initiation of the diauxic shift and the inactivation of the canonical Gcr1 isoform. Through mutant studies of GCV1, GCV2, and GCV3, it could also be determined the effect they have on ATP production, growth rates, and GCR1 transcript levels. Demonstrating the importance of the glycine cleavage complex in relation to high growth could also have an impact on studying cancer, due to cancer cells possibly relying on the glycine cleavage complex for greater ATP generation.

4.4.5 Icp4 domain validation as a protein pyrophosphorylation domain

The Icp4 domain was identified as a conversed domain according to DELTA-BLAST, with organisms containing this domain ranging from viruses to mammals. In Gcr1, this domain contains a site that is pyrophosphorylated. The mechanism by which a protein becomes pyrophosphorylated is unclear as well as why certain proteins are targeted. Phosphoproteome data indicated that the other proteins in yeast with this domain had similar phosphorylation patterns on peptides in the region of where pyrophosphorylation potentially could occur, with the most statistically significant protein being Ppz1. It should be validated if Ppz1 can indeed be pyrophosphorylated. If so, it should also be determined what the structure of this domain is and how the mechanism of pyrophosphorylation occurs. If this domain does characterize a protein region that becomes pyrophosphorylated by 5PP-IP5, this would have profound impacts for studying proteins that have this domain. It would also allow for the better understanding of the

influence that pyrophosphorylation by 5PP-IP5 has. In addition, when a pyrophosphorylation reagent becomes publically available, this will be able to clearly identify all proteins that are pyrophosphorylated [110]. In addition, the community should strive to create a better method to determine which proteins bind 5PP-IP5 and 1,5PP-IP4 to fully understand the interplay of the proteins that associate with these high energy metabolites.

4.4.6   Analysis of pyrophosphates on cancer metabolism

Inositol pyrophosphates help regulate cellular homeostasis mechanisms and the cell cycle, which is why they become deregulated in cancer cells to allow for uncontrolled growth. 5PP-IP5 targets many cancer-related proteins. 5PP-IP5 inactivates AKT, thereby downregulating glycolytic metabolism, and it promotes the phosphorylation of p53 upon DNA damage, thereby triggering apoptosis [133, 135]. However, inositol pyrophosphates may also have oncogenic potential, as they upregulate the EMT [155]. Studies need to better demonstrate the functions of IP6K isoform in cancer cells, if these functions vary across tissue type, and if these roles differ among cells within a tumor.

4.4.7   Drug targets related to inositol pyrophosphates

Once the field has a more complete understanding of the role of inositol pyrophosphates, as well as their roles in preventing tumor formation and later in tumorigenesis, pharmaceuticals can be developed to better target the pathway. One target may be Hsp90, since it binds to and inactivates IP6k2. There have been 18 drugs that have entered clinical trials that have targeted Hsp90. None of these drugs have entered a phase 3 clinical trial. Currently, only 5 remain in clinical trials [164]. By upregulating 5PP-IP5 levels in HCT119 cells by deleting PPIP5K and preventing 1,5PP-IP4 formation, cellular proliferation slowed, as most of the cells became stalled in G1 [122]. If 5PP-IP5 can be upregulated and the cellular machinery which regulates apoptosis can be reactivated, then the tumors can be destroyed. However, there are likely many ways in which some tumor cells use 5PP-IP5 to promote tumorigenesis and various ways that the apoptotic machinery has become inactivated based upon a patients' mutational and copy-number differences.

5PP-IP5 has been demonstrated to be a master regulator of cell cycle control and homeostasis. Preventative research could also be performed to use 5PP-IP5 or an earlier formed inositol phosphate pathway derivate to target cancer cells early and inhibit tumorigenesis. If studies are successful in animal models, the metabolite could be introduced in a clinical trial. Unless intellectual property is developed around targeting cancer cells specifically though, it is unlikely that such a study would be undertaken by a drug company, as generic drugs would undercut the market. Thus a multicompany federally sponsored approach would be necessary, but such a large-scale study would be unlikely.

4.5 Figures and tables

Figure 4.5.1 Gcr1 structure and similar proteins to the Icp4 domain

(**a**) Domains and regions of the Gcr1 protein. The Icp4 domain is highlighted by the red box. (**b**) Modelled structured of Icp4 domain. (c) Icp4 domain alignment and phosphorylation sites. Reference proteins below the significance threshold of 0.05 for DELTA-BLAST identified as aligning to Gcr1 (from 340-640 amino acids). The Icp4 conserved multi-domain (from 362-589 amino acids) is highlighted by the red box. Phosphorylation patterns close to the peptide containing the pyrophosphorylation site of GCR1 are from PhosphoPep and indicated by the gray box.

51
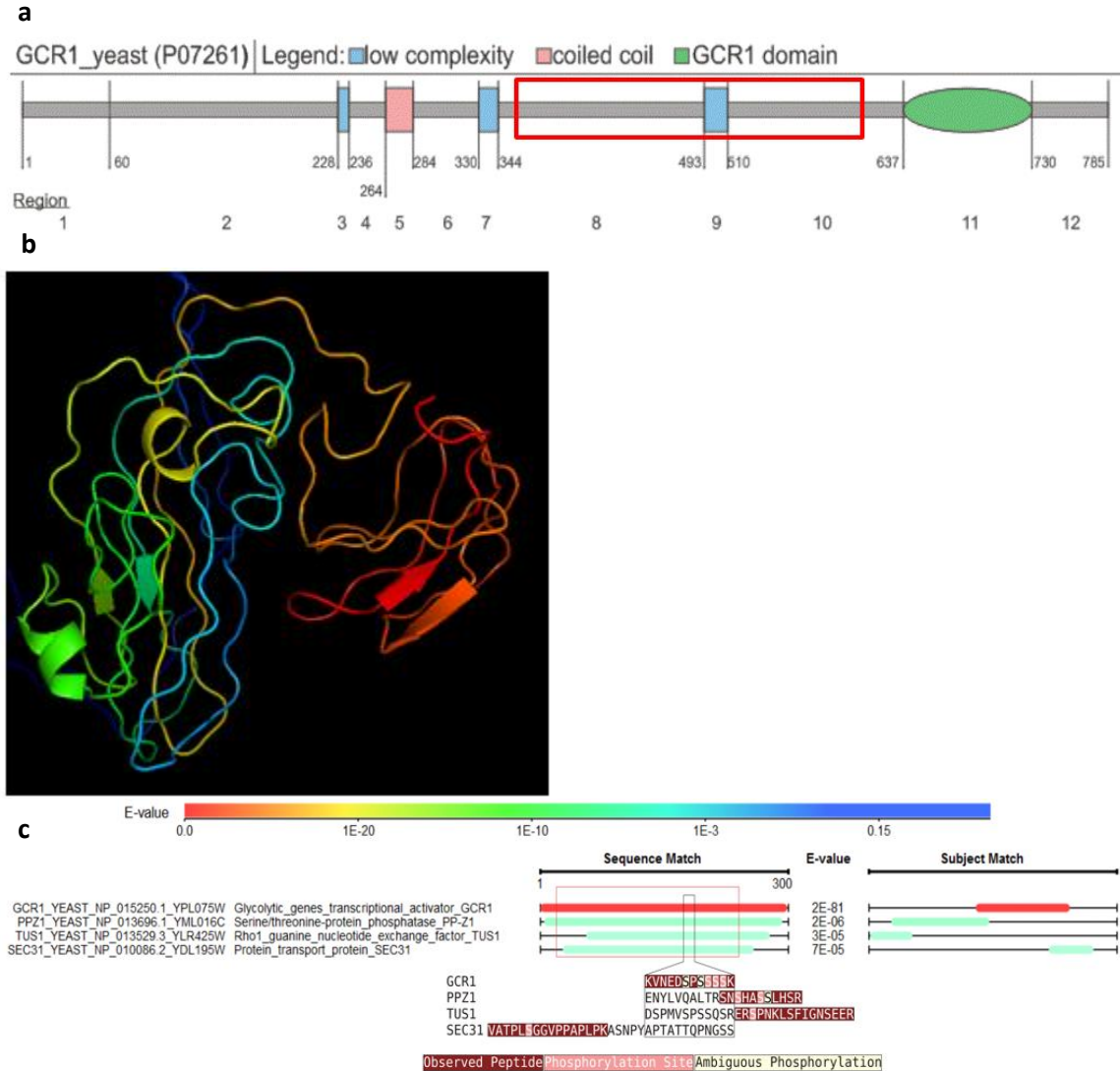
Bibliography

1. Pronk JT, Yde Steensma H, Van Dijken JP. Pyruvate metabolism in Saccharomyces cerevisiae. Yeast. 1996;12:1607–33.

2. Zampar GG, Kummel A, Ewald J, Jol S, Niebel B, Picotti P, et al. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol Syst Biol. 2013;9:651.

3. Murphy JP, Stepanova E, Everley RA, Paulo JA, Gygi SP. Comprehensive Temporal Protein Dynamics during the Diauxic Shift in Saccharomyces cerevisiae. Mol Cell Proteomics. 2015;14:2454–65.

4. Deminoff SJ, Santangelo GM. Rap1p requires Gcr1p and Gcr2p homodimers to activate ribosomal protein and glycolytic genes, respectively. Genetics. 2001;158:133–43.

5. Gray J V, Petsko GA, Johnston GC, Ringe D, Singer RA, Werner-Washburne M. "Sleeping beauty": quiescence in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. 2004;68:187–206.

6. Hossain MA, Claggett JM, Edwards SR, Shi A, Pennebaker SL, Cheng MY, et al. Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for Metabolic Adjustment in Response to Glucose Availability. Mol Cell. 2016;62:346–58.

7. Warburg O, Posener K, Negelein E. Ueber den stoffwechsel der tumoren. Biochem Z. 1924;152:319–44.

8. Crabtree HG. Observations on the carbohydrate metabolism of tumours. Biochem J. 1929;23:536–45.

9. Doherty JR, Cleveland JL. Targeting lactate metabolism for cancer therapeutics. J Clin Invest. 2013;123:3685–92.

10. Sonveaux P, Vegran F, Schroeder T, Wergin MC, Verrax J, Rabbani ZN, et al. Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. J Clin Invest. 2008;118:3930–42.

11. Ho J, de Moura MB, Lin Y, Vincent G, Thorne S, Duncan LM, et al. Importance of glycolysis and oxidative phosphorylation in advanced melanoma. Mol Cancer. 2012;11:76.

12. Haq R, Shoag J, Andreu-Perez P, Yokoyama S, Edelman H, Rowe GC, et al. Oncogenic BRAF regulates oxidative metabolism via PGC1alpha and MITF. Cancer Cell. 2013;23:302–15.

13. Roesch A, Vultur A, Bogeski I, Wang H, Zimmermann KM, Speicher D, et al. Overcoming intrinsic multidrug resistance in melanoma by blocking the mitochondrial respiratory chain of slow-cycling JARID1B(high) cells. Cancer Cell. 2013;23:811–25.

14. Corazao-Rozas P, Guerreschi P, Jendoubi M, Andre F, Jonneaux A, Scalbert C, et al. Mitochondrial oxidative stress is the Achille's heel of melanoma cells resistant to Braf-mutant inhibitor. Oncotarget. 2013;4:1986–98.

15. Kell DB, Goodacre R. Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. Drug Discov Today. 2014;19:171–82.

16. Jerby L, Ruppin E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. Clin Cancer Res. 2012;18:5572–84.

17. Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol. 2010;13:344–9.

18. Resendis-Antonio O, Gonzalez-Torres C, Jaime-Munoz G, Hernandez-Patino CE, Salgado-Munoz CF. Modeling metabolism: a window toward a comprehensive interpretation of networks in cancer. Semin Cancer Biol. 2015;30:79–87.

19. Wiechert W. 13C metabolic flux analysis. Metab Eng. 2001;3:195–206.

20. Zamboni N, Fendt SM, Ruhl M, Sauer U. (13)C-based metabolic flux analysis. Nat Protoc. 2009;4:878–92.

21. Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol. 2015;42:317–25.

22. Bordbar A, Yurkovich JT, Paglia G, Rolfsson O, Sigurjonsson OE, Palsson BO. Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. Sci Rep. 2017;7:46249.

23. Mahadevan R, Edwards JS, Doyle 3rd FJ. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J. 2002;83:1331–40.

24. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol. 1994;60:3724–31.

25. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol. 2010;28:245–8.

26. Beard DA, Liang SD, Qian H. Energy balance for analysis of complex metabolic networks. Biophys J. 2002;83:79–86.

27. Price ND, Famili I, Beard DA, Palsson BO. Extreme pathways and Kirchhoff's second law. Biophys J. 2002;83:2879–82.

28. Schellenberger J, Lewis NE, Palsson BO. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J. 2011;100:544–53.

29. De Martino D, Capuani F, Mori M, De Martino A, Marinari E. Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks. Metabolites. 2013;3:946–66.

30. Mao L, Verwoerd WS. Model-driven elucidation of the inherent capacity of Geobacter sulfurreducens for electricity generation. J Biol Eng. 2013;7:14.

31. Desouki AA, Jarre F, Gelius-Dietrich G, Lercher MJ. CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. Bioinformatics. 2015.

32. Rossell S, Huynen MA, Notebaart RA. Inferring metabolic states in uncharacterized environments using gene-expression measurements. PLoS Comput Biol. 2013;9:e1002988.

33. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol. 2012;3:299.

34. Machado D, Herrgard M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol. 2014;10:e1003580.

35. Kim MK, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. Comput Struct Biotechnol J. 2014;11:59–65.

36. Dandekar T, Fieselmann A, Majeed S, Ahmed Z. Software applications toward quantitative metabolic flux analysis and modeling. Brief Bioinform. 2014;15:91–107.

37. Simeonidis E, Price ND. Genome-scale modeling for metabolic engineering. J Ind Microbiol Biotechnol. 2015;42:327–38.

38. Pacheco MP, Pfau T, Sauter T. Benchmarking Procedures for High-Throughput Context Specific Reconstruction Algorithms. Front Physiol. 2016;6:410.

39. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. Cell Syst. 2017;4:318–329.e6.

40. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics. 2010;26:3140–2.

41. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol. 2008;4:e1000082.

42. Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, et al. Improving metabolic flux predictions using absolute gene expression data. BMC Syst Biol. 2012;6:73.

43. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. PLoS Comput Biol. 2009;5:e1000489.

44. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2010;107:17845–50.

45. Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics. 2011;27:541–7.

46. Kim J, Reed JL. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol. 2012;13:R78-2012-13-9-r78.

47. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol Syst Biol. 2010;6:401.

48. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst Biol. 2012;6:153.

49. Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. PLoS Comput Biol. 2014;10:e1003424.

50. Gruning NM, Rinnerthaler M, Bluemlein K, Mulleder M, Wamelink MM, Lehrach H, et al. Pyruvate kinase triggers a metabolic feedback loop that controls redox metabolism in respiring cells. Cell Metab. 2011;14:415–27.

51. Qian H, Beard DA. Metabolic futile cycles and their functions: a systems analysis of energy and control. Syst Biol (Stevenage). 2006;153:192–200.

52. Jouhten P, Rintala E, Huuskonen A, Tamminen A, Toivari M, Wiebe M, et al. Oxygen dependence of metabolic fluxes and energy generation of Saccharomyces cerevisiae CEN.PK113-

1A. BMC Syst Biol. 2008;2:60.

53. Young JD, Shastri AA, Stephanopoulos G, Morgan JA. Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis. Metab Eng. 2011;13:656–65.

54. Ahn WS, Antoniewicz MR. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab Eng. 2011;13:598–609.

55. Murphy TA, Dang C V, Young JD. Isotopically nonstationary 13C flux analysis of Myc-induced metabolic reprogramming in B-cells. Metab Eng. 2013;15:206–17.

56. Carinhas N, Pais DA, Koshkin A, Fernandes P, Coroadinha AS, Carrondo MJ, et al. Metabolic flux profiling of MDCK cells during growth and canine adenovirus vector production. Sci Rep. 2016;6:23529.

57. Bornstein BJ, Keating SM, Jouraku A, Hucka M. LibSBML: an API library for SBML. Bioinformatics. 2008;24:880–1.

58. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013;7:74.

59. Mo ML, Palsson BO, Herrgard MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol. 2009;3:37.

60. Funahashi A, Morahashi M, Kitano H, Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. Biosilico. 2003;1:159–62.

61. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol. 2008;26:1003–10.

62. Szappanos B, Kovacs K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet. 2011;43:656–62.

63. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013;31:419–25.

64. Wilson WA, Hawley SA, Hardie DG. 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. Curr Biol. 1996;6:1426–34.

65. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40 Database issue:D700-5.

66. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000;11:4241–57.

67. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418:387–91.

68. Snitkin ES, Dudley AM, Janse DM, Wong K, Church GM, Segre D. Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. Genome Biol. 2008;9:R140–2008–9–9–r140. Epub 2008 Sep 22.

69. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23:1846–7.

70. Bolstad B. Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization. University of California, Berkeley; 2004.

71. Vieira E, Brandao T, Ferreira IM. Evaluation of Brewer's spent yeast to produce flavor enhancer nucleotides: influence of serial repitching. J Agric Food Chem. 2013;61:8724–9.

72. Rintala E, Toivari M, Pitkanen JP, Wiebe MG, Ruohonen L, Penttila M. Low oxygen levels as a trigger for enhancement of respiratory metabolism in Saccharomyces cerevisiae. BMC Genomics. 2009;10:461.

73. Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast. 1992;8:501–17.

74. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, et al. Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. J Bacteriol. 1999;181:6679–88.

75. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 2003;4:117–2003–4–9–117. Epub 2003 Aug 29.

76. Barbara KE, Haley TM, Willis KA, Santangelo GM. The transcription factor Gcr1 stimulates cell growth by participating in nutrient-responsive gene expression on a global level. Mol Genet Genomics. 2007;277:171–88.

77. Lopez MC, Baker H V. Understanding the growth phenotype of the yeast gcr1 mutant in terms of global genomic expression patterns. J Bacteriol. 2000;182:4970–8.

78. Sasaki H, Uemura H. Influence of low glycolytic activities in gcr1 and gcr2 mutants on the expression of other metabolic pathway genes in Saccharomyces cerevisiae. Yeast. 2005;22:111–27.

79. Pennebaker SL. Features of the GCR1 Intron that Contribute to its Retention and Gcr1 Regulation. 2013.

80. Shah KH, Zhang B, Ramachandran V, Herman PK. Processing body and stress granule assembly occur by independent and differentially regulated pathways in Saccharomyces cerevisiae. Genetics. 2013;193:109–23.

81. Tornow J, Santangelo G. The GCR1 gene of Saccharomyces cerevisiae is a split gene with an unusually long intron. Genetics. 1994;138:973–4.

82. Szijgyarto Z, Garedew A, Azevedo C, Saiardi A. Influence of inositol pyrophosphates on cellular energy dynamics. Science. 2011;334:802–5.

83. Tornow J, Zeng X, Gao W, Santangelo GM. GCR1, a transcriptional activator in Saccharomyces cerevisiae, complexes with RAP1 and can function without its DNA binding domain. EMBO J. 1993;12:2431–7.

84. Sasaki H, Kishimoto T, Mizuno T, Shinzato T, Uemura H. Expression of GCR1, the transcriptional activator of glycolytic enzyme genes in the yeast Saccharomyces cerevisiae, is

69. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23:1846–7.

70. Bolstad B. Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization. University of California, Berkeley; 2004.

71. Vieira E, Brandao T, Ferreira IM. Evaluation of Brewer's spent yeast to produce flavor enhancer nucleotides: influence of serial repitching. J Agric Food Chem. 2013;61:8724–9.

72. Rintala E, Toivari M, Pitkanen JP, Wiebe MG, Ruohonen L, Penttila M. Low oxygen levels as a trigger for enhancement of respiratory metabolism in Saccharomyces cerevisiae. BMC Genomics. 2009;10:461.

73. Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast. 1992;8:501–17.

74. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, et al. Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. J Bacteriol. 1999;181:6679–88.

75. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 2003;4:117–2003–4–9–117. Epub 2003 Aug 29.

76. Barbara KE, Haley TM, Willis KA, Santangelo GM. The transcription factor Gcr1 stimulates cell growth by participating in nutrient-responsive gene expression on a global level. Mol Genet Genomics. 2007;277:171–88.

77. Lopez MC, Baker H V. Understanding the growth phenotype of the yeast gcr1 mutant in terms of global genomic expression patterns. J Bacteriol. 2000;182:4970–8.

78. Sasaki H, Uemura H. Influence of low glycolytic activities in gcr1 and gcr2 mutants on the expression of other metabolic pathway genes in Saccharomyces cerevisiae. Yeast. 2005;22:111–27.

79. Pennebaker SL. Features of the GCR1 Intron that Contribute to its Retention and Gcr1 Regulation. 2013.

80. Shah KH, Zhang B, Ramachandran V, Herman PK. Processing body and stress granule assembly occur by independent and differentially regulated pathways in Saccharomyces cerevisiae. Genetics. 2013;193:109–23.

81. Tornow J, Santangelo G. The GCR1 gene of Saccharomyces cerevisiae is a split gene with an unusually long intron. Genetics. 1994;138:973–4.

82. Szijgyarto Z, Garedew A, Azevedo C, Saiardi A. Influence of inositol pyrophosphates on cellular energy dynamics. Science. 2011;334:802–5.

83. Tornow J, Zeng X, Gao W, Santangelo GM. GCR1, a transcriptional activator in Saccharomyces cerevisiae, complexes with RAP1 and can function without its DNA binding domain. EMBO J. 1993;12:2431–7.

84. Sasaki H, Kishimoto T, Mizuno T, Shinzato T, Uemura H. Expression of GCR1, the transcriptional activator of glycolytic enzyme genes in the yeast Saccharomyces cerevisiae, is

positively autoregulated by Gcr1p. Yeast. 2005;22:305–19.

85. Bhandari R, Saiardi A, Ahmadibeni Y, Snowman AM, Resnick AC, Kristiansen TZ, et al. Protein pyrophosphorylation by inositol pyrophosphates is a posttranslational event. Proc Natl Acad Sci U S A. 2007;104:15305–10.

86. Wundenberg T, Mayr GW. Synthesis and biological actions of diphosphoinositol phosphates (inositol pyrophosphates), regulators of cell homeostasis. Biol Chem. 2012;393:979–98.

87. Barbara KE, Willis KA, Haley TM, Deminoff SJ, Santangelo GM. Coiled coil structures and transcription: an analysis of the S. cerevisiae coilome. Mol Genet Genomics. 2007;278:135–47.

88. Pommier I, Liu H, Reinhold W. Microarray Data Analysis. http://dis.

89. Barrett CL, Schwab RB, Jung H, Crain B, Goff DJ, Jamieson CH, et al. Transcriptome sequencing of tumor subpopulations reveals a spectrum of therapeutic options for squamous cell lung cancer. PLoS One. 2013;8:e58714.

90. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. PLoS One. 2009;4:e7767.

91. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.

92. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

93. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 2005;21:2067–75.

94. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics. 2008;9:321–32.

95. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

96. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57:289–300. doi:10.2307/2346101.

97. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23:257–8.

98. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

99. Otulakowski G, Robinson BH. Isolation and sequence determination of cDNA clones for porcine and human lipoamide dehydrogenase. Homology to other disulfide oxidoreductases. J Biol Chem. 1987;262:17313–8.

100. Kikuchi G, Motokawa Y, Yoshida T, Hiraga K. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Japan Acad B, Phys Biol Sci. 2008;84:246–63.

101. Razak MA, Begum PS, Viswanath B, Rajagopal S. Multifarious Beneficial Effect of

Nonessential Amino Acid, Glycine: A Review. Oxid Med Cell Longev. 2017;2017:1716701.

102. Cheng C, Kacherovsky N, Dombek KM, Camier S, Thukral SK, Rhim E, et al. Identification of potential target genes for Adr1p through characterization of essential nucleotides in UAS1. Mol Cell Biol. 1994;14:3842–52.

103. He XJ, Mulford KE, Fassler JS. Oxidative stress function of the Saccharomyces cerevisiae Skn7 receiver domain. Eukaryot Cell. 2009;8:768–78.

104. Giannattasio S, Liu Z, Thornton J, Butow RA. Retrograde response to mitochondrial dysfunction is separable from TOR1/2 regulation of retrograde gene expression. J Biol Chem. 2005;280:42528–35.

105. Shirra MK, Patton-Vogt J, Ulrich A, Liuta-Tehlivets O, Kohlwein SD, Henry SA, et al. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the INO1 gene in a snf1 mutant strain of Saccharomyces cerevisiae. Mol Cell Biol. 2001;21:5710–22.

106. Ye C, Bandara WM, Greenberg ML. Regulation of inositol metabolism is fine-tuned by inositol pyrophosphates in Saccharomyces cerevisiae. J Biol Chem. 2013;288:24898–908.

107. Henry SA, Kohlwein SD, Carman GM. Metabolism and regulation of glycerolipids in the yeast Saccharomyces cerevisiae. Genetics. 2012;190:317–49.

108. Young BP, Shin JJ, Orij R, Chao JT, Li SC, Guan XL, et al. Phosphatidic acid is a pH biosensor that links membrane biogenesis to metabolism. Science. 2010;329:1085–8.

109. Yates LM, Fiedler D. Establishing the stability and reversibility of protein pyrophosphorylation with synthetic peptides. Chembiochem. 2015;16:415–23.

110. Conway JH. Chemical Tools to Study Protein Pyrophosphorylation . 2015.

111. Saiardi A, Bhandari R, Resnick AC, Snowman AM, Snyder SH. Phosphorylation of proteins by inositol pyrophosphates. Science. 2004;306:2101–5.

112. Bodenmiller B, Campbell D, Gerrits B, Lam H, Jovanovic M, Picotti P, et al. PhosphoPep--a database of protein phosphorylation sites in model organisms. Nat Biotechnol. 2008;26:1339–40.

113. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using MODELLER. Curr Protoc protein Sci. 2007;Chapter 2:Unit 2.9.

114. Hogle JM, Maeda A, Harrison SC. Structure and assembly of turnip crinkle virus. I. X-ray crystallographic structure analysis at 3.2 A resolution. J Mol Biol. 1986;191:625–38.

115. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol. 1999;287:797–815.

116. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396–404.

117. Wagner LM, Lester JT, Sivrich FL, DeLuca NA. The N terminus and C terminus of herpes simplex virus 1 ICP4 cooperate to activate viral gene expression. J Virol. 2012;86:6862–74.

118. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7:12.

119. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16:404–5.

120. Santos A, Wernersson R, Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res. 2015;43 Database issue:D1140-4.

121. Jadav RS, Chanduri M V, Sengupta S, Bhandari R. Inositol pyrophosphate synthesis by inositol hexakisphosphate kinase 1 is required for homologous recombination repair. J Biol Chem. 2013;288:3312–21.

122. Gu C, Nguyen HN, Ganini D, Chen Z, Jessen HJ, Gu Z, et al. KO of 5-InsP7 kinase activity transforms the HCT116 colon cancer cell line into a hypermetabolic, growth-inhibited phenotype. Proc Natl Acad Sci U S A. 2017;114:11968–73.

123. Barker CJ, Wright J, Hughes PJ, Kirk CJ, Michell RH. Complex changes in cellular inositol phosphate complement accompany transit through the cell cycle. Biochem J. 2004;380 Pt 2:465–73.

124. Posas F, Camps M, Arino J. The PPZ protein phosphatases are important determinants of salt tolerance in yeast cells. J Biol Chem. 1995;270:13036–41.

125. Clotet J, Gari E, Aldea M, Arino J. The yeast ser/thr phosphatases sit4 and ppz1 play opposite roles in regulation of the cell cycle. Mol Cell Biol. 1999;19:2408–15.

126. de Nadal E, Clotet J, Posas F, Serrano R, Gomez N, Arino J. The yeast halotolerance determinant Hal3p is an inhibitory subunit of the Ppz1p Ser/Thr protein phosphatase. Proc Natl Acad Sci U S A. 1998;95:7357–62.

127. Canadell D, Gonzalez A, Casado C, Arino J. Functional interactions between potassium and phosphate homeostasis in Saccharomyces cerevisiae. Mol Microbiol. 2015;95:555–72.

128. Pattison-Granberg J, Persson BL. Regulation of cation-coupled high-affinity phosphate uptake in the yeast Saccharomyces cerevisiae. J Bacteriol. 2000;182:5017–9.

129. Lee YS, Huang K, Quiocho FA, O'Shea EK. Molecular basis of cyclin-CDK-CKI regulation by reversible binding of an inositol pyrophosphate. Nat Chem Biol. 2008;4:25–32.

130. Lee YS, Mulugu S, York JD, O'Shea EK. Regulation of a cyclin-CDK-CDK inhibitor complex by inositol pyrophosphates. Science. 2007;316:109–12.

131. Crocco P, Saiardi A, Wilson MS, Maletta R, Bruni AC, Passarino G, et al. Contribution of polymorphic variation of inositol hexakisphosphate kinase 3 (IP6K3) gene promoter to the susceptibility to late onset Alzheimer's disease. Biochim Biophys Acta. 2016;1862:1766–73.

132. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

133. Chakraborty A, Koldobskiy MA, Bello NT, Maxwell M, Potter JJ, Juluri KR, et al. Inositol pyrophosphates inhibit Akt signaling, thereby regulating insulin sensitivity and weight gain. Cell. 2010;143:897–910.

134. Morrison BH, Bauer JA, Hu J, Grane RW, Ozdemir AM, Chawla-Sarkar M, et al. Inositol hexakisphosphate kinase 2 sensitizes ovarian carcinoma cells to multiple cancer therapeutics. Oncogene. 2002;21:1882–9.

135. Rao F, Cha J, Xu J, Xu R, Vandiver MS, Tyagi R, et al. Inositol pyrophosphates mediate the DNA-PK/ATM-p53 cell death pathway by regulating CK2 phosphorylation of Tti1/Tel2. Mol Cell. 2014;54:119–32.

136. Gobert RP, Joubert L, Curchod ML, Salvat C, Foucault I, Jorand-Lebrun C, et al. Convergent functional genomics of oligodendrocyte differentiation identifies multiple autoinhibitory signaling circuits. Mol Cell Biol. 2009;29:1538–53.

137. Fu C, Xu J, Li RJ, Crawford JA, Khan AB, Ma TM, et al. Inositol Hexakisphosphate Kinase-3 Regulates the Morphology and Synapse Formation of Cerebellar Purkinje Cells via Spectrin/Adducin. J Neurosci. 2015;35:11056–67.

138. Morrison BH, Bauer JA, Kalvakolanu D V, Lindner DJ. Inositol hexakisphosphate kinase 2 mediates growth suppressive and apoptotic effects of interferon-beta in ovarian carcinoma cells. J Biol Chem. 2001;276:24965–70.

139. Nagata E, Luo HR, Saiardi A, Bae BI, Suzuki N, Snyder SH. Inositol hexakisphosphate kinase-2, a physiologic mediator of cell death. J Biol Chem. 2005;280:1634–40.

140. Tan J, Yu CY, Wang ZH, Chen HY, Guan J, Chen YX, et al. Genetic variants in the inositol phosphate metabolism pathway and risk of different types of cancer. Sci Rep. 2015;5:8473.

141. Kay NE, Eckel-Passow JE, Braggio E, Vanwier S, Shanafelt TD, Van Dyke DL, et al. Progressive but previously untreated CLL patients with greater array CGH complexity exhibit a less durable response to chemoimmunotherapy. Cancer Genet Cytogenet. 2010;203:161–8.

142. Chakraborty A, Koldobskiy MA, Sixt KM, Juluri KR, Mustafa AK, Snowman AM, et al. HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. Proc Natl Acad Sci U S A. 2008;105:1134–9.

143. Miyata Y, Nakamoto H, Neckers L. The therapeutic target Hsp90 and cancer hallmarks. Curr Pharm Des. 2013;19:347–65.

144. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

145. Koldobskiy MA, Chakraborty A, Werner Jr JK, Snowman AM, Juluri KR, Vandiver MS, et al. P53-Mediated Apoptosis Requires Inositol Hexakisphosphate Kinase-2. Proc Natl Acad Sci U S A. 2010;107:20947–51.

146. el-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, et al. WAF1, a potential mediator of p53 tumor suppression. Cell. 1993;75:817–25.

147. Miyashita T, Reed JC. Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. Cell. 1995;80:293–9.

148. Sengupta S, Harris CC. p53: traffic cop at the crossroads of DNA repair and recombination. Nat Rev cell Biol. 2005;6:44–55.

149. Amaral JD, Xavier JM, Steer CJ, Rodrigues CM. The role of p53 in apoptosis. Discov Med. 2010;9:145–52.

150. Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, et al. Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. Nucleic Acids Res. 1996;24:141–6.

151. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian S V, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat. 2007;28:622–9.

152. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2:a001008.

153. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

154. Chene P, Fuchs J, Carena I, Furet P, Garcia-Echeverria C. Study of the cytotoxic effect of a peptidic inhibitor of the p53-hdm2 interaction in tumor cells. FEBS Lett. 2002;529:293–7.

155. Aoki M, Sobek V, Maslyar DJ, Hecht A, Vogt PK. Oncogenic transformation by beta-catenin: deletion analysis and characterization of selected target genes. Oncogene. 2002;21:6983–91.

156. Gu C, Wilson MS, Jessen HJ, Saiardi A, Shears SB. Inositol Pyrophosphate Profiling of Two HCT116 Cell Lines Uncovers Variation in InsP8 Levels. PLoS One. 2016;11:e0165286.

157. Jadav RS, Kumar D, Buwa N, Ganguli S, Thampatty SR, Balasubramanian N, et al. Deletion of inositol hexakisphosphate kinase 1 (IP6K1) reduces cell migration and invasion, conferring protection from aerodigestive tract carcinoma in mice. Cell Signal. 2016;28:1124–36.

158. Ghoshal S, Zhu Q, Asteian A, Lin H, Xu H, Ernst G, et al. TNP [N2-(m-Trifluorobenzyl), N6-(p-nitrobenzyl)purine] ameliorates diet induced obesity and insulin resistance via inhibition of the IP6K1 pathway. Mol Metab. 2016;5:903–17.

159. Luo HR, Huang YE, Chen JC, Saiardi A, Iijima M, Ye K, et al. Inositol pyrophosphates mediate chemotaxis in Dictyostelium via pleckstrin homology domain-PtdIns(3,4,5)P3 interactions. Cell. 2003;114:559–72.

160. Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt stimulates aerobic glycolysis in cancer cells. Cancer Res. 2004;64:3892–9.

161. Di Pietro E, Sirois J, Tremblay ML, MacKenzie RE. Mitochondrial NAD-dependent methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase is essential for embryonic development. Mol Cell Biol. 2002;22:4158–66.

162. Vazquez A, Markert EK, Oltvai ZN. Serine biosynthesis with one carbon catabolism and the glycine cleavage system represents a novel pathway for ATP generation. PLoS One. 2011;6:e25881.

163. Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev. 2013;13:572–83.

164. Yuno A, Lee MJ, Lee S, Tomita Y, Rekhtman D, Moore B, et al. Clinical Evaluation and Biomarker Profiling of Hsp90 Inhibitors. Methods Mol Biol. 2018;1709:423–41.