EVALUATION OF THE CONSEQUENCES OF TOPOLOGICAL AND FLUX CONSTRAINTS IN SIMULATING YEAST METABOLISM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Hnin Wuttyi Aung

May 2015

EVALUATION OF THE CONSEQUENCES OF TOPOLOGICAL AND FLUX CONSTRAINTS IN SIMULATING YEAST METABOLISM

Hnin Wuttyi Aung, Ph.D.

Cornell University 2015

This dissertation presents an evaluation of the reconstruction of *Saccharomyces cerevisiae*'s metabolic network in the Yeast consensus model. A particular focus in this curation was fatty acid, glycerophospholipid, and neutral glycerolipid metabolism. By comparing the information contained within the Yeast v6.0 model to that in literature, I found inaccuracies which included incorrect gene-reaction associations, improper definition of substrates/products in reactions, missing connections between metabolites, inappropriate assignments of reaction directionality, and nonfunctional β-oxidation pathways. Changes to address these issues were implemented to generate Yeast v7.0.

To understand the impact of the changes made throughout the evolution of the Yeast consensus model, I examined differences in the maximum pathway yield of eight industrially relevant metabolites and six lipid species that were calculated by the iMM904 model, a progenitor to the consensus models, and the three latest major version releases of the consensus model at that time: Yeast v5.01, v6.06, and v7.11. The factors that led to differences in the calculated maximum pathway yield were the assumed proton stoichiometry for the electron transport chain and ATP synthase; the assumed cofactor specificity of enzymes; the inclusion/exclusion of reactions involved in NADH/NADPH homeostasis; assumed reversibility of reactions; and assumed transport reactions.

While the reconstruction of the metabolic network represents the inherent capacity to process substrate into different metabolites and biomass, I also examined how experimental data on concentration of biomass and extracellular metabolites over time for batch cultivation can be used to constrain the input/output to the metabolic network. This data was first processed by determining models to describe concentration over time. Specific rates for growth, glucose consumption, and formation of ethanol, glycerol, and acetate were then calculated from these models. These specific rates were used to constrain fluxes for growth and exchange reactions in the Yeast v7.5 model. The use of lower and upper bounds to the fluxes, instead of equalities, were found to be useful to avoid over-constraining the linear programming problem with incompatible specific rates and to also represent the uncertainty in determining accurate specific rates.

BIOGRAPHICAL SKETCH

Hnin Aung was born May 30, 1987 in Rangoon, Burma and lived there for three years. Growing up, she lived in several different locations: Birmingham, England; Yonkers, NY; Cheboygan, MI; and Visalia, CA. She graduated from Golden West High School in Visalia, CA in 2004. Hnin's interest in the application of living organisms to address societal needs and demands led her to study Biological Systems Engineering at the University of California, Davis, where she earned a Bachelor of Science in 2008. Following a research rotation in her first year of graduate school at Cornell University, she chose to pursue a Doctorate of Philosophy under the guidance of Dr. Larry Walker. During her time at Cornell University, Hnin has served as treasurer and vice president of the Biological and Environmental Engineering Graduate Student Association, developed teaching tools for the Bioenergy & Bioproducts Education Program, and instructed rock climbing courses for Cornell Outdoor Education.

This thesis is dedicated to my friends and family.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Larry Walker for all the support he has provided me throughout my graduate career. His mentorship has helped foster my growth and development as a researcher and as a person figuring out her passions in life. His unwavering belief in my abilities and potential has meant a lot to me, especially during my self-critical moments. It is truly an honor to be his last Ph.D. student.

Sincerest thanks go to my other committee members: Dr. Susan Henry and Dr. Ruth Richardson. I appreciate the insight and advice they have given me on my research. Dr. Henry's expertise in lipid metabolism has been crucial in my modeling efforts.

I would like to thank the members of the Walker group for their advice, support, encouragement, and friendship. I am also grateful to the members of the Henry group. I am grateful to Manuel Villa for the training on yeast molecular genetics and to Laura Gaspar and Steve Jesch for answering questions that I had on lipid metabolism and for sharing their own insights.

Outside of the lab, I am grateful for the friends I have made through hiking, climbing, and soccer. I will continue my love for the outdoors and adventure beyond my time in Ithaca.

Finally, I would like to thank my family for the support they have provided me throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

Evidence for the use of yeast by mankind to produce desirable products dates back over 7000 years ago (Mortimer 2000). The ability of the yeast *Saccharomyces cerevisiae* to ferment sugar into ethanol and carbon dioxide has long been harnessed for production of alcoholic beverages and bread. Beyond this traditional role in biotechnology, *S. cerevisiae* has been used for the production of other industrially relevant biochemicals, such as glycerol, propanediol, organic acids, sugar alcohols, L-glycerol-3-phosphate, steroids, and isoprenoids (Nevoigt 2008). These biochemicals span a diverse range of applications that include cosmetics, pharmaceuticals, plastics, and food additives.

S. cerevisiae has multiple advantages for use in industrial biotechnology. It is classified as GRAS (Generally Regarded As Safe) by the US Food and Drug Administration; can tolerate low pH and high sugar and ethanol concentrations, which can allow for the use of cultivation conditions more inhibitive to the growth of other microorganisms; and has a well-established history of large-scale production (Ostergaard, Olsson et al. 2000; Nevoigt 2008). Furthermore, biotechnology tools developed for *S. cerevisiae* provide the opportunity to engineer cells to utilize a diverse set of substrates, to expand the set of metabolites that can be produced, and to improve titer, productivity, and yield of desired products (Ostergaard, Olsson et al. 2000). Due to the high efficiency of homologous recombination in *S. cerevisiae*, transformed DNA can be readily integrated into its genome to introduce foreign genes, disrupt genes, or modify promoters to adjust gene expression (Duina, Miller et al. 2014). Multiple strains of *S. cerevisiae*, including commercial, laboratory, and wild isolate strains, have been sequenced (Engel and Cherry 2013). Analysis of the variations in the genomes of these strains could allow for understanding of the

divergence in phenotypes. Within the gene content of its genome, a collection of over 21,000 mutants containing deletions of about 6,000 open reading frames has been created and used to understand individual gene function, interactions between different genes, and the role of genes in responding to environmental conditions (Giaever and Nislow 2014).

Rational metabolic engineering based on direct knowledge of pathways and genes has had success in certain applications. For example, Runguphan and Keasling (2014) were able to increase the lipid content to more than 17% dry cell weight, a 300% increase over the control *S. cerevisiae* strain, by overexpressing multiple genes responsible for *de novo* fatty acid and triacylglycerol synthesis: *ACC1*, *FAS1*/*FAS2*, and *DGA1*. They found that overexpression of each gene singularly led to much smaller increases in lipid content of 58%, 30%, and 150%, respectively. This observation is in line with the need to adopt a metabolic engineering approach beyond consideration of an individual rate-limiting reaction and towards an appreciation that "flux is a systemic property and questions of its control cannot be answered by looking at one step in isolation – or even each step in isolation" (Kacser and Burns 1995). In addition to a holistic attention to the pathway of interest, competing pathways which may redirect flux away from the desired metabolite should be accounted for (Tepper and Shlomi 2010). One tool available to facilitate a more complete comprehension of the interactions within the entire metabolic network is a genome-scale metabolic model.

The initial goal of my research was to utilize a genome-scale metabolic model for *S. cerevisiae* to suggest genetic engineering strategies to drive flux into triacylglycerol. In the interest of reaching this goal, I manually evaluated the representation of fatty acid, glycerophospholipid, and neutral glycerolipid metabolism in the Yeast v6.0 consensus model (Chapter 3). This evaluation revealed inaccuracies and missing information in the reconstruction

of these pathways in the model and led to an updated version of the model, Yeast v7.0. Given the incremental modifications made throughout the history of the Yeast consensus model, I addressed how differences in the reconstruction of the metabolic network alter simulated pathway yields of various lipids and industrially relevant metabolites (Chapter 4). Beyond the generic conditions included in the Yeast model, I examined how experimental data can be used to further constrain the predicted behavior in batch fermentation (Chapter 5). Although my research interest appears to have shifted focus towards a fundamental understanding of the constraints contained within the model, the choice and implementation of these constraints are important to outcomes predicted by the model and can ultimately affect suggestions in metabolic engineering strategies.

REFERENCES

Duina, A. A., M. E. Miller and J. B. Keeney (2014). "Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces cerevisiae Model System." Genetics **197**(1): 33-48.

Engel, S. R. and J. M. Cherry (2013). "The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database." Database **2013**.

Giaever, G. and C. Nislow (2014). "The Yeast Deletion Collection: A Decade of Functional Genomics." Genetics **197**(2): 451-465.

Kacser, H. and J. Burns (1995). "The control of flux." Biochemical Society Transactions **23**(2): 341-366.

Mortimer, R. K. (2000). "Evolution and Variation of the Yeast (Saccharomyces) Genome." Genome Research **10**(4): 403-409.

Nevoigt, E. (2008). "Progress in Metabolic Engineering of Saccharomyces cerevisiae." Microbiology and Molecular Biology Reviews **72**(3): 379-412.

Ostergaard, S., L. Olsson and J. Nielsen (2000). "Metabolic Engineering of Saccharomyces cerevisiae." Microbiology and Molecular Biology Reviews **64**(1): 34-50.

Runguphan, W. and J. D. Keasling (2014). "Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals." Metabolic Engineering **21**(0): 103-113.

Tepper, N. and T. Shlomi (2010). "Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways." Bioinformatics **26**(4): 536-543.

Chapter 2: Literature Review

2.1 Triacylglycerol and Phospholipid Metabolism in *Saccharomyces cerevisiae*

2.1.1 Relationship between Triacylglycerol and Phospholipid Metabolism

The major membrane phospholipids and triacylglycerol (TAG) share common metabolic intermediates: phosphatidic acid (PA) and diacylglycerol (DAG). PA can be activated with cytidine triphosphate (CTP) to form cytidine diphosphate-diacylglycerol (CDP-DAG), which can be used for synthesis of all the major membrane phospholipids. PA can also be dephosphorylated to form DAG, which is a direct precursor of TAG and can be used for synthesis of the phospholipids phosphatidylethanolamine and phosphatidylcholine through the Kennedy Pathway (see Figure 2.1 and Section 2.1.2 for more details).

In addition to being a central hub between TAG and phospholipid metabolism, PA plays a role in the transcriptional control of genes involved in phospholipid biosynthesis. PA in the endoplasmic reticulum (ER) membrane can bind to the Opi1p repressor, thus restricting the translocation of Opi1p to the nucleus where it can repress specific phospholipid biosynthetic genes (Loewen, Gaspar et al. 2004). Genetic modifications affecting PA metabolism have been shown to influence the levels of TAG and phospholipids. Mutants deleted of the *PAH1* gene, whose gene product dephosphorylates PA to yield DAG, accumulated PA and had reduced levels of DAG and TAG (Han, Wu et al. 2006). These mutants also exhibited pronounced nuclear membrane proliferation, which was attributed to increased phospholipid synthesis through transcriptional upregulation (Santos-Rosa, Leung et al. 2005).

It is hypothesized that "the cell coordinates the synthesis and breakdown of storage lipids [i.e., TAG and steryl esters] with its demand for membrane lipid synthesis" (Gaspar, Hofbauer et

al. 2011). The connection between TAG and phospholipids is exemplified in a study where phospholipid synthesis was decreased following a block in membrane trafficking from the ER and a concomitant increase in TAG synthesis was observed (Gaspar, Jesch et al. 2008). This divergence is assumed to channel excess lipid intermediates arising from reduced phospholipid synthesis into TAG production.

TAG can be degraded to provide fatty acids for generation of metabolic energy via β-oxidation and/or to provide DAG and fatty acid for phospholipid synthesis (Taylor and Parks 1979; Gray, Petsko et al. 2004; Kurat, Natter et al. 2006; Zanghellini, Natter et al. 2008; Gaspar, Hofbauer et al. 2011). It should be noted that β-oxidation occurs only in the peroxisome for *S. cerevisiae* (Hiltunen, Mursula et al. 2003) and that peroxisomes are repressed in the presence of glucose (Veenhuis, Mateblowski et al. 1987; Skoneczny and Rytka 1996). Thus, β-oxidation of fatty acids from TAG is not significant for *S. cerevisiae* in media containing glucose. Rather, in this situation, it is assumed that TAG degradation serves the purpose of providing building blocks for membrane lipid synthesis. The role of β-oxidation in glucose cultures emerges during the postdiauxic and stationary phase when both the glucose supply is exhausted and the ethanol produced from fermentation has been consumed (Lefevre, van Roermund et al. 2013).

The importance of TAG to growth is evident with mutants unable to utilize or produce intracellular pools of TAG. For instance, exponentially growing wild type yeast cells deprived of inositol exhibit a burst in phosphatidylinositol (PI) synthesis when supplemented with its precursor inositol (Kelley, Bailis et al. 1988; Gaspar, Aregullin et al. 2006). However, mutant strains unable to synthesize or degrade TAG display a significantly reduced rate of PI synthesis and reduced levels of PI (Gaspar, Hofbauer et al. 2011). This suggests that TAG degradation plays a role in achieving rapid synthesis of PI. Similarly, when stationary phase yeast cells are

transferred to fresh media, degradation of TAG occurs to support rapid growth (Taylor and Parks 1979; Kurat, Natter et al. 2006; Fakas, Konstantinou et al. 2011). Mutants defective in TAG degradation exhibited diminished initial growth upon transfer to fresh media (Kurat, Natter et al. 2006; Fakas, Konstantinou et al. 2011).

In general, the TAG content of yeast cells is low during early exponential growth but becomes substantial by late exponential growth phase, reaching the highest level during stationary phase (Taylor and Parks 1979; Kohei and Satoshi 1984; Kurat, Natter et al. 2006; Czabany, Wagner et al. 2008). As discussed in the previous paragraph, TAG is quickly broken down to supply lipid precursors for rapid initial growth. As exponential growth proceeds, the rate of TAG synthesis exceeds the rate of degradation and the level of TAG increases. TAG accumulates in late exponential growth and stationary phase as net phospholipid synthesis diminishes.

2.1.2 Metabolic Pathways for Synthesis of Triacylglycerol and Phospholipids

A visual overview of the metabolic pathways for synthesis of triacylglycerol and phospholipids is shown in Figure 2.1.

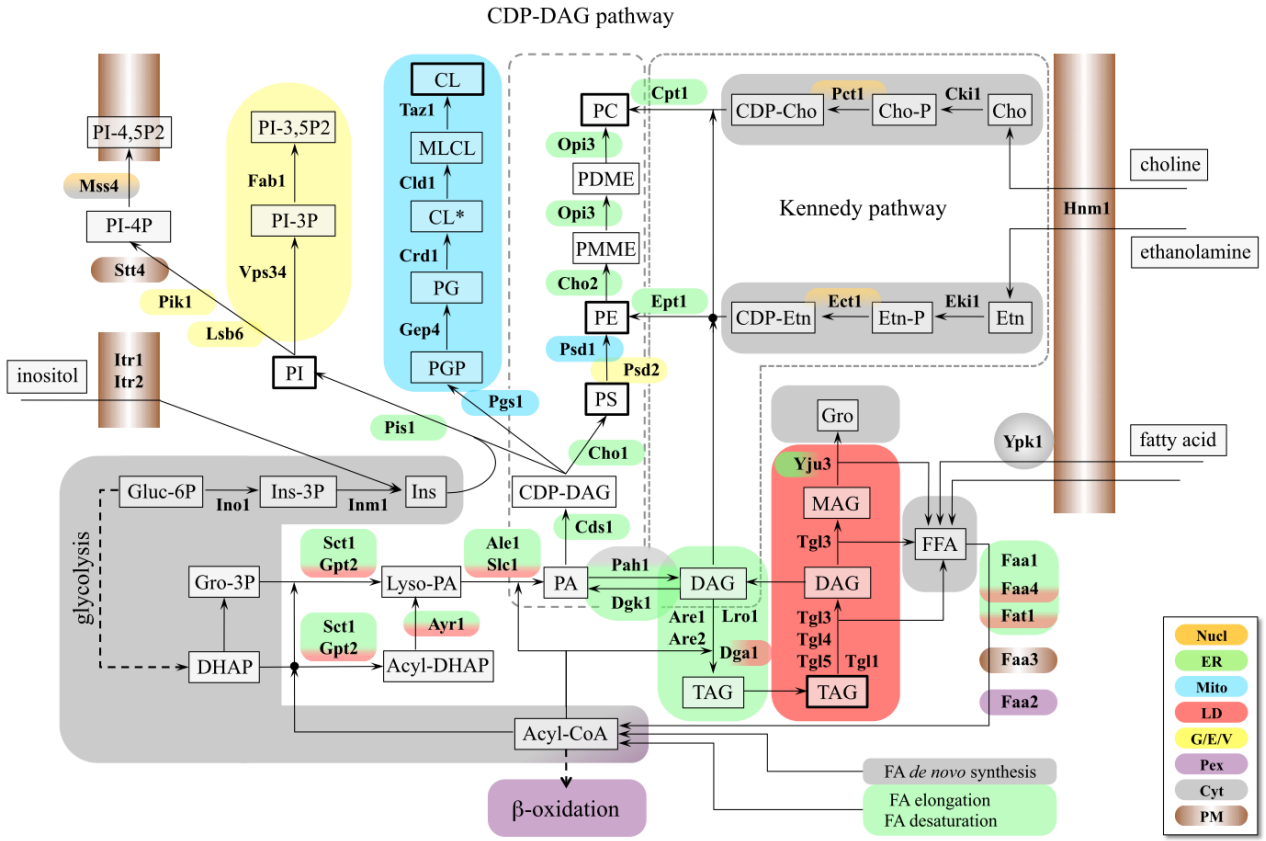


Figure 2.1: Pathways for the synthesis of glycerolipids and their subcellular localization. Figure from (Henry, Kohlwein et al. 2012). Abbreviations: TAG, triacylglycerols; PI, phosphatidylinositol; PA, phosphatidic acid; CDP-DAG, CDP-diacylglycerol; DAG, diacylglycerol; MAG, monoacylglycerol; Gro, glycerol; DHAP, dihydroxyacetone phosphate; PS, phosphatidylserine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PGP, phosphatidylglycerol phosphate; CL*, precursor cardiolipin; MLCL, monolyso-cardiolipin; CL, mature cardiolipin; PMME, phosphatidylmonomethylethanolamine; PDME, phosphatidyl-dimethylethanolamine; PC, phosphatidylcholine; FFA, free fatty acids; Cho, choline; Etn, ethanolamine; Ins, inositol; Cho-P, choline phosphate; CDP-Cho, CDP-choline; Etn-P, ethanolamine phosphate; CDP-Etn, CDP-ethanolamine; PI 3-P, phosphatidylinositol 3-phosphate; PI 4-P, phosphatidylinositol 4-phosphate; PI 4,5-P_2, phosphatidylinositol 4,5-bisphosphate; PI 3,5-P_2, phosphatidylinositol 3,5-bisphosphate. Nucl, nucleus; ER, endoplasmic reticulum; Mito, mitochondria; LD, lipid droplets; G/E/V, Golgi, endosomes, vacuole; Pex, peroxisomes; Cyt, cytoplasm; PM, plasma membrane.

De novo synthesis of PA is accomplished by sequential acylation of the *sn-1* and *sn-2* positions of glycerol 3-phosphate. The first acylation at the *sn-1* position of glycerol 3-phosphate is catalyzed by either Sct1p at the ER membrane or Gpt2p at the ER membrane and lipid particle (Zheng and Zou 2001; Kim, Melénn et al. 2006). This acylation yields lyso-PA. These two acyltransferases can also acylate dihydroxyacetone phosphate (DHAP) to produce acyl-DHAP (Zheng and Zou 2001); acyl-DHAP can then be reduced to lyso-PA by the NADPH-dependent reductase, Ayr1p, at the ER membrane and lipid particle (Athenstaedt and Daum 2000). Sct1p displays a preference for the 16 carbon acyl-CoAs palmitoyl-CoA (16:0) and palmitoleoyl-CoA (16:1), while Gpt2p displays a broader substrate specifity, utilizing the aforementioned acyl-CoAs and the 18 carbon acyl-CoAs stearoyl-CoA (18:0) and oleoyl-CoA (18:1) (Zheng and Zou 2001). Disruption of either *SCT1* or *GPT2* leads to reduced glycerol 3-phosphate acyltransferase activity, whereas disruption of both genes concurrently is lethal (Zheng and Zou 2001).

The second acylation step, which converts lyso-PA to PA, is catalyzed by either Ale1p at the ER membrane or Slc1p at the ER membrane and lipid particle (Athenstaedt and Daum 1997; Kim, Melénn et al. 2006; Benghezal, Roubaty et al. 2007; Tamaki, Shimada et al. 2007). Similar to the glycerol 3-phosphate acyltransferases, cells with single deletion of either *ALE1* or *SLC1* are still viable but simultaneous deletion of both genes is lethal (Benghezal, Roubaty et al. 2007; Tamaki, Shimada et al. 2007). *slc1*Δ cells exhibit decreased content of C18:1 fatty acids in most phospholipids and lyso-phospholipids which suggests that C18:1 acyl-CoAs are the major substrate for Slc1p (Shui, Guan et al. 2010). Ale1p displays a preference for unsaturated fatty acids of 16- and 18-carbon chain length (Tamaki, Shimada et al. 2007).

PA can be used to synthesize various phospholipids through the intermediate metabolite CDP-DAG. The gene *CDS1* encodes for CDP-diacylglycerol synthase activity in ER and

mitochondrial membranes (Shen, Heacock et al. 1996). This enzyme activates PA in a CTP-dependent reaction to produce CDP-DAG. The phosphatidyl moiety of CDP-DAG can then be transferred to different metabolites to yield different phospholipids. At the mitochondrial membrane, phosphatidylglycerolphosphate synthase, encoded by *PGS1*, transfers the phosphatidyl moiety of CDP-DAG onto glycerol 3-phosphate to form phosphatidylglycerol phosphate (PGP) (Chang, Heacock et al. 1998). PGP can then be utilized to form mature cardiolipin through a series of reactions (Osman, Voelker et al. 2011). At the ER membrane, the phosphatidyl moiety of CDP-DAG can be transferred to inositol to form phosphatidylinositol (PI) via PI synthase Pis1p (Nikawa and Yamashita 1984; Nikawa, Kodaki et al. 1987) or transferred to L-serine to form phosphatidylserine (PS) via PS synthase Cho1p (Letts, Klig et al. 1983; Bae-Lee and Carman 1984; Kiyono, Miura et al. 1987; Nikawa, Tsukagoshi et al. 1987). PI can serve as a precursor for polyphosphoinositides (Strahl and Thorner 2007), sphingolipids (Cowart and Obeid 2007; Dickson 2008; Dickson, Chalfant et al. 2010), and glycosylphosphatidylinositol anchors (Pittet and Conzelmann 2007). PS can be used to synthesize phosphatidylethanolamine (PE) and phosphatidylcholine (PC) through the CDP-DAG pathway. PS can be decarboxylated to form PE by the PS decarboxylases Psd1p at the inner mitochondrial membrane (Clancey, Chang et al. 1993) and Psd2p at the Golgi/vacuolar membrane (Trotter and Voelker 1995). PE can then be converted to PC through three S-adenosyl-L-methionine-dependent methylation reactions at the ER membrane. The first methylation step is catalyzed by the PE methyltransferase Cho2p and the final two methylation steps are catalyzed by the phospholipid methyltransferase Opi3p (Kodaki and Yamashita 1987; Summers, Letts et al. 1988; Kodaki and Yamashita 1989); Opi3p can also catalyze the first

methylation step, however this is a minor contribution *in vivo* (Kodaki and Yamashita 1987; Kodaki and Yamashita 1989).

PA can also be used to synthesize TAG and the phospholipids PE and PC through the intermediate metabolite diacylglycerol (DAG). Dephosphorylation of PA to DAG can be catalyzed by Mg^{2+}-dependent PA phosphatase Pah1p at the ER membrane (Han, Wu et al. 2006; Karanasios, Han et al. 2010), diacylglycerol pyrophosphate phosphatase Dpp1p at the vacuolar membrane (Wu, Liu et al. 1996; Han, Johnston et al. 2001), and lipid phosphate phosphatase Lpp1p at the Golgi membrane (Toke, Bennett et al. 1998; Huh, Falvo et al. 2003). However, it is assumed that the DAG utilized in synthesis of TAG, PE, and PC largely arises from Pah1p activity (Han, Wu et al. 2006). Consistent with this assumption is the finding that the DAG and TAG content of the *pah1*Δ mutant is greatly lower than wild-type while the content for the *dpp1*Δ *lpp1*Δ mutant is similar to wild-type (Han, Wu et al. 2006).

The DAG produced from PA can be used to synthesize PE and PC through the Kennedy pathway. The reaction with DAG and CDP-ethanolamine to form PE is catalyzed by Ept1p while the reaction with DAG and CDP-choline to form PC is catalyzed by Cpt1p or Ept1p (Hjelmstad and Bell 1987; Hjelmstad and Bell 1988; Hjelmstad and Bell 1991). The choline and ethanolamine component of these reactions can be derived from either exogenous choline and ethanolamine or turnover of PC and sphingolipids (Henry, Kohlwein et al. 2012).

DAG can serve as a direct precursor of TAG via transfer of an acyl group from either acyl-CoA or phospholipids. The four enzymes that account for TAG synthesis are Dga1p, Lro1p, Are1p, and Are2p. Dga1p catalyzes the majority of the acyl-CoA-dependent DAG acyltransferase activity while Are1p and Are2p provide only minor activity (Oelkers, Cromley et al. 2002; Sandager, Gustavsson et al. 2002; Sorger and Daum 2002). Lro1p catalyzes the transfer

of the acyl group at the *sn-2* position of PC or PE to DAG (Dahlqvist, Ståhl et al. 2000; Oelkers, Tinkelenberg et al. 2000). It has been hypothesized that Lro1p is largely responsible for TAG synthesis during exponential growth and that Dga1p is largely responsible for TAG synthesis during stationary phase (Oelkers, Tinkelenberg et al. 2000; Oelkers, Cromley et al. 2002). However, a *dga1Δare1Δare2Δ* mutant with Lro1p active exhibited lower TAG amounts in late exponential phase compared to wild-type and this mutant's TAG levels increased from late-exponential growth to mid-stationary phase (Czabany, Wagner et al. 2008). Additionally, a *lro1Δare1Δare2Δ* mutant with Dga1p active exhibited TAG levels similar to wild type in both late exponential and mid-stationary phase (Czabany, Wagner et al. 2008). These observations qualify the aforementioned hypothesis of the role of the individual DAG acyltransferases throughout growth phases.

2.2 Modeling of Metabolism

Metabolic models help to structure our knowledge of metabolism. These models are crucial to characterize systems and to offer predictive insights. The simplest approach to modeling cell metabolism is to treat the cell as a black box. This approach considers material flows of substrate into biomass and products without direct consideration of biochemical processes. These models allow for calculation of theoretical biomass and product yields for a given substrate; however, the fundamental mechanism for obtaining these yields is not able to be discerned due to the restricted scope of the model. In contrast, the most detailed approach to modeling cell metabolism is to characterize the mechanism of each enzymatic reaction through kinetic modeling. This approach allows for determination of reaction rates through consideration of enzyme maximum reaction velocity and binding affinities for substrates and inhibitors; however, cell-scale kinetic modeling is a great challenge due to the considerable number of

kinetic parameters needed to characterize the system and the difficulties in determining kinetic constants consistent with *in vivo* conditions (Jamshidi and Palsson 2008).

An intermediate approach is to use a constraint-based network model of cell metabolism. This approach examines how input nutrients are processed through the reactions of the metabolic network to collectively produce a phenotype. The basis of this model is a metabolic genome-scale reconstruction or metabolic GENRE. Metabolic GENREs are built using information from an organism's annotated genome and, correspondingly, information on reactions catalyzed by the set of metabolic enzymes encoded by the genome. Metabolic GENREs have currently been used for "(1) contextualization of high-throughput data, (2) guidance of metabolic engineering, (3) directing hypothesis-driven discovery, (4) interrogation of multi-species relationships, and (5) network property discovery" (Oberhardt, Palsson et al. 2009).

2.2.1 Genome-Scale Reconstruction (GENRE)

Initial reconstruction of the organism-specific metabolic network is created using coupled information from genome annotation and online databases. Table 2.1 summarizes the information contained in various databases used in the reconstructions of *S. cerevisiae*'s metabolic network.

Table 2.1: Databases used to reconstruct *S. cerevisiae*'s metabolic network

Database	Website	Description
Braunschweig Enzyme Database (BRENDA)	http://www.brenda-enzymes.org/	Collection of enzyme information including substrate/products, cofactors, functional and kinetic parameters, and localization
Chemical Entities of Biological Interest (ChEBI)	http://www.ebi.ac.uk/chebi/	Dictionary of molecular entities
Comprehensive Yeast Genome Database (CYGD)	http://mips.helmholtz-muenchen.de/genre/proj/yeast/	Collection of genetic, biochemical, and cell biological information for all individual elements of *S. cerevisiae*'s genomic structure
ENZYME	http://enzyme.expasy.org/	Repository of information relative to the nomenclature of enzymes
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/	Integrated resource consisting of 17 databases related to genomic, pathway, and chemical information
MetaCyc	http://metacyc.org/	Metabolic encyclopedia containing information on metabolic pathways, enzymatic reactions, enzymes, genes, and chemicals
PubChem	http://pubchem.ncbi.nlm.nih.gov/	Database of chemical structures and biological activities of small organic molecules
Reactome	www.reactome.org	Knowledgebase of biological pathways covering metabolism, signal transduction, DNA replication, cell cycle regulation, and other biologic processes
Saccharomyces Genome Database (SGD)	www.yeastgenome.org/	Encyclopedic information about the genome, genes and their encoded proteins, and other chromosomal features of *S. cerevisiae*
Universal Protein Resource (UniProt)	http://www.uniprot.org/	Database of protein sequences with annotations and functional information
Yeast Protein Database (YPD)	http://www.proteome.com/YPDhome.html	Database for *S. cerevisiae*'s proteome

The process of creating an initial reconstruction has been automated through tools such as SimPheny (Genomatica, San Diego, CA), PathoLogic (Karp, Paley et al. 2002), Model SEED (Henry, DeJongh et al. 2010) and SuBliMinaL Toolbox (Swainston, Smallbone et al. 2011). Despite the capacity and value of these automated tools, the generated reconstructions are

typically incomplete and contain errors due to limitations, misinformation, and knowledge gaps in the database sources. This can generally be ascribed to one or more of the items listed below:

- *Missing reactions*

  Reactions may be missing from the initial reconstruction because the gene encoding a reaction's catalytic enzyme is unknown and thus the reaction is not included in the database. Transport reactions based on diffusion or without known transporters are also missing since there are no corresponding genes and proteins for these reactions. In addition, the mapping of genes to specific reactions can be complicated because annotations may not include full Enzyme Commission numbers or may just be textual or ontology-based (Durot, Bourguignon et al. 2009).

- *Generic or synonymous terms*

  Certain reactions may contain generic metabolites, such as 'an aldehyde' or 'long-chain-acyl-CoA', which requires interpretation to be linked to specific metabolites. Given the various possibilities for these generic metabolites, the reconstruction must reflect the actual specificity of the organism-specific enzyme.

  Certain compounds may have multiple entries in the same database. For example, KEGG associates four separate compound ID's for glucose: 'glucose', 'D-glucose', 'alpha-D-glucose', and 'beta-D-glucose'. Poolman, Bonde et al. (2006) raised the concern that such entries could fragment the model if the equivalency or relationship between the distinct entries was not recognized.

- *Incorrect Gene-Protein-Reaction (GPR) associations*

  Incorrect GPR assignments can arise from misannotation of the genome. "Genome annotation is a multilevel process, and errors can emerge at different stages:

during sequencing, as a result of gene-calling procedures, and in the process of assigning gene functions." (Poptsova and Gogarten 2010). During their reconstruction of the *Pseudomonas putida* metabolic network, Puchałka, Oberhardt et al. (2008) found several instances of improperly annotated genes in the databases of Pseudomonas Genome Database, KEGG, and NCBI. These misannotations occurred due to lack of information at the time of the original annotation or overlooked literature evidence.

Incorrect GPR assignments can also arise due to "lack of defined multimeric enzymatic complexes in KEGG and MetaCyc" (Swainston, Smallbone et al. 2011). In this respect, the genes associated with the component subunits are each assigned the reaction(s) catalyzed by the complex. This association yields OR relationships, instead of the appropriate AND relationships.

- *Uncertain or erroneous reaction information*

The following may not be known for a reaction or may not be correct in a database: cofactor specificity, reversibility, or subcellular localization. In addition, these traits may differ between different organisms which may not be reflected in databases.

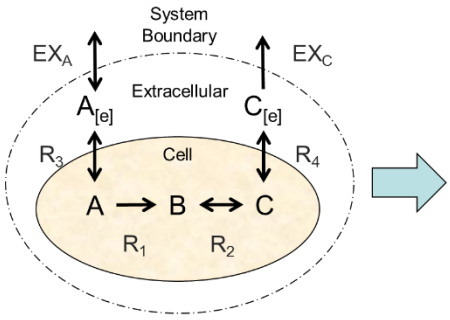
To resolve these issues, the reconstruction must undergo manual curation in which one refers to primary literature, review articles, and/or organism-specific textbooks. These resources can offer organism-specific information on reaction details, indicate what reactions are known to occur but whose GPR associations are unknown, and help identify reactions that should not be in the reconstruction due to contradiction with literature evidence.

2.2.2 Genome-Scale Model (GEM)

Metabolic GENREs can be viewed as databases of verified and putative biochemical reactions occurring in a specific organism. However, just knowing the collection of utilizable

reactions does not directly lead one to understand how the cell operates. In order to understand the coordination of various reactions to produce biological function, the reconstruction is then converted into a genome-scale model (GEM) which is amenable to computational methods.

The information contained in the reconstruction is condensed into a stoichiometric matrix (S) which describes the stoichiometry of the reactants and products involved in each reaction of the cell. The columns of S represent the reactions and the rows represent the metabolites in the metabolic network. For each reaction in the network, the flux or reaction rate is constrained to either a particular value or to a range of values in order to limit the possible flux distributions in the metabolic network. These flux constraints are chosen based on "thermodynamic and capacity constraints or experimental data" (Covert, Schilling et al. 2001). Thermodynamic constraints refer to whether a reaction is reversible or irreversible. Capacity constraints refer to the maximum achievable reaction rate, such as the maximum uptake rate for a transporter. For reactions with unknown reversibility or capacity, a lower bound and upper bound of arbitrarily large magnitude, such as [-1000, 1000] $\frac{mmol}{gdW-hr}$, is assigned so as to not impose explicit constraints on the possible flux values. A simplified example of this process is shown in Figure 2.2.



Stoichiometric Matrix (S)

Reactions

Metabolites	R1	R2	R3	R4	EXA	EXC
A	-1	0	1	0	0	0
B	1	-1	0	0	0	0
C	0	1	0	-1	0	0
A[e]	0	0	-1	0	-1	0
C[e]	0	0	0	1	0	-1

Constraints to Reaction Rates
$0 \le V_{R1} \le 10^3$
$-10^3 \le V_{R2} \le 10^3$
$-10^3 \le V_{R3} \le 10^3$
$-10^3 \le V_{R4} \le 10^3$
$-1 \le V_{EXA} \le 10^3$
$0 \le V_{EXC} \le 10^3$

Figure 2.2: Simplified illustration of constraint based modeling. In this example, R_1 and R_2 represent irreversible and reversible metabolic reactions, respectively; R_3 and R_4 represent transport reactions that uptake or excrete metabolites A and C to the extracellular environment; and Ex_A and Ex_C represents source/sink reactions for metabolites A and C.

In addition to the reactions already contained in the reconstruction, the metabolic GEM includes the following pseudo-reactions: exchange reactions and a biomass equation. These reactions do not reflect actual biochemical reactions, but are rather modeling artifacts. Exchange reactions are source/sink reactions that allow metabolites to be supplied or removed from the system. These exchange reactions ultimately allow for uptake and excretion of metabolites into the media. Different media and oxygen environments can thus be represented through different flux bounds of the exchange reactions.

The biomass equation represents a drain of metabolites needed to produce 1 gram dry weight of biomass. This equation thus allows for simulation of cell growth in the model. The metabolites drained are the precursors of vital cellular components (e.g., proteins, DNA, RNA, carbohydrates, and lipids). In addition to these precursors, the biomass equation also includes ATP needed for growth associated maintenance. The non-growth associated maintenance is represented through a separate ATP hydrolysis reaction. The growth and non-growth associated maintenance represent the energy costs beyond that already accounted for in the model's metabolic reactions, such as costs for maintenance of concentration and electrical gradients, futile cycles, and turnover of macromolecules (Stephanopoulos, Aristidou et al. 1998). The approach to creating a biomass equation and determining maintenance cost has been described in (Feist, Herrgard et al. 2009; Thiele and Palsson 2010).

A key concern that arises is that the biomass composition is not constant for a microorganism and can differ based on environmental conditions and mutations. To address this concern, several studies have conducted sensitivity analysis to determine the influence of biomass composition on model predictions. The details on how the model predictions were generated will be discussed in the next section. Varma and Palsson (1993) found that the biomass

yield was insensitive to changes in stoichiometry of biomass precursors. For instance, increasing the stoichiometry of ATP, the most influential metabolite in their biomass equation, by 10% would lead to only a 2% decrease in predicted biomass yield. Feist, Henry et al. (2007) studied how different biomass equations with varying protein content from 50-80 wt%, RNA content from 10-25 wt%, and lipid content from 7-15 wt% affected growth rate and oxygen uptake rate. While they noticed an effect due to the changing stoichiometries, the variance stayed below 5% for growth rate and 8% for oxygen uptake rate. In contrast to the relative insensitivity to protein, RNA, and lipid content, varying the growth associated maintenance cost from 59.81 ± 50% mmol ATP gdW^{-1} affected the growth rate by 16% and the oxygen uptake rate by 31%. Thus, growth associated maintenance cost was deemed an influential parameter. This differs from the conclusion in Varma and Palsson (1993) since this study looked at a larger range of values, which is reflective of the large uncertainty in quantifying growth associated maintenance costs. Pramanik and Keasling (1997) studied how a growth-rate dependent biomass equation affected intracellular fluxes. They observed significant differences in flux distribution through key reactions in the citric acid cycle and glycolysis between biomass compositions corresponding to a 70-minute and 145-minute doubling time. Use of an experimentally inconsistent biomass composition generated flux distributions that contradicted experimental observations. Based on the results of these sensitivity analyses, caution should be exercised before applying a single biomass equation for various scenarios and one must consider whether the magnitude of differences in biomass composition under different conditions is negligible.

Another consideration for the biomass equation is what metabolites should be included and what should be left out. Feist, Henry et al. (2007) conducted a sensitivity analysis on how the model predictions are affected by the set of metabolites included in the biomass equation.

Testing two different biomass equations, one based on the wild-type biomass composition and the other containing only the "minimal set of macromolecular molecules needed for cell viability", they found that the growth rate and oxygen uptake rate were essentially the same between the simulations using the two different biomass equations. However, they did notice that the simulation using the wild-type biomass composition required more non-zero flux reactions, which is logical since more metabolites in the biomass equation would require more active reactions for their synthesis.

In addition to providing a driving force for utilization of reactions, the choice of metabolites included in the biomass equation can affect predictions on gene essentiality. For instance, the inclusion of conditionally essential metabolites can lead to false identification of essential genes that are needed to synthesize the metabolites in the biomass equation. Imieliński, Belta et al. (2005) studied the relationship between metabolite essentiality and gene essentiality using a metabolic GEM for *E. coli*. Of the 81 mutants found to be not viable *in silico* due to inability to produce one or more components of the biomass equation, only half of these mutants were found to be lethal *in vivo*. They suggested that "either the production of these biomass metabolites is non-essential for survival or that the metabolic network annotation is incomplete with respect to pathways facilitating their production." In regards to non-essential metabolites, they proposed that "more complex associations may exist between metabolite production and survival" such that metabolite requirements may include OR relationships as opposed to strictly AND relationships. These relationships would allow for more flexibility in metabolite requirements. In addition to false positives for essential genes, they observed 39 false negatives wherein mutants that are capable of producing all biomass components *in silico* are nonviable *in vivo*. Eight of the 39 mutants had no effect on any reactions *in silico*, 11 mutants failed to knock

out any metabolites despite knocking out one or more reactions, and 20 mutants knocked out one or more metabolites outside of the biomass set. They gave four possible explanations for the inconsistencies between *in vivo* and *in silico* results: 1) isozymes and pathways capable of complementing the lost metabolic function "may be potentially suppressed *in vivo* via transcriptional regulation and other epigenetic mechanisms", 2) incorrect experimental assessment of lethality due to extremely slow growth, 3) all the roles of the genes may not be accounted for in the model, and 4) metabolites outside of the biomass equation may be essential. Thus, the choice of biomass equations in the metabolic GEM is non-trivial since it is a potential source of error.

2.2.3 Simulation of Phenotypes using Metabolic GEMs and Flux Balance Analysis (FBA)

In order to calculate the phenotypic state of the cell, a system of differential equations describing the mass balance for all the metabolites in the model is created. In a network with *m* metabolites and *n* reactions, the mass balance for each metabolite is given by the following equation:

$$\frac{dX_i}{dt} = \sum_{j=1}^{n} S_{ij} v_j \ , \qquad \text{for i} = 1, \dots, \text{m}$$

$$v_{j,min} \leq v_j \leq v_{j,max} \qquad \text{Equation 2.1}$$

where:

X_i = concentration of the *i*th metabolite, $(\frac{mmol}{g\ dW})$

dX_i/dt = time derivative of the *i*th metabolite concentration, $(\frac{mmol}{g\ dW-hr})$

$S_{i,j}$ = stoichiometric coefficient of the *i*th metabolite in the *j*th reaction

v_j = flux for the *j*th reaction, $(\frac{mmol}{g\ dW-hr})$

$v_{j,min}$ = lower flux bound for the *j*th reaction, ($\frac{mmol}{g\ dW-hr}$)

$v_{j,max}$ = upper flux bound for the *j*th reaction, ($\frac{mmol}{g\ dW-hr}$)

Or in more succinct matrix notation:

$$\frac{dX}{dt} = S \cdot v$$

$$v_{min} \leq v \leq v_{max}$$

Equation 2.2

where:

X = vector of metabolite concentrations, ($\frac{mmol}{g\ dW}$)

dX/dt = time derivative of the concentration vector, ($\frac{mmol}{g\ dW-hr}$)

S = stoichiometric matrix of *m* metabolites and *n* reactions

v = flux vector, ($\frac{mmol}{g\ dW-hr}$)

$v_{\min}$ = lower bound flux vector, ($\frac{mmol}{g\ dW-hr}$)

$v_{\max}$ = upper bound flux vector, ($\frac{mmol}{g\ dW-hr}$)

Pseudo-steady state is commonly assumed so concentration of all internal metabolites and fluxes for all reactions are treated as constants. This reduces the set of mass balances to a system of linear homogenous equations:

$$S \cdot v = 0$$

$$v_{min} \leq v \leq v_{max}$$

Equation 2.3

The pseudo-steady state assumption is justified by the very high turnover of most metabolites which allow for "the concentrations of the different metabolite pools [to] rapidly adjust to new levels, even after large perturbations in the environment experienced by the cells"

(Stephanopoulos, Aristidou et al. 1998). In the case of batch growth, the batch time is discretized into uniform time intervals and a new steady state is determined for each time interval, reflecting the response to changes in the environment (Varma and Palsson 1995). Thus a dynamic system is modeled as a series of steady-states.

Equation 2.3 is typically underdetermined since the number of unknown fluxes exceeds the number of mass balance equations. As such, this underdetermined system has an infinite but bounded set of solutions that mathematically satisfy all imposed constraints. The actual *in vivo* flux distribution theoretically lies within this space of feasible flux distributions. The approach that flux balance analysis (FBA) takes to determine a likely flux distribution is to impose an objective function on the metabolic network. This establishes a linear programming problem as described in Equation 2.4:

$$Maximize\ Z = c^T v$$
$$s.t.\quad S \cdot v = 0 \qquad \text{Equation 2.4}$$
$$v_{min} \leq v \leq v_{max}$$

where:

Z = objective function

c = vector of weights indicating how much each reaction contributes to the objective function

The objective function represents a hypothetical optimization goal that governs cellular behavior and that directs metabolic activities to achieve the optimum value of the objective function. The most frequently used objective function is maximization of growth rate. This objective function assumes that evolutionary processes select for cells with maximized growth performance. Using maximization of biomass growth as the objective function, Famili, Förster et

al. (2003) were able to predict glucose uptake rate and production rates of ethanol, glycerol, succinate, and biomass consistent with independently obtained experimental data. This objective function has also been applied to simulate growth of strains with gene deletions (Fong and Palsson 2004) and growth on uncommon carbon substrates (Ibarra, Edwards et al. 2002). After repeated cultivation to allow for adaptive evolution, the experimental growth rate converged over time to the predicted optimal growth rate. It was concluded that although "the adaptive evolutionary path itself cannot be predicted [by metabolic GEMs], the final outcome can be" (Ibarra, Edwards et al. 2002).

In addition to maximization of growth rate, other physiologically relevant objective functions have been posited. Schuetz, Kuepfer et al. (2007) examined how the predictive capability of FBA is affected by the objective function used. They utilized 11 different objective functions and compared the predictions for central carbon metabolism to ^{13}C-based *in vivo* flux distributions under different growth conditions. They found that "no single objective predicted the experimental data for wild-type *E. coli* under all conditions." Batch cultures with excess nutrient supply were found to be best modeled by maximization of ATP yield per unit of flux. The rationale behind this objective function is that "cells operate to maximize ATP yield while minimizing enzyme usage". Glucose- or ammonium-limited chemostat cultures were found to be best modeled by maximization of overall ATP yield or biomass yield. The rationale behind these objective functions is that evolution drives selection for maximal energetic efficiency or maximal biomass yield, respectively.

The use of FBA can allow for insight and understanding of cellular behavior. Simeonidis, Murabito et al. (2010) were able to simulate different fermentation phenomenon through the use of different objective functions and constraints. Overflow metabolism, in which "*S. cerevisiae* is

subjected to an increased glycolytic flux and metabolism shifts from pure respiration to a mix of respiratory and fermentative behavior", was able to be simulated by maximizing for biomass growth, restricting the oxygen uptake rate, and constraining the glucose uptake above a certain threshold. The Crabtree effect, in which "*S. cerevisiae* produces ethanol aerobically in the presence of high external glucose concentrations rather than producing biomass through the TCA cycle", was able to be simulated using the hypothesis that cells "minimize the resources needed to guarantee a certain biomass production rate". This was implemented by constraining growth rate at its optimum value, minimizing the number of active (i.e. non-zero flux) reactions, and constraining the glucose uptake above a certain threshold.

FBA has also been applied to model triacylglycerol homeostasis. Zanghellini, Natter et al. (2008) examined how the products of TAG hydrolysis are utilized in exit from stasis back into active growth and how this affects membrane synthesis. Through the use of different objective functions and modeling constructs, they were able to generate and test hypotheses on TAG metabolism. For instance, they found that the objective functions of maximizing TAG mobilization and maximizing membrane production were interchangeable and yielded similar TAG degradation rates and membrane particle production rates between the two simulations. This suggests that membrane synthesis is tightly linked with the generation of membrane-precursors from TAG hydrolysis.

2.2.4 Refinements to FBA

FBA returns a singular flux distribution that satisfies the optimization of the objective function. However, there can be multiple valid flux distributions that are able to achieve the same optimal value for the objective function. A simple example of this is illustrated in Figure 2.3.

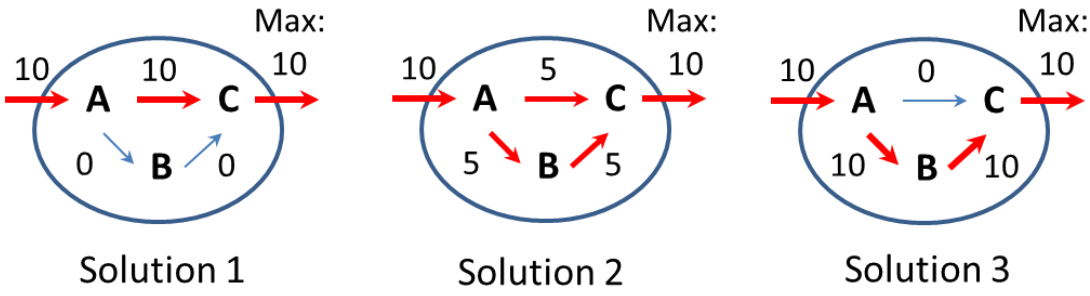


Figure 2.3: Simple example of multiple solutions achieving the same optimal value. While linear programming is able to determine a unique optimal value for the objective function, multiple solutions capable of achieving this optimal value may exist. In this example, different flux distributions yield the same maximum excretion of compound C.

These alternate optimal flux distributions arise from the functional redundancies found in the metabolic network which enhance robustness against genetic and environmental perturbations (Wang and Zhang 2009). In order to examine the alternate optimal flux distributions, Mahadevan and Schilling (2003) created the method of flux variability analysis. This approach determines the range of fluxes possible for each reaction such that all imposed constraints are satisfied and the same optimal value for the objective function is obtained. This is accomplished using a series of linear programming problems with one case maximizing each flux and the second case minimizing each flux. The mathematical formulation of this approach is shown in Equation 2.5 and 2.6:

$$Case\ 1:$$

$$Max\ v_j\ , \qquad \text{for j} = 1, \ldots, \text{n}$$

$$s.t. \quad S \cdot v = 0$$

$$c^T v = Z_{obj}$$

$$v_{j,min} \leq v_j \leq v_{j,max}$$

Equation 2.5

$$Case\ 2:$$

$$Min\ v_j\ , \qquad \text{for j} = 1, \ldots, \text{n}$$

$$s.t. \quad S \cdot v = 0$$

$$c^T v = Z_{obj}$$

$$v_{j,min} \leq v_j \leq v_{j,max}$$

Equation 2.6

where Z_{obj} = optimal value of the objective function calculated previously from Equation 2.4.

In comparing internal flux values derived from ^{13}C labeling experiments with the flux ranges generated from flux variability analysis, Chen, Alonso et al. (2011) found that the experimentally derived fluxes were contained within the predicted ranges while Puchałka, Oberhardt et al. (2008) found a few discrepancies. These discrepancies were attributed to *in vivo* cellular behavior deviating from optimal function, a key assumption in FBA. After conducting suboptimal flux variability analysis in which cell growth was allowed to vary from 90-100% of the maximum value, the ^{13}C-derived fluxes were all found to be contained within the predicted ranges. Despite the validity of flux variability analysis, both studies found that the generated flux ranges for certain reactions were fairly wide which belies its predictive power for those reactions. Nevertheless, Chen, Alonso et al. (2011) cautioned against placing full confidence in the internal flux values obtained from the singular flux distribution output of FBA without consideration of the feasible flux ranges or further flux constraints.

Application of additional constraints can further restrict the space of feasible fluxes. One approach is to incorporate experimentally measured fluxes to reduce the degrees of freedom in the solution (Wiback, Mahadevan et al. 2004; Teusink, Wiersma et al. 2006). However, "it is

important to realize that when fluxes are constrained by additional observations, the model becomes more and more descriptive, and less and less predictive" (Teusink, Wiersma et al. 2006). In this sense, the applicability of these additional flux constraints to simulate metabolic behavior under conditions different than those for which they were measured is unclear.

An area of active research is the integration of transcriptional regulation to further restrict the space of feasible fluxes to more biologically consistent solutions. Covert, Schilling et al. (2001) developed a framework to describe transcriptional regulation using Boolean logic equations which includes AND, OR, and NOT relationships. These Boolean logic equations were used to determine whether transcription of a gene occurs based on the "presence or surplus of an intracellular metabolite, an extracellular metabolite, regulatory proteins, signaling molecule, or any combination of these or other factors" (Covert, Schilling et al. 2001). Only the transcribed genes were able to contribute to the flux solution. Using this framework, Covert and Palsson (2002) created a model for central metabolism in *E. coli* that accounts for regulation of gene expression. Incorporation of regulatory rules improved predictions of mutant lethality due to the consideration of down-regulated reactions and also allowed for simulation of diauxic growth on glucose and lactose, which would have otherwise been simulated as simultaneous uptake with the stand-alone FBA model. Since this paper, metabolic GEMs incorporating transcriptional regulation via Boolean logic have been built for *E. coli* (Covert, Knight et al. 2004) and *S. cerevisiae* (Herrgård, Lee et al. 2006).

A main weakness with the Boolean logic approach is its simplification of transcriptional regulation to binary on/off responses which do not allow for intermediate gene expression. In contrast, Chandrasekaran and Price (2010) utilized probabilities to represent the relationship between gene expression and gene-transcription factor interactions; the probability of gene

expression given a transcription factor state was inferred from large sets of high-throughput microarray data obtained for various environmental and genetic perturbations. The use of probabilities allowed for differentiation between strong and weak regulators, as opposed to completely on/off representation in Boolean logic. This difference yielded improved predictions on the phenotype of transcription factor mutants.

2.2.5 Advancements in Metabolic GEMs for *S. cerevisiae*

The process of creating a metabolic GEM is often iterative and ongoing (Thiele and Palsson 2010). Since the first metabolic GEM for *S. cerevisiae* was published in 2003, numerous models have been released that incorporate new knowledge on biochemical reactions and genes and that introduce different modeling assumptions/representations (see Figure 2.4).

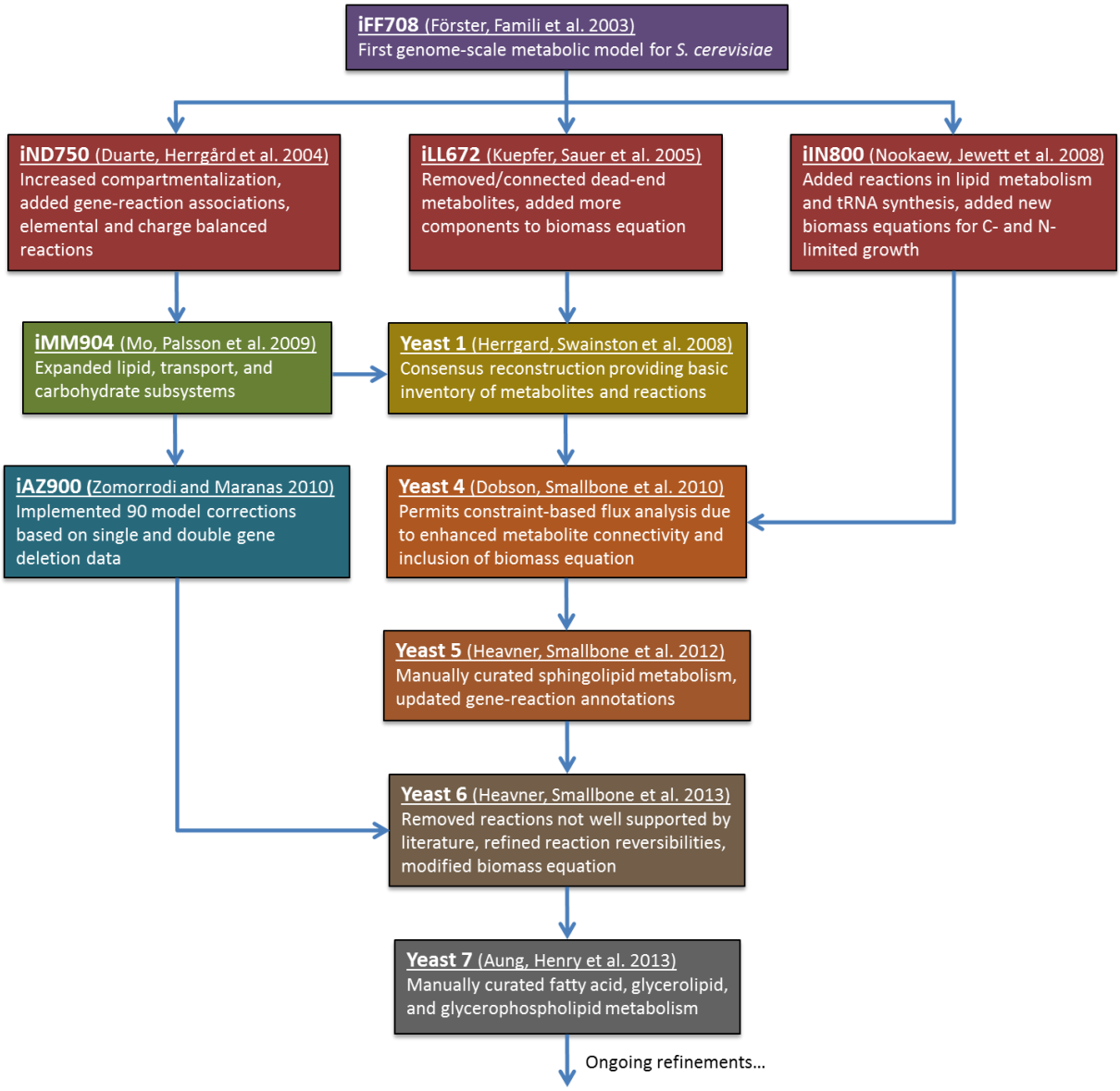


Figure 2.4: Evolution of *S. cerevisiae* metabolic GEMs. The metabolic GEM for *S. cerevisiae* has undergone numerous updates and expansions. Since iFF708, each subsequent model has used a prior model as a template to build upon. In the case of Yeast 1.0, the information contained in two models was utilized to create a consensus model. This Yeast consensus model has since been updated to reflect changes suggested by other models and by the research community.

The first metabolic GEM for *S. cerevisiae* was iFF708 (Förster, Famili et al. 2003a), in which the nomenclature refers to "i" for *in silico* model, "FF" for the initials of the authors, and "708" for the total number of genes accounted for in the model. The reconstruction process began by generating a gene catalog from the KEGG database for *S. cerevisiae*. Additional genes were added from the databases CYGD and SGD. Genes involved in regulation, such as activators or negative regulators, and genes assigned function based on similarity searches were excluded from the gene catalog. Reactions were then constructed from this catalog using the associated Enzyme Commision numbers and the ENZYME nomenclature database or using primary literature; reactions without an annotated open reading frame (ORF) but with biochemical and physiological evidence for its existence were also included in the model. Reactions were localized to the cytosol, extracellular space, or mitochondria in the model based on information from the databases CYGD and YPD; if a reaction was localized in another compartment *in vivo* or if localization was unknown, the reaction was placed in the cytosol in the model. The work in building iFF708 culminated in a metabolic network containing 1175 reactions and 733 metabolites. As a metric of the model's predictive ability, the phenotype of *in silico* gene deletions were compared to experimental data from a competitive growth assay (Winzeler, Shoemaker et al. 1999) and available data from the databases CYGD and SGD. For the 599 experimental growth phenotypes, iFF708 was able to correctly predict 89.4% of the viable phenotypes and 68.2% of the lethal phenotypes for an overall accuracy of 87.8% (Förster, Famili et al. 2003b).

iFF708 has served as a direct template for building three independent metabolic GEMs (iND750, iLL672, and iIN800). The first of these models was iND750 (Duarte, Herrgård et al. 2004). Major revisions to the iFF708 model included: 1) increased compartmentalization from

three cellular locations (extracellular space, cytosol, and mitochondrion) to eight cellular locations with the addition of peroxisome, nucleus, Golgi apparatus, vacuole, and endoplasmic reticulum, 2) elementally and charge balanced reactions that impose cell-wide proton balance, and 3) integration of a logic expression for each reaction indicating gene(s) involved in the reaction. The number of reactions in iND750 increased because of the addition of new genes and their corresponding reactions, the expansion of lumped reactions and generic metabolites, and the addition of intercompartmental transport reactions. This led to a metabolic GEM containing 1489 reactions and 646 metabolites. As part of its validation, iND750's phenotype predictions were compared to two published large-scale gene deletion studies (Giaever, Chu et al. 2002; Steinmetz, Scharfe et al. 2002) for seven different media conditions. These studies allowed for analysis of 682 of the total 750 genes in the model. For the 4154 experimental growth phenotypes, iND750 was able to correctly predict 96.6% of the viable phenotypes and 23.0% of the lethal phenotypes for an overall accuracy of 82.6%. The decreased lethal and overall predictive performance of iND750 as compared to iFF708 was mainly attributed to "the involvement of metabolic genes in other cellular processes that are not accounted for in the current model" (Duarte, Herrgård et al. 2004). For instance, iND750 does not include mRNA and protein synthesis but does include the pathways for synthesis of various RNA species such as transfer RNAs. Thus, without reactions utilizing these species, no flux can go through these pathways and consequently these pathways are indispensable in simulations despite *in vivo* essentiality.

iLL672 sought to improve iFF708's lethality predictions by including additional metabolites in the biomass equation, removing 110 reactions leading to dead-end metabolites, and closing the gap for 33 dead-end reactions (Kuepfer, Sauer et al. 2005). As a result of these

changes, iLL672 consists of 1038 reactions and 636 metabolites. Comparison of model predictions with experimental data obtained from plate growth of single-gene deletion mutants under five different media conditions showed improved predictive accuracy for iLL672. The precursor model, iFF708, was able to correctly predict viable phenotypes for 90-96% of the cases and lethal phenotypes for 40-53% of the cases depending on media considered. iLL672's predictive capability was higher with accuracies of 96-98% for viable phenotypes and 68-80% for lethal phenotypes. However, since iLL672 pruned dead-end metabolites from the original network by removal of reactions, iLL672 had a smaller scope than iFF708, accounting for 672 ORFs versus 708 ORFs.

Considering the incomplete representation of lipid metabolism in the three aforementioned models, iIN800 added more reactions and metabolites to the iFF708 model to provide increasing detail and comprehension of lipid metabolism (Nookaew, Jewett et al. 2008). New or expanded reactions in cytosolic and mitochondrial fatty acid synthesis, fatty acid elongation, fatty acid activation, β-oxidation, sphingolipid synthesis, ergosterol esterificiation, and lipid degradation were incorporated, which yielded at least 143 more reactions in lipid metabolism than previous models. In addition to these new reactions, iIN800 presented two distinct biomass equations for carbon-limited growth and for nitrogen-limited growth. In total, iIN800 consists of 1446 reactions and 1013 metabolites.

Additional models have since been derived from these second generation metabolic GEMs. iMM904 expanded on the content of iND750 and increased the level of coverage by 154 genes (Mo, Palsson et al. 2009). These new additions were predominantly in lipid, transport, and carbohydrate subsystems. Based on iLL672's improved lethal predictions, iMM904 also reformulated the biomass equation to include the cofactors and metabolites added in iLL672's

biomass equation. Overall, iMM904 contains 1412 reactions and 1228 metabolites. When compared to the experimental dataset obtained in the iLL672 study (Kuepfer, Sauer et al. 2005), iMM904 was able to correctly predict 97.6% of the viable phenotypes and 32.5% of the lethal phenotypes for an overall accuracy of 83.0% (Zomorrodi and Maranas 2010).[*] Compared to iLL672, iMM904 had decreased predictive capability for lethal phenotypes. This was attributed to the inclusion of dead-end pathways in iMM904, which were removed in iLL672. Despite the decrease in accuracy, iMM904 retained these dead-end pathways because "it captures a more complete picture of currently known metabolic functions and provides a framework for network expansion as new pathways are elucidated" (Mo, Palsson et al. 2009).

In order to remedy the discrepancy between iMM904's predictions and experimental data, Zomorrodi and Maranas (2010) reconciled many of the differences for single and double gene deletions through a procedure called GrowMatch (Kumar and Maranas 2009). Inconsistencies in which iMM904 predicted no growth while the mutant was able to grow *in vivo* were resolved through combinations of: adding new reactions or isozymes by using bidirectional protein-protein BLAST search to find genes in *S. cerevisiae* potentially capable of catalyzing a particular reaction, relaxing reaction irreversibility based on Gibbs free energy change or literature evidence, adding more metabolites to the *in silico* media that represents YP media, and adding transport reactions for certain metabolites between compartments. Inconsistencies with *in silico* growth and *in vivo* lethality were resolved through combinations of: reevaluating GPR relationship for reactions catalyzed by isozymes (i.e., should a gene be included and is the

[*] When Mo, Palsson et al. (2009) calculated these parameters, they achieved accuracies of 94.8% for viable phenotypes and 59.9% for lethal phenotypes for an overall accuracy of 90.9%. However, these values were obtained by adjusting the threshold for predicted growth rate considered to be viable so that the true viable rate and false viable rate is optimized. This approach may not be a good representation of the model's accuracy since it attempts to minimize the discrepancy between experimental data and predictions through parameter adjustment. Additionally, these accuracies were reported for the testing of 722 genes, which does not give a complete representation for all 904 ORFs in the model.

appropriate logic rule applied), suppressing isozymes through regulatory rules based on gene expression data and literature evidence, and modifying the biomass equation (e.g., including charged and uncharged tRNA molecules). This GrowMatch procedure led to 120 model modifications which consisted of 90 model corrections and 30 added regulatory constraints. Although unnamed in the paper, this model can be considered as iAZ900 based on the established model nomenclature. iAZ900 consists of 1597 reactions and 1404 metabolites. When compared to the single gene deletion experimental dataset obtained in the iLL672 study (Kuepfer, Sauer et al. 2005), the changes implemented in iAZ900 increased the accuracy for viable phenotypes from 97.6% to 99.3%, 32.5% to 47.5% for lethal phenotypes, and 83.0% to 87.7% for overall accuracy.

As can be seen in Figure 2.4, there has been divergence in the development of *S. cerevisiae* metabolic GEMs which has created inconsistencies in representation and scope between the different models. Thus, several research groups worked together to consolidate "the community knowledge of yeast metabolism"; the first version of this consensus model, Yeast 1.0, was created by merging the content of iLL672 and iMM904 (Herrgard, Swainston et al. 2008). In the generation of the consensus network, ambiguous generic metabolites (i.e. molecules with R-groups) were removed, which consequently created fragments in the metabolic network and "led to the underrepresentation of lipids" (Herrgard, Swainston et al. 2008). Due to this fragmentation and the exclusion of a pre-assumed biomass equation, Yeast 1.0 was not able to be used for constraint –based analysis but rather served as a "basic inventory of metabolites and network structure" (Herrgard, Swainston et al. 2008). However, since Yeast 1.0, many improvements have been made to increase the connectivity of the metabolic network and the coverage of known metabolic reactions. The expansion in lipid metabolism made by iIN800 has

since been incorporated into the consensus model. With "its enhanced representation of lipid metabolism and greater connectivity", Yeast 4.0 was the first version of the consensus model utilizable for constraint-based analysis (Dobson, Smallbone et al. 2010). The Yeast consensus model continues to be improved upon through corrections and updates by the research community. The representation of sphingolipid, fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model has since been improved through manual curation (Heavner, Smallbone et al. 2012; Aung, Henry et al. 2013). Additional manual curation has also led to the removal of reactions not supported by literature review and the refinement of reaction reversibility (Heavner, Smallbone et al. 2013). The latest version of the Yeast consensus model can be found at yeast.sourceforge.net/.

2.2.6 Sources of False Predictions

Comparisons between model predictions and experimental data show high predictive accuracies for viable single gene deletions but reduced accuracies for lethal single gene deletions (Table 2.2). The challenges in predicting essential genes can be attributed to several factors such as biomass composition, incomplete pathways, difference in activities of isozymes, the role of genes in non-metabolic processes, and missing regulatory mechanisms.

Table 2.2: Accuracy for single gene deletions

Model	True Viable *	True Lethal **	Source of Experimental Data
iFF708	89.4%	68.2%	(Winzeler, Shoemaker et al. 1999), CYGD, SGD
iND750	96.6%	23.0%	(Giaever, Chu et al. 2002; Steinmetz, Scharfe et al. 2002)
iLL672	96-98%	68-80%	(Kuepfer, Sauer et al. 2005)
iIN800	95.5%	38.7%	(Winzeler, Shoemaker et al. 1999; Kuepfer, Sauer et al. 2005)
iMM904	97.6%	32.5%	(Kuepfer, Sauer et al. 2005)
iAZ900	99.3%	47.5%	(Kuepfer, Sauer et al. 2005)
Yeast 7.11	98.4%	67.9%	(Winzeler, Shoemaker et al. 1999), SGD

The percentages for iLL672 show the range of accuracies over different media conditions. For the rest of the models, the single accuracy is calculated by combining the data for all media conditions considered.

$$* \, True\ viable = \frac{TP}{TP+FN} \, , ** \, True\ lethal = \frac{TN}{TN+FP}$$

TP = true positive (correctly predicted viable phenotype), TN = true negative (correctly predicted lethal phenotype), FP = false positive (wrongly predicted viable phenotype), FN = false negative (wrongly predicted lethal phenotype).

As discussed in the sub-section "Genome-Scale Model (GEM)", the choice of biomass composition can affect simulation results since it dictates what metabolites are essential to the cell and in what quantities they are present. Thus, one established approach to increase the predictive accuracy is to increase the number and specificity of the metabolites included in the biomass equation. Some accuracy improvements have been shown using this approach (Kuepfer, Sauer et al. 2005). Another issue that affects predictive accuracy is incomplete representation of metabolic pathways. For example, in iND750, "many of the dead ends are in phospholipids metabolism in which the corresponding genes participate in the biosynthesis of complex phospholipids that are not used within the model, but that are probably converted to essential membrane phospholipids" (Duarte, Herrgård et al. 2004). These gaps continue to be filled by model updates that reflect the growing knowledge of yeast metabolism.

The predictive capability of constraint-based models can sometimes be hindered by the models' simplifying assumptions and scope. As is, the GPR relationships included in the model do not indicate how much each isozyme contributes to the reaction activity. Thus, the knockout of a dominant isozyme would be completely compensated by other isozymes *in silico* despite significant decrease in activity *in vivo*. Also, as is, the models can only determine the effect of gene deletions on the metabolic network, as opposed to the additional impacts on other components of cellular function like genetic and environmental information processing or the cell cycle. Nevertheless, progression has been made in incorporating selected regulatory interactions into the *S. cerevisiae* metabolic GEM (Herrgård, Lee et al. 2006; Zomorrodi and Maranas 2010). These additions allow for down-regulation of genes under certain conditions, which consequently alters the flux distribution throughout the metabolic network.

REFERENCES

Athenstaedt, K. and G. Daum (1997). "Biosynthesis of phosphatidic acid in lipid particles and endoplasmic reticulum of Saccharomyces cerevisiae." Journal of Bacteriology **179**(24): 7611-7616.

Athenstaedt, K. and G. n. Daum (2000). "1-Acyldihydroxyacetone-phosphate Reductase (Ayr1p) of the Yeast Saccharomyces cerevisiae Encoded by the Open Reading Frame YIL124w Is a Major Component of Lipid Particles." Journal of Biological Chemistry **275**(1): 235-240.

Aung, H. W., S. A. Henry and L. P. Walker (2013). "Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism." Industrial Biotechnology **9**(4): 215-228.

Bae-Lee, M. S. and G. M. Carman (1984). "Phosphatidylserine synthesis in Saccharomyces cerevisiae. Purification and characterization of membrane-associated phosphatidylserine synthase." Journal of Biological Chemistry **259**(17): 10857-10862.

Benghezal, M., C. Roubaty, V. Veepuri, J. Knudsen and A. Conzelmann (2007). "SLC1 and SLC4 Encode Partially Redundant Acyl-Coenzyme A 1-Acylglycerol-3-phosphate O-Acyltransferases of Budding Yeast." Journal of Biological Chemistry **282**(42): 30845-30855.

Chandrasekaran, S. and N. D. Price (2010). "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis." Proceedings of the National Academy of Sciences **107**(41): 17845-17850.

Chang, S.-C., P. N. Heacock, C. J. Clancey and W. Dowhan (1998). "The PEL1 Gene (Renamed PGS1) Encodes the Phosphatidylglycerophosphate Synthase of Saccharomyces cerevisiae." Journal of Biological Chemistry **273**(16): 9829-9836.

Chen, X., A. P. Alonso, D. K. Allen, J. L. Reed and Y. Shachar-Hill (2011). "Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. coli." Metabolic Engineering **13**(1): 38-48.

Clancey, C. J., S. C. Chang and W. Dowhan (1993). "Cloning of a gene (PSD1) encoding phosphatidylserine decarboxylase from Saccharomyces cerevisiae by complementation of an Escherichia coli mutant." Journal of Biological Chemistry **268**(33): 24580-24590.

Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard and B. O. Palsson (2004). "Integrating high-throughput and computational data elucidates bacterial networks." Nature **429**(6987): 92-96.

Covert, M. W. and B. Ø. Palsson (2002). "Transcriptional Regulation in Constraints-based Metabolic Models of Escherichia coli." Journal of Biological Chemistry **277**(31): 28058-28064.

Covert, M. W., C. H. Schilling and B. Palsson (2001). "Regulation of Gene Expression in Flux Balance Models of Metabolism." Journal of Theoretical Biology **213**(1): 73-88.

Cowart, L. A. and L. M. Obeid (2007). "Yeast sphingolipids: Recent developments in understanding biosynthesis, regulation, and function." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 421-431.

Czabany, T., A. Wagner, D. Zweytick, K. Lohner, E. Leitner, E. Ingolic and G. n. Daum (2008). "Structural and Biochemical Properties of Lipid Particles from the Yeast Saccharomyces cerevisiae." Journal of Biological Chemistry **283**(25): 17065-17074.

Dahlqvist, A., U. Ståhl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne and S. Stymne (2000). "Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants." Proceedings of the National Academy of Sciences **97**(12): 6487-6492.

Dickson, R. C. (2008). "Thematic Review Series: Sphingolipids. New insights into sphingolipid metabolism and function in budding yeast." Journal of Lipid Research **49**(5): 909-921.

Dickson, R. C., C. Chalfant and M. D. Poeta (2010). Roles for Sphingolipids in Saccharomyces cerevisiae, Springer New York. **688:** 217-231.

Dobson, P., K. Smallbone, D. Jameson, E. Simeonidis, K. Lanthaler, P. Pir, C. Lu, N. Swainston, W. Dunn, P. Fisher, D. Hull, M. Brown, O. Oshota, N. Stanford, D. Kell, R. King, S. Oliver, R. Stevens and P. Mendes (2010). "Further developments towards a genome-scale metabolic model of yeast." BMC Systems Biology **4**(1): 145.

Duarte, N. C., M. J. Herrgård and B. Ø. Palsson (2004). "Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model." Genome Research **14**(7): 1298-1309.

Durot, M., P.-Y. Bourguignon and V. Schachter (2009). "Genome-scale models of bacterial metabolism: reconstruction and applications." FEMS Microbiology Reviews **33**(1): 164-190.

Fakas, S., C. Konstantinou and G. M. Carman (2011). "DGK1-encoded Diacylglycerol Kinase Activity Is Required for Phospholipid Synthesis during Growth Resumption from Stationary Phase in Saccharomyces cerevisiae." Journal of Biological Chemistry **286**(2): 1464-1474.

Famili, I., J. Förster, J. Nielsen and B. Ø. Palsson (2003). "Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network." Proceedings of the National Academy of Sciences **100**(23): 13134-13139.

Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. Ø. Palsson (2007). "A genome-scale metabolic

reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**.

Feist, A. M., M. J. Herrgard, I. Thiele, J. L. Reed and B. Ø. Palsson (2009). "Reconstruction of biochemical networks in microorganisms." Nat Rev Micro **7**(2): 129-143.

Fong, S. S. and B. Ø. Palsson (2004). "Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes." Nat Genet **36**(10): 1056-1058.

Förster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen (2003a). "Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network." Genome Research **13**(2): 244-253.

Förster, J., I. Famili, B. Ø. Palsson and J. Nielsen (2003b). "Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae." Omics A Journal of Integrative Biology **7**(2): 193-202.

Gaspar, M. L., M. A. Aregullin, S. A. Jesch and S. A. Henry (2006). "Inositol Induces a Profound Alteration in the Pattern and Rate of Synthesis and Turnover of Membrane Lipids in Saccharomyces cerevisiae." Journal of Biological Chemistry **281**(32): 22773-22785.

Gaspar, M. L., H. F. Hofbauer, S. D. Kohlwein and S. A. Henry (2011). "Coordination of Storage Lipid Synthesis and Membrane Biogenesis." Journal of Biological Chemistry **286**(3): 1696-1708.

Gaspar, M. L., S. A. Jesch, R. Viswanatha, A. L. Antosh, W. J. Brown, S. D. Kohlwein and S. A. Henry (2008). "A Block in Endoplasmic Reticulum-to-Golgi Trafficking Inhibits Phospholipid Synthesis and Induces Neutral Lipid Accumulation." Journal of Biological Chemistry **283**(37): 25735-25751.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C.-y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis and M. Johnston (2002). "Functional profiling of the Saccharomyces cerevisiae genome." Nature **418**(6896): 387-391.

Gray, J. V., G. A. Petsko, G. C. Johnston, D. Ringe, R. A. Singer and M. Werner-Washburne (2004). ""Sleeping Beauty": Quiescence in Saccharomyces cerevisiae." Microbiology and Molecular Biology Reviews **68**(2): 187-206.

Han, G.-S., C. N. Johnston, X. Chen, K. Athenstaedt, G. n. Daum and G. M. Carman (2001). "Regulation of the Saccharomyces cerevisiae DPP1-encoded Diacylglycerol Pyrophosphate Phosphatase by Zinc." Journal of Biological Chemistry **276**(13): 10126-10133.

Han, G.-S., W.-I. Wu and G. M. Carman (2006). "The Saccharomyces cerevisiae Lipin Homolog Is a Mg2+-dependent Phosphatidate Phosphatase Enzyme." Journal of Biological Chemistry **281**(14): 9210-9218.

Heavner, B. D., K. Smallbone, B. Barker, P. Mendes and L. P. Walker (2012). "Yeast 5–an expanded reconstruction of the Saccharomyces cerevisiae metabolic network." BMC Systems Biology **6**(1): 55.

Heavner, B. D., K. Smallbone, N. D. Price and L. P. Walker (2013). "Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance." Database **2013**.

Henry, C. S., M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay and R. L. Stevens (2010). "High-throughput generation, optimization and analysis of genome-scale metabolic models." Nat Biotech **28**(9): 977-982.

Henry, S. A., S. D. Kohlwein and G. M. Carman (2012). "Metabolism and Regulation of Glycerolipids in the Yeast Saccharomyces cerevisiae." Genetics **190**(2): 317-349.

Herrgård, M. J., B.-S. Lee, V. Portnoy and B. Ø. Palsson (2006). "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae." Genome Research **16**(5): 627-635.

Herrgard, M. J., N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Büthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasie, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kurdar, M. Penttila, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen and D. B. Kell (2008). "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology." Nat Biotech **26**(10): 1155-1160.

Hiltunen, J. K., A. M. Mursula, H. Rottensteiner, R. K. Wierenga, A. J. Kastaniotis and A. Gurvitz (2003). "The biochemistry of peroxisomal β-oxidation in the yeast Saccharomyces cerevisiae." FEMS Microbiology Reviews **27**(1): 35-64.

Hjelmstad, R. H. and R. M. Bell (1987). "Mutants of Saccharomyces cerevisiae defective in sn-1,2-diacylglycerol cholinephosphotransferase. Isolation, characterization, and cloning of the CPT1 gene." Journal of Biological Chemistry **262**(8): 3909-3917.

Hjelmstad, R. H. and R. M. Bell (1988). "The sn-1,2-diacylglycerol ethanolaminephosphotransferase activity of Saccharomyces cerevisiae. Isolation of

mutants and cloning of the EPT1 gene." Journal of Biological Chemistry **263**(36): 19748-19757.

Hjelmstad, R. H. and R. M. Bell (1991). "sn-1,2-diacylglycerol choline- and ethanolaminephosphotransferases in Saccharomyces cerevisiae. Mixed micellar analysis of the CPT1 and EPT1 gene products." Journal of Biological Chemistry **266**(7): 4357-4365.

Huh, W.-K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-691.

Ibarra, R. U., J. S. Edwards and B. Ø. Palsson (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." Nature **420**(6912): 186-189.

Imieliński, M., C. Belta, Ã. Halász and H. Rubin (2005). "Investigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities." Bioinformatics **21**(9): 2008-2016.

Jamshidi, N. and B. Ø. Palsson (2008). "Formulating genome-scale kinetic models in the post-genome era." Mol Syst Biol **4**.

Karanasios, E., G.-S. Han, Z. Xu, G. M. Carman and S. Siniossoglou (2010). "A phosphorylation-regulated amphipathic helix controls the membrane translocation and function of the yeast phosphatidate phosphatase." Proceedings of the National Academy of Sciences **107**(41): 17539-17544.

Karp, P. D., S. Paley and P. Romero (2002). "The Pathway Tools software." Bioinformatics **18**(suppl 1): S225-S232.

Kelley, M. J., A. M. Bailis, S. A. Henry and G. M. Carman (1988). "Regulation of phospholipid biosynthesis in Saccharomyces cerevisiae by inositol. Inositol is an inhibitor of phosphatidylserine synthase activity." Journal of Biological Chemistry **263**(34): 18078-18085.

Kim, H., K. Melénn, M. Österberg and G. von Heijne (2006). "A global topology map of the Saccharomyces cerevisiae membrane proteome." Proceedings of the National Academy of Sciences **103**(30): 11142-11147.

Kiyono, K., K. Miura, Y. Kushima, T. Hikiji, M. Fukushima, I. Shibuya and A. Ohta (1987). "Primary Structure and Product Characterization of the Saccharomyces cerevisiae CHO1 Gene That Encodes Phosphatidylserine Synthase." Journal of Biochemistry **102**(5): 1089-1100.

Kodaki, T. and S. Yamashita (1987). "Yeast phosphatidylethanolamine methylation pathway. Cloning and characterization of two distinct methyltransferase genes." Journal of Biological Chemistry **262**(32): 15428-15435.

Kodaki, T. and S. Yamashita (1989). "Characterization of the methyltransferases in the yeast phosphatidylethanolamine methylation pathway by selective gene disruption." European Journal of Biochemistry **185**(2): 243-251.

Kohei, H. and Y. Satoshi (1984). "Regulatory role of phosphatidate phosphatase in triacylglycerol synthesis of Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **796**(1): 110-117.

Kuepfer, L., U. Sauer and L. M. Blank (2005). "Metabolic functions of duplicate genes in Saccharomyces cerevisiae." Genome Research **15**(10): 1421-1430.

Kumar, V. S. and C. D. Maranas (2009). "GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions." PLoS Comput Biol **5**(3): e1000308.

Kurat, C. F., K. Natter, J. Petschnigg, H. Wolinski, K. Scheuringer, H. Scholz, R. Zimmermann, R. Leber, R. Zechner and S. D. Kohlwein (2006). "Obese Yeast: Triglyceride Lipolysis Is Functionally Conserved from Mammals to Yeast." Journal of Biological Chemistry **281**(1): 491-500.

Lefevre, S. D., C. W. van Roermund, R. J. A. Wanders, M. Veenhuis and I. J. van der Klei (2013). "The significance of peroxisome function in chronological aging of Saccharomyces cerevisiae." Aging Cell **12**(5): 784-793.

Letts, V. A., L. S. Klig, M. Bae-Lee, G. M. Carman and S. A. Henry (1983). "Isolation of the yeast structural gene for the membrane-associated enzyme phosphatidylserine synthase." Proceedings of the National Academy of Sciences **80**(23): 7279-7283.

Loewen, C. J. R., M. L. Gaspar, S. A. Jesch, C. Delon, N. T. Ktistakis, S. A. Henry and T. P. Levine (2004). "Phospholipid Metabolism Regulated by a Transcription Factor Sensing Phosphatidic Acid." Science **304**(5677): 1644-1647.

Mahadevan, R. and C. H. Schilling (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." Metabolic Engineering **5**(4): 264-276.

Mo, M., B. Palsson and M. Herrgard (2009). "Connecting extracellular metabolomic measurements to intracellular flux states in yeast." BMC Systems Biology **3**(1): 37.

Nikawa, J.-i., Y. Tsukagoshi, T. Kodaki and S. Yamashita (1987). "Nucleotide sequence and characterization of the yeast PSS gene encoding phosphatidylserine synthase." European Journal of Biochemistry **167**(1): 7-12.

Nikawa, J.-i. and S. Yamashita (1984). "Molecular cloning of the gene encoding CDPdiacylglycerol–inositol 3-phosphatidyl transferase in Saccharomyces cerevisiae." European Journal of Biochemistry **143**(2): 251-256.

Nikawa, J., T. Kodaki and S. Yamashita (1987). "Primary structure and disruption of the phosphatidylinositol synthase gene of Saccharomyces cerevisiae." Journal of Biological Chemistry **262**(10): 4876-4881.

Nookaew, I., M. Jewett, A. Meechai, C. Thammarongtham, K. Laoteng, S. Cheevadhanarak, J. Nielsen and S. Bhumiratana (2008). "The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism." BMC Systems Biology **2**(1): 71.

Oberhardt, M. A., B. Ø. Palsson and J. A. Papin (2009). "Applications of genome-scale metabolic reconstructions." Mol Syst Biol **5**.

Oelkers, P., D. Cromley, M. Padamsee, J. T. Billheimer and S. L. Sturley (2002). "The DGA1 Gene Determines a Second Triglyceride Synthetic Pathway in Yeast." Journal of Biological Chemistry **277**(11): 8877-8881.

Oelkers, P., A. Tinkelenberg, N. Erdeniz, D. Cromley, J. T. Billheimer and S. L. Sturley (2000). "A Lecithin Cholesterol Acyltransferase-like Gene Mediates Diacylglycerol Esterification in Yeast." Journal of Biological Chemistry **275**(21): 15609-15612.

Osman, C., D. R. Voelker and T. Langer (2011). "Making heads or tails of phospholipids in mitochondria." The Journal of Cell Biology **192**(1): 7-16.

Pittet, M. and A. Conzelmann (2007). "Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 405-420.

Poolman, M. G., B. K. Bonde, A. Gevorgyan, H. H. Patel and D. A. Fell (2006). "Challenges to be faced in the reconstruction of metabolic networks from public databases." IEE Proceedings-Systems Biology **153**: 379.

Poptsova, M. S. and J. P. Gogarten (2010). "Using comparative genome analysis to identify problems in annotated microbial genomes." Microbiology **156**(7): 1909-1917.

Pramanik, J. and J. D. Keasling (1997). "Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements." Biotechnology and Bioengineering **56**(4): 398-421.

Puchałka, J., M. A. Oberhardt, M. Godinho, A. Bielecka, D. Regenhardt, K. N. Timmis, J. A. Papin and V. A. P. M. dos Santos (2008). "Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology." PLoS computational biology **4**(10): e1000210.

Sandager, L., M. H. Gustavsson, U. Ståhl, A. Dahlqvist, E. Wiberg, A. Banas, M. Lenman, H. Ronne and S. Stymne (2002). "Storage Lipid Synthesis Is Non-essential in Yeast." Journal of Biological Chemistry **277**(8): 6478-6482.

Santos-Rosa, H., J. Leung, N. Grimsey, S. Peak-Chew and S. Siniossoglou (2005). "The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth." EMBO J **24**(11): 1931-1941.

Schuetz, R., L. Kuepfer and U. Sauer (2007). "Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli." Mol Syst Biol **3**.

Shen, H., P. N. Heacock, C. J. Clancey and W. Dowhan (1996). "The CDS1 Gene Encoding CDP-diacylglycerol Synthase In Saccharomyces cerevisiae Is Essential for Cell Growth." Journal of Biological Chemistry **271**(2): 789-795.

Shui, G., X. L. Guan, P. Gopalakrishnan, Y. Xue, J. S. Y. Goh, H. Yang and M. R. Wenk (2010). "Characterization of Substrate Preference for Slc1p and Cst26p in Saccharomyces cerevisiae Using Lipidomic Approaches and an LPAAT Activity Assay." PLoS ONE **5**(8): e11956.

Simeonidis, E., E. Murabito, K. Smallbone and H. Westerhoff (2010). "Why does yeast ferment? A flux balance analysis study." Biochemical Society Transactions **38**: 1225-1229.

Skoneczny, M. and J. Rytka (1996). "Maintenance of the peroxisomal compartment in glucose-repressed and anaerobically grown Saccharomyces cerevisiae cells." Biochimie **78**(2): 95-102.

Sorger, D. and G. Daum (2002). "Synthesis of Triacylglycerols by the Acyl-Coenzyme A:Diacyl-Glycerol Acyltransferase Dga1p in Lipid Particles of the Yeast Saccharomyces cerevisiae." Journal of Bacteriology **184**(2): 519-524.

Steinmetz, L. M., C. Scharfe, A. M. Deutschbauer, D. Mokranjac, Z. S. Herman, T. Jones, A. M. Chu, G. Giaever, H. Prokisch, P. J. Oefner and R. W. Davis (2002). "Systematic screen for human disease genes in yeast." Nat Genet **31**(4): 400-404.

Stephanopoulos, G., A. A. Aristidou and J. Nielsen (1998). Metabolic engineering: principles and methodologies. San Diego, Academic Press.

Strahl, T. and J. Thorner (2007). "Synthesis and function of membrane phosphoinositides in budding yeast, Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 353-404.

Summers, E. F., V. A. Letts, P. McGraw and S. A. Henry (1988). "Saccharomyces cerevisiae cho2 mutants are deficient in phospholipid methylation and cross-pathway regulation of inositol synthesis." Genetics **120**(4): 909-922.

Swainston, N., K. Smallbone, P. Mendes, D. B. Kell and N. W. Paton (2011). "The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks." Journal of integrative bioinformatics **8**(2): 186.

Tamaki, H., A. Shimada, Y. Ito, M. Ohya, J. Takase, M. Miyashita, H. Miyagawa, H. Nozaki, R. Nakayama and H. Kumagai (2007). "LPT1 Encodes a Membrane-bound O-Acyltransferase Involved in the Acylation of Lysophospholipids in the Yeast Saccharomyces cerevisiae." Journal of Biological Chemistry **282**(47): 34288-34298.

Taylor, F. R. and L. W. Parks (1979). "Triacylglycerol metabolism in Saccharomyces cerevisiae relation to phospholipid synthesis." Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **575**(2): 204-214.

Teusink, B., A. Wiersma, D. Molenaar, C. Francke, W. M. de Vos, R. J. Siezen and E. J. Smid (2006). "Analysis of Growth of Lactobacillus plantarum WCFS1 on a Complex Medium Using a Genome-scale Metabolic Model." Journal of Biological Chemistry **281**(52): 40041-40048.

Thiele, I. and B. Ø. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nat. Protocols **5**(1): 93-121.

Toke, D. A., W. L. Bennett, J. Oshiro, W.-I. Wu, D. R. Voelker and G. M. Carman (1998). "Isolation and Characterization of the Saccharomyces cerevisiae LPP1 Gene Encoding a Mg2+-independent Phosphatidate Phosphatase." Journal of Biological Chemistry **273**(23): 14331-14338.

Trotter, P. J. and D. R. Voelker (1995). "Identification of a Non-mitochondrial Phosphatidylserine Decarboxylase Activity (PSD2) in the Yeast Saccharomyces cerevisiae." Journal of Biological Chemistry **270**(11): 6062-6070.

Varma, A. and B. Ø. Palsson (1993). "Metabolic Capabilities of Escherichia coli II. Optimal Growth Patterns." Journal of Theoretical Biology **165**(4): 503-522.

Varma, A. and B. Ø. Palsson (1995). "Parametric sensitivity of stoichiometric flux balance models applied to wild-type Escherichia coli metabolism." Biotechnology and Bioengineering **45**(1): 69-79.

Veenhuis, M., M. Mateblowski, W. H. Kunau and W. Harder (1987). "Proliferation of microbodies in Saccharomyces cerevisiae." Yeast **3**(2): 77-84.

Wang, Z. and J. Zhang (2009). "Abundant Indispensable Redundancies in Cellular Metabolic Networks." Genome Biology and Evolution **1**: 23-33.

Wiback, S. J., R. Mahadevan and B. Ø. Palsson (2004). "Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum." Biotechnology and Bioengineering **86**(3): 317-331.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. o. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Véronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston and R. W. Davis (1999). "Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis." Science **285**(5429): 901-906.

Wu, W.-I., Y. Liu, B. Riedel, J. B. Wissing, A. S. Fischl and G. M. Carman (1996). "Purification and Characterization of Diacylglycerol Pyrophosphate Phosphatase from Saccharomyces cerevisiae." Journal of Biological Chemistry **271**(4): 1868-1876.

Zanghellini, J., K. Natter, C. Jungreuthmayer, A. Thalhammer, C. F. Kurat, G. Gogg-Fassolter, S. D. Kohlwein and H.-H. von Grünberg (2008). "Quantitative modeling of triacylglycerol homeostasis in yeast – metabolic requirement for lipolysis to promote membrane lipid synthesis and cellular growth." FEBS Journal **275**(22): 5552-5563.

Zheng, Z. and J. Zou (2001). "The Initial Step of the Glycerolipid Pathway." Journal of Biological Chemistry **276**(45): 41710-41716.

Zomorrodi, A. and C. Maranas (2010). "Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data." BMC Systems Biology **4**(1): 178.

Chapter 3: Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism†

3.1 Abstract

Genome-scale metabolic models are built using information from an organism's annotated genome and, correspondingly, information on reactions catalyzed by the set of metabolic enzymes encoded by the genome. These models have been successfully applied to guide metabolic engineering to increase production of metabolites of industrial interest. Congruity between simulated and experimental metabolic behavior is influenced by the accuracy of the representation of the metabolic network in the model.

In the interest of applying the consensus model of *Saccharomyces cerevisiae* metabolism for increased productivity of triglycerides, we manually evaluated the representation of fatty acid, glycerophospholipid, and glycerolipid metabolism in the consensus model (Yeast v6.0). These areas of metabolism were chosen due to their tightly interconnected nature to triglyceride synthesis. Manual curation was facilitated by custom MATLAB® functions that return information contained in the model for reactions associated with genes and metabolites within the stated areas of metabolism. Through manual curation, we have identified inconsistencies between information contained in the model and literature knowledge. These inconsistencies include incorrect gene-reaction associations, improper definition of substrates/products in

† This chapter has previously been published as: Aung, H. W., S. A. Henry and L. P. Walker (2013). "Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism." Industrial Biotechnology **9**(4): 215-228.

reactions, inappropriate assignments of reaction directionality, nonfunctional β-oxidation pathways, and missing reactions relevant to the synthesis and degradation of triglycerides. Suggestions to amend these inconsistencies in the Yeast v6.0 model can be implemented through the MATLAB script "YN6Edit" provided at online.liebertpub.com/doi/suppl/10.1089/ind.2013.0013/suppl_file/Supp_Data.zip.

3.2 Introduction

Biodiesel can serve as a renewable alternative to petroleum-derived diesel. The current major feedstock for commercial production of biodiesel in the United States is soybean oil (U.S. Environmental Protection Agency 2011). However, the use of soybean oil poses several issues such as the necessity for arable cropland, competition with food use of the oil, and high feedstock costs. Alternative feedstocks grown on marginal lands or generated from waste or agricultural and forest residues could address these issues. The microbial conversion of these alternative feedstocks into lipids has been demonstrated as being technically feasible (Kosa and Ragauskas 2011). Nonetheless, high lipid productivity will be essential for commercial production of biodiesel that is economically and environmentally sustainable.

Metabolic engineering can serve to increase productivity by driving more metabolic flux into lipids. A challenge in developing effective schemes for metabolic engineering is predicting the effects of genetic manipulations *a priori*. This is due to the complexities of the metabolic network (e.g. redundancy, high level of interconnectedness, presence of alternative metabolic routes, and control mechanisms). Thus, it is imperative that we move towards a systems approach to understanding the workings of cells. Genome-scale metabolic models can be used to predict or describe flux distributions throughout the entirety of a cell's known metabolic

network. By extension, these models can also be used to predict the outcome of genetic perturbations and to determine optimal engineering strategies (Oberhardt, Palsson et al. 2009).

An accurate representation of the metabolic network is crucial for simulation of cellular behavior. The endeavor to create an accurate representation is often an iterative process which requires repeated expansion or correction of the model based upon comparison of model predictions to experimental data and/or the discovery of errors in the model (Thiele and Palsson 2010). In this paper, we manually evaluated the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism (Yeast v6.0) (Herrgard, Swainston et al. 2008). These areas of metabolism were chosen due to the connection and coordination between glycerophospholipid and glycerolipid (e.g. triglyceride) metabolism and the function of fatty acids as precursors to both lipids. The description of these metabolic pathways in the model was compared to the knowledge described in scientific literature. Through this comparison, we identified biologically inconsistent representations (e.g. substrate specificity, compartmentalization of reactions, reaction directionality), missing connections between metabolites, and limitations in the lumped representation of metabolites. We have compiled a collection of suggested changes to the Yeast v6.0 model to contribute to the community approach to further develop a genome-scale reconstruction of yeast metabolism.

3.3 Methods

3.3.1 Yeast v6.0 Model

The Yeast v6.0 model was downloaded from http://yeast.sourceforge.net/. This SBML file was read into MATLAB® R2009a (MathWorks, Natick, MA) using the “readCbModel” function in the COBRA Toolbox 2.0.5 (Schellenberger, Que et al. 2011) via libSBML 5.6.0 (Bornstein, Keating et al. 2008) and the SBMLToolbox 4.1.0 (Keating, Bornstein et al. 2006).

3.3.2 Curation Process

To facilitate model investigation, MATLAB functions were developed to explore the reactions, metabolites, and genes contained in the model. The following custom functions were frequently utilized during the curation process: “metInfo”, “rxnInfo”, and “cgpr”. These functions extract information of interest from the COBRA-format model and display this information in a human-readable format. The function “metInfo” displays the names of all the reactions that a metabolite is involved with. “rxnInfo” displays the genes associated with a reaction, lower/upper flux bounds, and the metabolites involved in the reaction. The sequential use of “metInfo” and “rxnInfo” allows one to leapfrog throughout the model. For instance, “rxnInfo” can be used to learn information on a particular reaction and “metInfo” can be used to learn more about the metabolites in the reaction by generating a list of other reactions the metabolite participates in. This cycle can be repeated again for the newly generated list of reactions. The function “cgpr”, which stands for common gene name-protein-reaction relationship, displays information on the reactions associated with a gene. The common gene name utilized for this function should be consistent with the standard name specified in the Saccharomyces Genome Database (http://www.yeastgenome.org/). All three of these functions can be found at online.liebertpub.com/doi/suppl/10.1089/ind.2013.0013/suppl_file/Supp_Data.zip.

Examination of the reconstruction of fatty acid, glycerolipid, and glycerophospholipid metabolism was initiated by first defining the set of genes involved in this area of metabolism. A recent review article served as a source for generating this list of genes (Henry, Kohlwein et al. 2012). The “cgpr” function was utilized to compare the information contained in the model with current knowledge from scientific literature for each of the genes of interest. To ensure

consideration of reactions not properly annotated with the appropriate gene(s) in the model, the model was also queried using metabolites as the search basis using the functions “metInfo” and “rxnInfo”.

Discrepancies between the model and literature evidence were addressed by revising the COBRA-format Yeast v6.0 model using functions contained in the COBRA toolbox. The function “changeGeneAssociation” was used to amend false gene-reaction relationships in the model. “changeRxnBounds” was used to change the lower and/or upper bound constraints for reaction flux. “addReaction” was used to add reactions for catalytic activity and for transport of metabolites between compartments. “removeRxns” was typically used in conjunction with “addReaction” to remove the previous representation of a reaction and to replace with an updated representation. The collection of all changes made to the COBRA-format model was documented in a MATLAB code that when executed applies all the changes.

3.3.3 Determining Blocked Reactions and Essential Genes

Blocked reactions and essential genes in the model were determined *in silico* using an updated version of the “testYeastModel” MATLAB function included in the Yeast 5 paper (Heavner, Smallbone et al. 2012). Blocked reactions were identified using flux variability analysis to search for reactions incapable of carrying fluxes under any media conditions (i.e. the flux bounds on all exchange reactions are relaxed to allow for the unconstrained uptake of all the extracellular metabolites included in the model). Essential genes were predicted by screening all possible gene knockout strains for inability to produce biomass at or above the growth rate threshold of 10^{-6} hr^{-1} in glucose minimal media; this simulation was performed using flux balance analysis. The solver used for both flux balance analysis and flux variability analysis was Gurobi 5.0.2 (Gurobi Optimization, Houston, TX).

3.4 Results

3.4.1 Revised Representation of Fatty Acid Metabolism

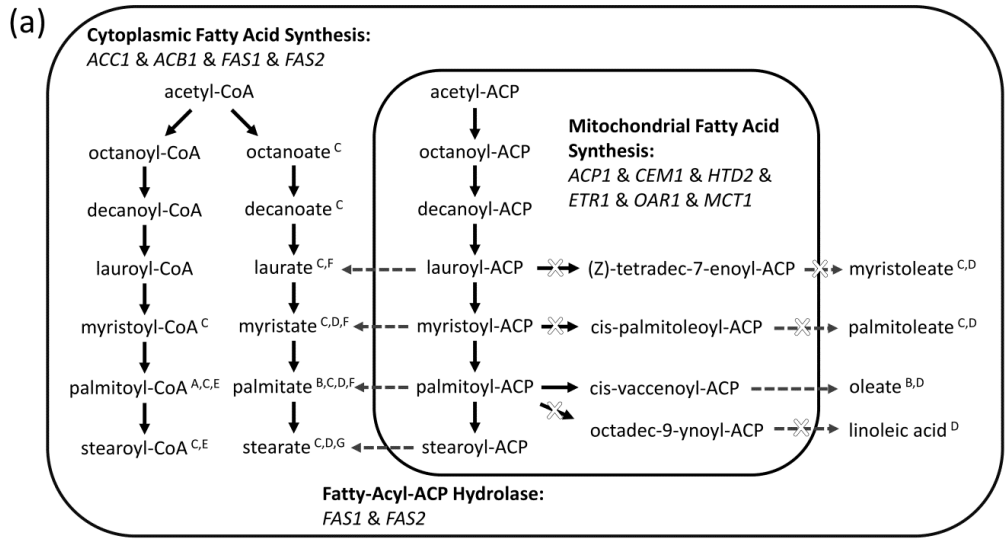
3.4.1.1 Changes to representation of fatty acid synthesis

Fatty acids serve as building blocks for membrane lipids and as stores of chemical energy. The profile of fatty acids found in *S. cerevisiae* is mostly dominated by C16:0, C16:1, C18:0, and C18:1 fatty acids; minor amounts of C14:0 and very long-chain fatty acids, such as C26:0, are also present (Henry, Kohlwein et al. 2012). Cellular fatty acids can be obtained through uptake from media, *de novo* synthesis, and lipid turnover (Tehlivets, Scheuringer et al. 2007). In this section, changes made to the Yeast v6.0 model that are relevant to *de novo* fatty acid synthesis are discussed.

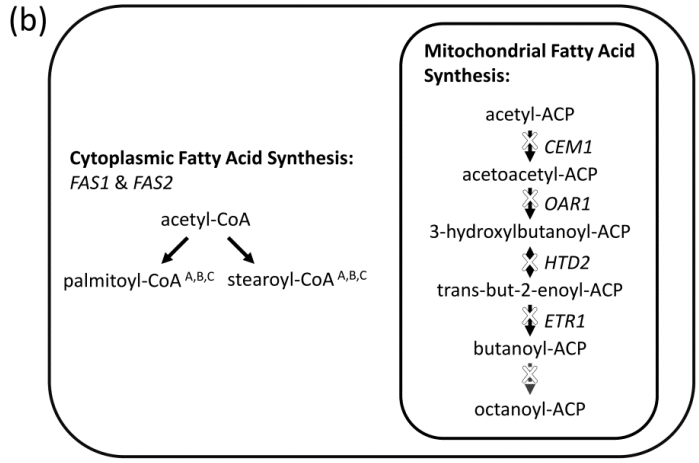
S. cerevisiae has two fatty acid synthase (FAS) systems to generate fatty acids through successive additions of two carbons to a starting acetyl moiety. Cytoplasmic FAS is a complex composed of two subunits, Fas1p and Fas2p, which releases C16:0 and C18:0 fatty acids esterified to coenzyme A (Lomakin, Xiong et al. 2007). In contrast, mitochondrial fatty acid synthesis is achieved by sequential action of discrete, individual enzymes and generates C8:0 fatty acids bound to an acyl carrier protein (ACP) (Hiltunen, Schonauer et al. 2009). The enzymes involved in this process are Acp1p, acyl-carrier protein; Mct1p, malonyl-CoA:ACP transferase; Cem1p, β-ketoacyl-ACP synthase; Oar1p, 3-oxoacyl-ACP reductase; Htd2p, 3-hydroxyacyl-thioester dehydratase; and Etr1p, enoyl-ACP reductase (Hiltunen, Schonauer et al. 2009). Although there is evidence to suggest *in vitro* catalytic capacity for longer fatty acids, *in vivo* evidence and understanding of the roles of mitochondrial FAS products with chain lengths greater than 8 carbons is lacking (Hiltunen, Schonauer et al. 2009).

The Yeast v6.0 model depicts cytoplasmic fatty acid synthesis with a lumped reaction for the extension of acetyl-CoA (C2:0) to 8 carbons and individual reactions for the addition of two carbons going from 8 carbons up to 18 carbons in length (Figure 3.1a). This representation is misleading since the reaction intermediates are shuttled within the complex and are not released into the cytosol *in vivo*. Cytoplasmic FAS is also modeled as being capable of utilizing and producing both free fatty acids and fatty acids esterified to coenzyme A. The erroneous inclusion of free fatty acids as products in the model may have arisen from comparison to mammalian cells (Tehlivets, Scheuringer et al. 2007). In order to amend both issues, the aforementioned reactions could be replaced with two reactions describing the net equation for generation of the dominant products of cytoplasmic FAS (i.e. palmitoyl-CoA (C16:0) and stearoyl-CoA (C18:0)) from acetyl-CoA (Figure 3.1b).

In the Yeast v6.0 model, mitochondrial fatty acid synthesis is depicted similarly to cytoplasmic fatty acid synthesis which allows for the mitochondrial system to create products of the same carbon lengths as the cytoplasmic system (Figure 3.1a). Acyl-ACP's from mitochondrial FAS are portrayed as being able to be transported to the cytoplasm where they can be hydrolyzed to free fatty acids through action of Fas1p and Fas2p. Yeast v6.0 also includes reactions to generate unsaturated acyl chains of 14, 16, and 18 carbon lengths through mitochondrial FAS. Given that the utility and destination of mitochondrial FAS products besides octanoyl-ACP is unknown (Hiltunen, Schonauer et al. 2009), the reactions for longer acyl-chains may be considered inappropriate (Figure 3.1b). Octanoyl-ACP serves as a precursor for lipoic acid, an essential cofactor in oxidative decarboxylation reactions (Tehlivets, Scheuringer et al. 2007). However, it should be noted that gaps remain in the reconstructed pathway for *de novo* synthesis of lipoic acid.



A: Transported to ER where it is then classified as an acyl-CoA through an "isa acyl-CoA" reaction; "acyl-CoA" can be used to acylate glycerol 3-phosphate, dihydroxyacetone phosphate, or 1-acyl-sn-glycerol 3-phosphate
B: Activated with CoA and subsequent transport of acyl-CoA to ER where it is then classified as an acyl-CoA through an "isa acyl-CoA" reaction
C: Transported to peroxisome
D: Activated with CoA and subsequent transport of acyl-CoA to peroxisome
E: Desaturated in the cytoplasm
F: Elongated by 2 carbons and desaturated in the cytoplasm
G: Elongated to lignoceric acid (C24:0)



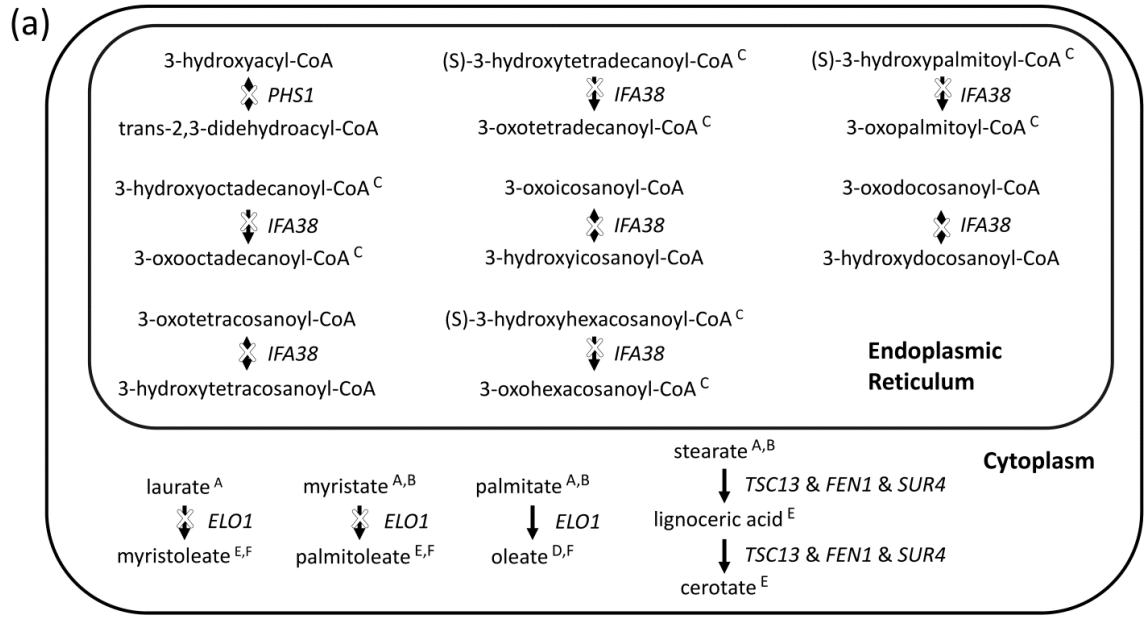
A: Transported to ER membrane (where it can be elongated; desaturated; or used to acylate glycerol 3-phosphate, dihydroxyacetone phosphate, or diglyceride)
B: Transported to lipid particle (where it can be used to acylate glycerol 3-phosphate, dihydroxyacetone phosphate, or diglyceride)
C: Transported to peroxisome

Figure 3.1: Curation of cytoplasmic and mitochondrial fatty acid synthesis (FAS) in the Yeast v6.0 model. (a) Reactions in the Yeast v6.0 model relevant to cytoplasmic and mitochondrial FAS. Dashed arrows indicate the net outcome of two reactions (i.e. transport and hydrolysis). Mitochondrial FAS reactions leading to production of myristoleate, palmitoleate, and linoleic acid are blocked reactions, as indicated by X's in the figure. (b) Proposed representation of cytoplasmic and mitochondrial FAS. For the sake of space, the repeated reaction sequence of *CEM1*, *OAR1*, *HTD2*, and *ETR1* for butanoyl-ACP (C4:0) to octanoyl-ACP is omitted in the figure, as indicated by the dashed arrow. The reactions for mitochondrial FAS are blocked reactions, as indicated by X's in the figure.

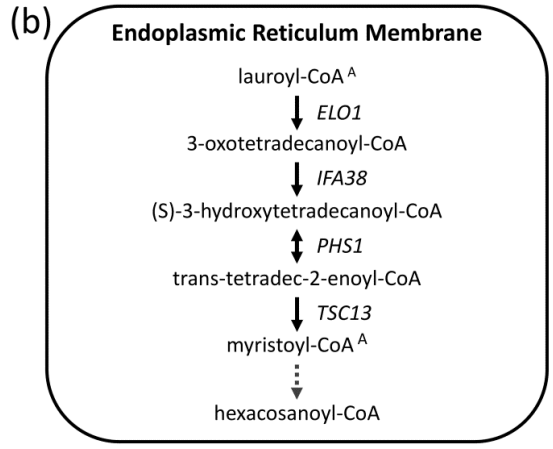
3.4.1.2 Changes to representation of fatty acid elongation

Another set of enzymes are needed to produce acyl chains longer than 18 carbons. These very long-chain fatty acids are synthesized through reaction mechanisms similar to fatty acid synthesis; the difference is that medium to very long-chain acyl-CoA's serve as substrates for addition of two-carbon units. Fatty acid elongation is catalyzed by the sequential action of the following ER membrane-localized enzymes: elongases, Elo1p, Fen1p, or Sur4p (Toke and Martin 1996; Oh, Toke et al. 1997); β-ketoacyl-CoA reductase, Ifa38p (Han, Gable et al. 2002); β-hydroxyacyl-CoA dehydratase, Phs1p (Denic and Weissman 2007); and enoyl-CoA reductase, Tsc13p (Kohlwein, Eder et al. 2001; Paul, Gable et al. 2007).

The Yeast v6.0 model describes fatty acid elongation in a biologically inconsistent manner. The model currently has seven reactions depicting the activity of Ifa38p on various metabolites; a reaction describing the activity of Phs1p on the generic metabolite "3-hydroxyacyl-CoA"; three reactions for the elongation and desaturation of free fatty acids from C12:0 to C14:1, C14:0 to C16:1, and C16:0 to C18:1 that are associated with *ELO1*; and two reactions for the elongation of free fatty acids from C18:0 to C24:0 and C24:0 to C26:0 that are associated with "*TSC13* and *FEN1* and *SUR4*" (Figure 3.2a). The current representation does not capture the interplay between the different enzymes involved in fatty acid elongation. For instance, the individual reactions for *IFA38* and for *PHS1* are not connected to the rest of fatty acid elongation since the other reactions consider net reactions as opposed to including intermediate steps. Evaluation of the net reactions reveals several inaccuracies. The gene associations for these reactions are incomplete since β-ketoacyl-CoA reductase and β-hydroxyacyl-CoA dehydratase are also needed to catalyze the net reaction. Another error is that the elongation process should utilize and generate fatty acids esterified to coenzyme A, not free



A: Comes from cytoplasmic or mitochondrial fatty acid synthesis
B: Comes from transport from the extracellular compartment
C: Classified as an acyl-CoA through an "isa acyl-CoA" reaction; "acyl-CoA" can be used to acylate glycerol 3-phosphate, dihydroxyacetone phosphate, or 1-acyl-sn-glycerol 3-phosphate
D: Activated with CoA and subsequent transport of acyl-CoA to ER where it is then classified as an acyl-CoA through an "isa acyl-CoA" reaction
E: Transported to peroxisome
F: Activated with CoA and subsequent transport of acyl-CoA to peroxisome



A: Comes from activation of fatty acid originating from the extracellular compartment

Figure 3.2: Curation of fatty acid elongation in the Yeast v6.0 model. (a) Reactions in the Yeast v6.0 model relevant to fatty acid elongation. As indicated by X's in the figure, all of the reactions associated with *PHS1* or *IFA38* are blocked and two of the three reactions associated with *ELO1* are blocked. (b) Proposed representation of fatty acid elongation. For the sake of space, the repeated reaction sequence of (*ELO1*, *FEN1*, or *SUR4*), *IFA38*, *PHS1*, and *TSC13* for myristoyl-CoA (C14:0) to hexacosanoyl-CoA (C26:0) is omitted in the figure, as indicated by the dashed arrow. For this omission, the embedded intermediates of palmitoyl-CoA (C16:0) and stearoyl-CoA (C18:0) are assumed to be also derived from activation of fatty acid originating from the extracellular compartment or from cytoplasmic fatty acid synthesis.

fatty acids. In addition to the two issues discussed above, Elo1p is not able to introduce a double bond to the fatty acid chain.

A solution to address the aforementioned issues is to remove the desaturation reactions associated with *ELO1*, to expand upon all the steps involved in the elongation process with each individual gene associated with the appropriate reaction, and to use acyl-CoA species, instead of free fatty acids, in the reactions (Figure 3.2b). This modification removes dead-ends in the model by improving metabolite connections and ensuring that the products of one reaction would serve as the reactants in another reaction. The expansion of the individual reaction steps allows for the description of the acyl-CoA specificity of the different elongase enzymes. The elongases have different reactant preferences and also differ in the length of the acyl-chain produced by the elongation cycle (Rössler, Rieck et al. 2003; Denic and Weissman 2007). The proposed modification assumes that Elo1p acts upon C12:0 and C14:0 acyl-CoA; Fen1p acts upon even length acyl-CoA's from C16:0 to C22:0; and Sur4p acts upon even length acyl-CoA's from C18:0 to C24:0. Overall, the expanded representation allows for resolution of the individual reactions associated with each gene.

3.4.1.3 Changes to representation of fatty acid desaturation

The production of C16:1 and C18:1 fatty acids is catalyzed by the Δ9-desaturase Ole1p, which catalyzes the insertion of a double bond between carbons 9 and 10. Ole1p acts upon palmitoyl-CoA (C16:0) to form palmitoleoyl-CoA (C16:1) and upon stearoyl-CoA (C18:0) to from oleoyl-CoA (C18:1) (Stukey, McDonough et al. 1990). As described in prior sections, reactions that produce monounsaturated fatty acids through mitochondrial fatty acid synthesis or fatty acid elongation are removed from the model after applying the suggested changes for those processes. This modification restricts *de novo* synthesis of C16:1 and C18:1Δ^9 fatty acids to

occur only through Ole1p. The two reactions for *OLE1* in the Yeast v6.0 model also have reconstruction errors. The introduction of a double bond in the fatty acid chain requires reducing equivalents from NADH (Martin, Oh et al. 2007). NADH/NAD+ cofactors for these reactions are not included in the Yeast v6.0 model. In addition, the product of desaturation of palmitoyl-CoA should be corrected to be palmitoleoyl-CoA, instead of hexadec-2-enoyl-CoA. Although both palmitoleoyl-CoA and hexadec-2-enoyl-CoA contain C16:1 fatty acids, the location of the double bond differs between the two metabolites.

3.4.1.4 Changes to repair blocked reactions in β-Oxidation

Fatty acids can be broken down through a process called β-oxidation. Each round of β-oxidation removes two carbons from the fatty acid chain in the form of acetyl-CoA. This acetyl-CoA can be utilized for energy production through the citric acid cycle and for carbohydrate biosynthesis through the glyoxylate cycle (Hiltunen, Mursula et al. 2003). For *S. cerevisiae*, β-oxidation occurs only in peroxisomes (Hiltunen, Mursula et al. 2003). Since the peroxisomal membrane is impermeable to NAD(H) and acetyl-CoA, there must exist mechanisms to regenerate NAD+ for continued β-oxidation and to transport acetyl-CoA to mitochondria for energy production (van Roermund, Elgersma et al. 1995). Attention to this issue in curation of the model led to corrections that allow for simulated growth on fatty acids as sole carbon sources, which is reflective of *in vivo* metabolic capacity of *S. cerevisiae*. A visual comparison of the differences in depiction of β-oxidation before and after our curation is shown in Figure 3.3.

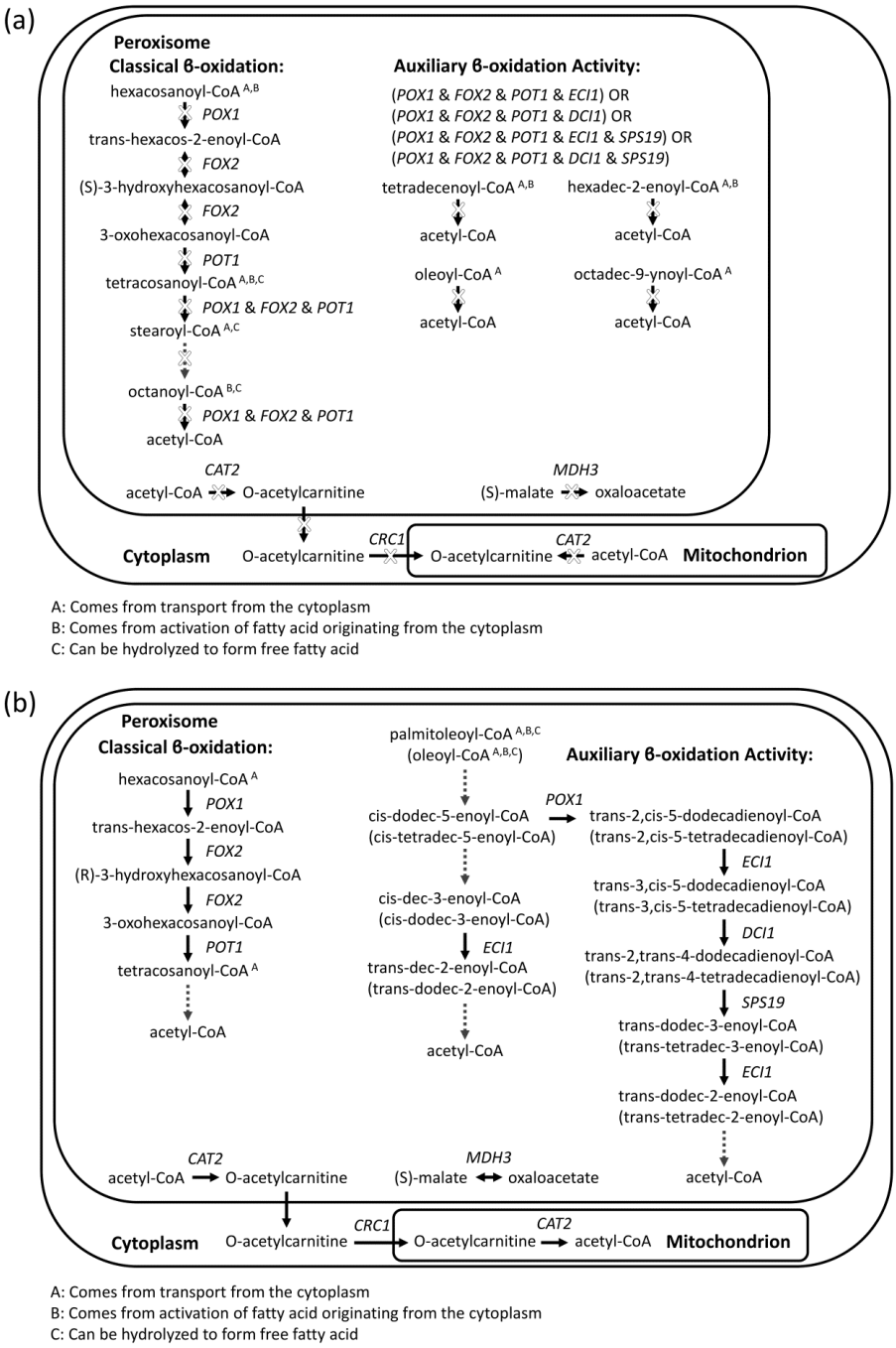


Figure 3.3: Curation of β-oxidation in the Yeast v6.0 model. (a) Reactions in the Yeast v6.0 model relevant to β-oxidation of saturated and unsaturated fatty acids. All of the reactions shown are blocked, as indicated by X's in the figure. For the sake of space, the repeated reaction sequence of *POX1*, *FOX2*, and *POT1* for stearoyl-CoA (C18:0) to octanoyl-CoA (C8:0) is omitted in the figure, as indicated by the dashed arrow. (b) Proposed representation of β-oxidation of saturated and unsaturated fatty acids. The metabolites relevant to β-oxidation of oleoyl-CoA are shown in parentheses in the figure and are written below the metabolites relevant to β-oxidation of palmitoleoyl-CoA. For the sake of space, dashed arrows are used in this figure to condense the combined action of the enzymes of classical β-oxidation (i.e. Pox1p, Fox2p, and Pot1p) into a singular illustrated reaction.

During β-oxidation, NAD+ is needed for the step catalyzed by *FOX2*-encoded 3-hydroxyacyl-CoA dehydrogenase (Hiltunen, Mursula et al. 2003). The NAD+ used for β-oxidation is able to be regenerated through action of *MDH3*-encoded malate dehydrogenase which reduces oxaloacetate into malate in an NADH-dependent manner (van Roermund, Elgersma et al. 1995). The Yeast v6.0 model has this reaction constrained strictly in the direction of conversion of malate to oxaloacetate. This direction constraint therefore prevents NAD+ from being able to be regenerated in the peroxisome. This can be remedied by making the malate dehydrogenase reaction reversible.

Since the acetyl-CoA end-product of β-oxidation is unable to diffuse across the peroxisomal membrane, *S. cerevisiae* utilizes two different pathways to utilize this metabolite (van Roermund, Elgersma et al. 1995). One pathway is through the glyoxylate cycle which yields succinate from two molecules of acetyl-CoA. Another pathway is through the carnitine shuttle. The carnitine acetyl-CoA transferase, Cat2p, is involved in transfer of acetyl units from the peroxisome to the mitochondria. Acetyl-CoA is converted to acetylcarnitine in the peroxisome. Acetylcarnitine can be transported to the mitochondria and subsequently the acetyl group can be transferred to a molecule of free coenzyme A for further metabolism in the TCA cycle. The Yeast v6.0 model represents both peroxisomal and mitochondrial carnitine acetyl-CoA transferase reactions in the same way i.e. (R)-carnitine + acetyl-CoA → coenzyme A + O-acetylcarnitine. This representation does not allow for mitochondrial production of acetyl-CoA from acetylcarnitine. Therefore, the mitochondrial *CAT2* reaction can be revised by switching the products/reactants.

3.4.2 Revised Representation of Glycerolipid and Glycerophospholipid Metabolism

3.4.2.1 Addition of new genes

Following manual curation, 15 additional genes were identified for addition to the Yeast v6.0 model (Table 3.1). Eleven of these genes introduce the potential for new catalytic activity. The addition of *AYR1* to the model ultimately allows for formation of phosphatidate from the substrate of dihydroxyacetone phosphate (Athenstaedt and Daum 2000) (Figure 3.4b). Phosphatidate is a key intermediate in the synthesis of glycerophospholipids and triglycerides. With the inclusion of *AYR1*, the two different pathways for phosphatidate biosynthesis (i.e. glycerol 3-phosphate pathway and dihydroxyacetone phosphate pathway) can be both accounted for in the model.

Table 3.1: List of new genes to be added to the Yeast v6.0 model.

New gene additions that catalyze reactions currently not included in the model	
AYR1	NADPH-dependent 1-acyl dihydroxyacetone phosphate reductase
CLD1	Mitochondrial cardiolipin-specific phospholipase
CST26	Protein that incorporates stearic acid into phosphatidylinositol
DGK1	Diacylglycerol kinase
GCY1	Putative NADP(+) coupled glycerol dehydrogenase
GEP4	Mitochondrial phosphatidylglycerophosphatase
IDP2	Cytosolic NADP-specific isocitrate dehydrogenase
LRO1	Acyltransferase that catalyzes diacylglycerol esterification
PAH1	Mg^{2+}-dependent phosphatidate (PA) phosphatase
PHM8	Lysophosphatidate (LPA) phosphatase
YJU3	Monoglyceride lipase
New gene additions that catalyze reactions already present in the model	
FRQ1	May have a role in intracellular signaling through its regulation of Pik1p
LDH1	Exhibits active esterase plus weak triacylglycerol lipase activities
VAC14	Protein involved in regulated synthesis of $PI(3,5)P_2$; interacts with Fig4p; activator of Fab1p
VPS15	Functions as a membrane-associated complex with Vps34p

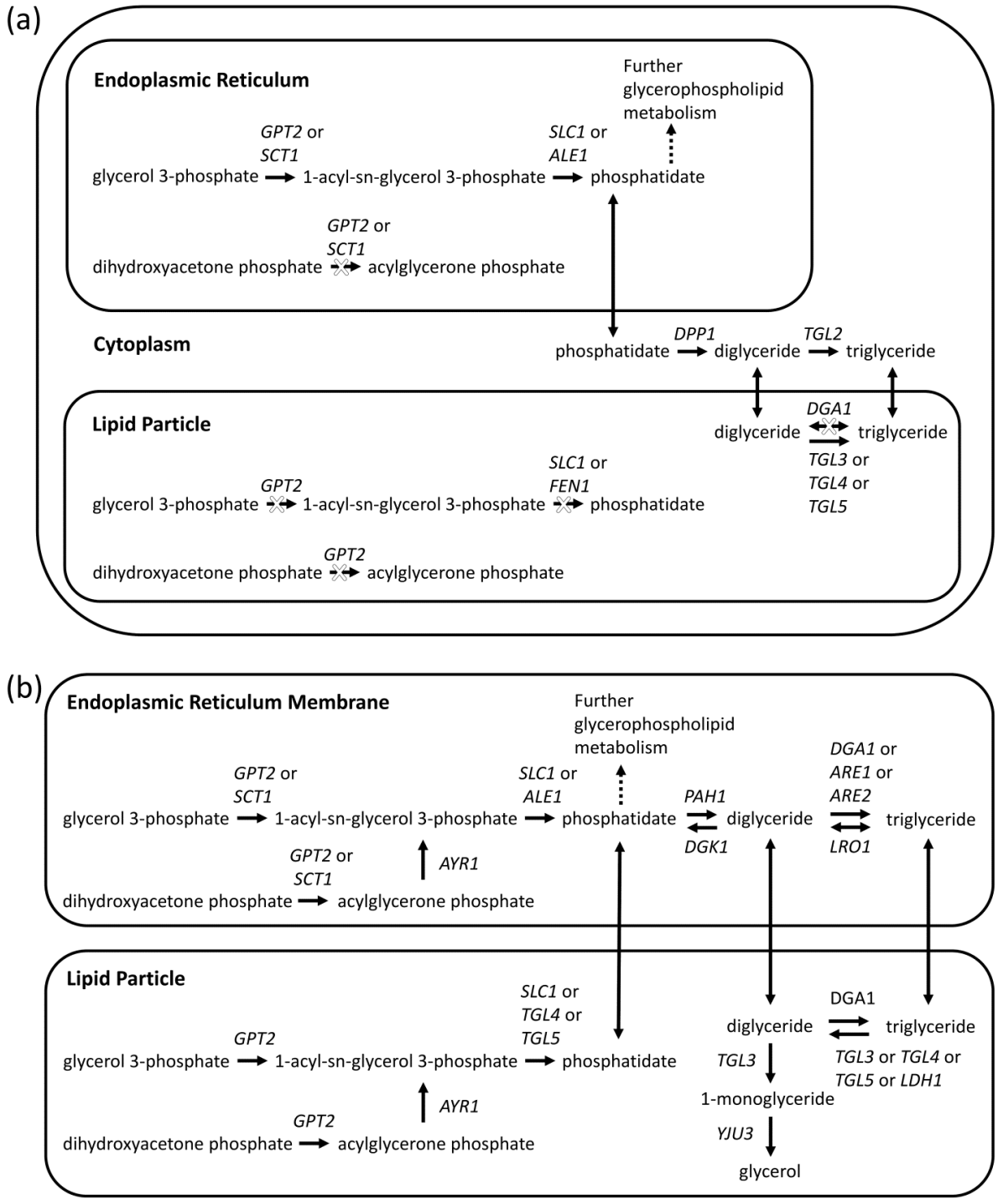


Figure 3.4: Curation of glycerolipid metabolism in the Yeast v6.0 model. (a) Reactions in the Yeast v6.0 model relevant to glycerolipid metabolism. Blocked reactions are indicated by X's in the figure. (b) Proposed representation of glycerolipid metabolism. The proposed modification expands each of the classes shown in this figure into its constituent species (not shown in figure). For instance, the term phosphatidate in the figure refers to a collection of individual species e.g., phosphatidate (1-16:0, 2-16:1), phosphatidate (1-16:1, 2-16:1), etc.

The genes *PAH1* and *LRO1*, which are relevant to triglyceride synthesis, can also be added to the model. The enzyme encoded by *PAH1* dephosphorylates phosphatidate to yield diglyceride which can be utilized in the synthesis of triglyceride or the glycerophospholipids phosphatidylethanolamine and phosphatidylcholine (Han, Wu et al. 2006). The Yeast v6.0 model associates the conversion of phosphatidate to diglyceride with the gene *DPP1* (Figure 3.4a). Although Dpp1p does act upon phosphatidate to yield diglyceride, this activity is believed to be involved in lipid signaling through the regulation of the amounts of phosphatidate and diacylglycerol pyrophosphate present in the vacuolar membrane as opposed to generating diglyceride for *de novo* synthesis of glycerophospholipids and triglyceride (Carman and Han 2006). Thus, the model can be amended to have Pah1p produce diglyceride at the ER membrane for the synthesis of triglyceride and the glycerophospholipids phosphatidylethanolamine and phosphatidylcholine (Figure 3.4b). The phosphatidate phosphatase activity associated with Dpp1p can be relocated to the vacuolar membrane in the model and kept separate from the *de novo* synthesis pathways.

The synthesis of triglyceride from diglyceride can occur through either a mechanism using acyl-CoA or glycerophospholipids as acyl donors. The four enzymes that account for all triglyceride synthesis are Dga1p, Lro1p, Are1p, and Are2p. Dga1p catalyzes the majority of the acyl-CoA-dependent diglyceride acyltransferase activity while Are1p and Are2p provide only minor activity (Oelkers, Cromley et al. 2002; Sandager, Gustavsson et al. 2002; Sorger and Daum 2002). Lro1p catalyzes the transfer of the acyl group at the sn-2 position of phosphatidylethanolamine or phosphatidylcholine to diglyceride (Dahlqvist, Ståhl et al. 2000; Oelkers, Tinkelenberg et al. 2000). The Yeast v6.0 model associates the conversion of diglyceride to triglyceride with the genes *DGA1*, *TGL2*, *TGL3*, *TGL4*, and *TGL5* (Figure 3.4a), instead of the

genes *DGA1*, *LRO1*, *ARE1*, and *ARE2* (Figure 3.4b). The association of the *TGL* (triacylglycerol lipase) genes with this activity is erroneous since these genes enable the breakdown, not synthesis, of triglyceride. Thus, the reaction directionality for the *TGL* genes should be changed so that it goes irreversibly towards triglyceride hydrolysis. The genes associated with triglyceride lipase activity should also be amended to be *TGL3*, *TGL4*, *TGL5*, and *LDH1* (Figure 3.4b). *TGL2* is not included in this association since triglyceride has not been observed at the mitochondria, which is where Tgl2p is localized (Ham, Rho et al. 2010; Grillitsch and Daum 2011). *LDH1* is a new gene to the model and is added based on recent evidence (Debelyy, Thoms et al. 2011).

Additional genes and reactions connected to triglyceride breakdown should be included in the model (Figure 3.4b). The inclusion of *DGK1* allows for utilization of the diglyceride generated from triglyceride hydrolysis. *DGK1* encodes for diacylglycerol kinase which converts diglyceride into phosphatidate for use in glycerophospholipid synthesis (Fakas, Konstantinou et al. 2011). The pathways for triglyceride degradation in the model should also be supplemented with reactions for the complete hydrolysis of triglyceride into glycerol and free fatty acids. This can be accomplished by the addition of *TGL3*-encoded diglyceride lipase activity (Kurat, Natter et al. 2006) and *YJU3*-encoded monoglyceride lipase activity (Heier, Taschler et al. 2010) to the model.

3.4.2.2 Expansion of species

Glycerolipids and glycerophospholipids can vary by the fatty acyl chain attached to the glycerol backbone of the molecule. The acyl chains can differ in chain length and/or the number of double bonds. These differences are influential in altering the physical properties of the biological membrane with respect to membrane thickness, intrinsic curvature, and fluidity (de Kroon 2007). These physical properties can in turn affect the membrane's permeability, the

activity of membrane associated enzymes, and membrane fusion and fission (de Kroon 2007). Furthermore, adjustment of the acyl chain composition is utilized by the cell to adapt to different conditions and has also been proposed as a means to increase tolerance to stress conditions (You, Rosenfield et al. 2003; Rodríguez-Vargas, Sánchez-García et al. 2007). Theoretically, the number of glycerolipid and glycerophospholipid species that can exist is extremely large due to the vast permutations possible for different acyl chains positioned along the glycerol backbone. Thus, a modeling strategy to represent this structural diversity is needed and is explored in the following discussion.

Most of the models prior to the Yeast consensus model (Herrgard, Swainston et al. 2008), i.e. iFF708 (Förster, Famili et al. 2003), iLL672 (Kuepfer, Sauer et al. 2005), iMM904 (Mo, Palsson et al. 2009), and iND750 (Duarte, Herrgård et al. 2004), represent the glycerolipid and glycerophospholipid classes as defined composites of specific species. For instance, in iND750, the glycerophospholipid class phosphatidate is given a singular defined molecular formula of $C_{3540}H_{6544}O_{800}P_{100}$ in the model‡. This formula arises from the stoichiometry of the various acyl-CoA's used in the two reactions describing the successive acylation of glycerol-3-phosphate. As an example, the first acylation is described in the iND750 model as:

Glycerol 3-phosphate + 0.02 Decanoyl-CoA (n-C10:0CoA) + 0.06 Dodecanoyl-CoA (n-C12:0CoA) + 0.17 Hexadecenoyl-CoA (n-C16:1CoA) + 0.09 Octadecynoyl-CoA (n-C18:2CoA) + 0.24 Octadecenoyl-CoA (n-C18:1CoA) + 0.27 Palmitoyl-CoA (n-C16:0CoA) + 0.05 Stearoyl-CoA (n-C18:0CoA) + 0.1 Tetradecanoyl-CoA (n-C14:0CoA) -> 0.01 1-Acyl-sn-glycerol 3-phosphate + Coenzyme A

This reaction in iND750 is an abstraction that depicts the creation of 1-acyl-sn-glycerol 3-phosphate through fixed fractional contribution of various acyl-CoA species in acylating the sn-1 position of glycerol 3-phosphate. The advantages of this approach are its succinctness and its

‡ Describes 100 molecules of phosphatidate. Reactions using this metabolite have the stoichiometric coefficient scaled by 1/100.

incorporation of information on reaction specificity through different stoichiometries for each species. Conversely, this approach implements rigid stoichiometries and imposes a requirement for all the species listed in the reaction equation. Thus, this approach does not capture the flexibility of the cell's lipidome and the possibility for remodeling of individual lipids because of its restrictive molecular formulas.

In contrast, the approach taken by the Yeast consensus model is to categorize individual specific species into general classes using "isa" reactions (e.g. "isa acyl-CoA": hexacosanoyl-CoA -> acyl-CoA) and to subsequently use the term for the general class throughout metabolic reactions in the model. For example, the reaction described above for iND750 is represented in the Yeast v6.0 model as:

Glycerol 3-phosphate + Acyl-CoA -> 1-Acyl-sn-glycerol 3-phosphate + Coenzyme A + 2 H+

The use of "isa acyl-CoA" reactions in defining the term "acyl-CoA" allows for logical OR operation in which any of the categorized individual acyl-CoA species can satisfy the "acyl-CoA" requirement in the reaction. The 1-acyl-sn-glycerol 3-phosphate produced in this reaction is consequently a generic term in the model since it is created using the generic "acyl-CoA" term.

The disadvantage of strictly using generic terms defined by "isa" reactions is that information on individual reaction specificity is lost. For instance, if a particular reaction only uses C16:0- and C16:1-CoA, this specificity cannot be indicated using the acyl-CoA generic term. Furthermore, species that would not be biochemically utilized in the reaction could be utilized in computer simulations due to this generic term. Another disadvantage is backward incompatibility in going from the generic term back to specific species. An illustration of this is the breakdown of triglyceride to yield free fatty acids. With the usage of the generic "fatty acid"

term, the exact molecular formula is unknown and thus the potential number of β-oxidation cycles that can occur cannot be determined.

Another approach to modeling the diversity of lipid species is to expand the general classes into their corresponding individual species and to utilize the appropriate species, as opposed to general class, in the equations for metabolic reactions. This approach has recently been applied to sphingolipid metabolism in the Yeast consensus model (Heavner, Henry et al. 2012) and has been continued in this paper for glycerolipid and glycerophospholipid metabolism. The general classes of glycerolipids and glycerophospholipids can be expanded to clarify the specific acyl group(s) attached to the glycerol backbone. This expansion can consider the dominant acyl species found in *S. cerevisiae* i.e. C16:0, C16:1, C18:0, and C18:1 (Henry, Kohlwein et al. 2012). An additional assumption can be applied to restrict the acyl chain found at the sn-2 position of the glycerol backbone to be either C16:1 or C18:1 (Wagner and Paltauf 1994). As an example, phosphatidylcholine can be expanded out to 8 different species: phosphatidylcholine (1-16:0, 2-16:1), (1-16:1, 2-16:1), (1-18:0, 2-16:1), (1-18:1, 2-16:1), (1-16:0, 2-18:1), (1-16:1, 2-18:1), (1-18:0, 2-18:1), and (1-18:1, 2-18:1). A major detriment of this representation is that it greatly increases the number of reactions, which may be deemed unnecessary for the modeling scope of certain users. However, incorporation of this level of detail will provide a platform for users interested in accounting for distribution and function of distinct molecular lipid species.

3.4.2.3 Compartmentalization

Multiple compartments, each housing specific enzymes, contribute to lipid synthesis in *S. cerevisiae* (Henry, Kohlwein et al. 2012). The catalogue of lipids in various membranes is a result of both local metabolism and membrane trafficking between different compartments. For

instance, most of the steps in synthesis of phosphatidylcholine through the CDP-diacylglycerol pathway occur at the ER membrane. However, the major source of the intermediate phosphatidylethanolamine comes from activity at the mitochondrial membrane. Consequently, lipid molecules must be transferred between different membranes in order to fulfill a sequence of metabolic reactions.

The Yeast v6.0 model classifies reactions and species into 16 compartments: boundary, cell envelope, cytoplasm, endoplasmic reticulum, endoplasmic reticulum membrane, extracellular, Golgi, Golgi membrane, lipid particle, mitochondrion, mitochondrial membrane, nucleus, peroxisome, peroxisomal membrane, vacuole, and vacuolar membrane. In classifying reactions involved in lipid metabolism, the Yeast v6.0 model opts to not localize the reactions to membrane compartments (e.g. localized to endoplasmic reticulum instead of endoplasmic reticulum membrane). Reactions in glycerolipid and glycerophospholipid metabolism could be updated to indicate if they occur at the membrane. The utilization of membrane compartments in the model provides more resolution on reaction localization. This change also requires the addition of transport reactions to allow transport of water soluble lipid precursors from the cytoplasm to the membrane and to allow reversible transfer of lipids between membrane compartments. The addition of these transport reactions can be based on modeling decisions to fill gaps within the model. As the interplay of organelles in lipid metabolism is further elucidated, these modeling decisions may be readdressed.

3.4.3 Blocked Reactions

Blocked reactions are defined as reactions that are incapable of carrying flux while still satisfying the constraints of flux bounds and the steady-state assumption. Upon implementing all of the suggested changes to the Yeast v6.0 model (see MATLAB script “YN6Edit.m” at

online.liebertpub.com/doi/suppl/10.1089/ind.2013.0013/suppl_file/Supp_Data.zip), the absolute number of blocked reactions actually increases by 8 but the relative number of blocked reactions decreases from 39.2% to 21.3% of all reactions (Table 3.2). This contrast arises from the large increase in the number of reactions in the model due to the expansion of classes into the individual species that can participate in various reactions. This expansion consequently biased the metric of percent blocked reactions to be more of an indicator of the topology of fatty acid, glycerolipid, and glycerophospholipid metabolism. Nevertheless, comparison between the set of blocked reactions for the model before and after modifications can allow assessment of gaps that were filled in the network and of areas of metabolism that can serve as focus for future model development.

Table 3.2: Statistics for the Yeast v6.0 model before and after implementing the changes suggested through manual curation.

Yeast v6.0 Model	Before Changes	After Changes
Genes	901	916
Reactions	1882	3493
Metabolites	1454	2218
Blocked Reactions	737 (39.2%)	745 (21.3%)

The major source of reductions in the number of blocked reactions was due to changes relevant to β-oxidation (Figure 3.3). These changes not only allowed for fatty acids to be broken down in the peroxisome but also allowed for the acetyl-CoA produced from β-oxidation to serve as a carbon and energy source for cell growth. Flux can now be carried through peroxisomal reactions utilizing acetyl-CoA i.e. carnitine acetyl-CoA transferase, citrate synthase, and malate synthase. Carnitine acetyl-CoA transferase mediates transfer of the acetyl unit from the peroxisome to the mitochondria, where it can then feed into the TCA cycle for further oxidation and energy release (Hiltunen, Mursula et al. 2003). The two glyoxylate cycle enzymes citrate

synthase and malate synthase use acetyl-CoA to generate citrate and malate, respectively. These metabolites can be exported to the cytosol where they can undergo further transformation and can ultimately be used for synthesis of carbohydrates through gluconeogenesis (Hiltunen, Mursula et al. 2003). Given the increase in utilizable pathways in the model, the model now has the capacity to simulate growth on fatty acids as the sole carbon source.

One area of metabolism that was originally capable of carrying flux in simulations but was blocked following changes to the model was mitochondrial fatty acid synthesis (Figure 3.1). In the original model, certain acyl-ACP species created in the mitochondria could be transported to the cytoplasm where they can then be hydrolyzed to free fatty acids. Removal of the acyl-ACP transport and hydrolysis reactions, as was done in the model changes, leads to blocked mitochondrial fatty acid synthesis. The removal of these reactions was justified by the observation that the mitochondrial fatty acid synthase system is unable to compensate for the loss of cytoplasmic fatty acid synthesis in *fas1* or *fas2* mutants (Schweizer and Hofmann 2004). The blocked nature of mitochondrial fatty acid synthesis in the modified model does not stem from gaps within the pathway but rather lack of utility of its end products. Although mutants in mitochondrial fatty acid synthesis are viable *in vivo*, these strains display deficiency in mitochondrial respiration (Hiltunen, Schonauer et al. 2009). This respiratory deficiency can be straightforwardly explained by lack of lipoic acid, which is an essential cofactor for α-keto acid dehydrogenase complexes and is produced from octanoyl-ACP from mitochondrial fatty acid synthesis (Hiltunen, Schonauer et al. 2009). The dysfunction in respiration is currently not captured by either the original model or the modified model, which can be attributed to several issues. The synthesis and attachment of the lipoic acid moiety to the glycine decarboxylase, pyruvate dehydrogenase, and α-ketoglutarate dehydrogenase enzyme complexes are not

represented in the original Yeast v6.0 model or in the modified model. Furthermore, in the reaction mechanisms for the enzyme complexes, the lipoic acid moiety undergoes transformations but is ultimately regenerated. Thus, the net reactions do not include any terms for lipoic acid which obviates the need for this metabolite and consequently yields a discrepancy between simulation and *in vivo* requirements.

Additional pathways are also blocked in both the original and revised Yeast v6.0 model due to the lack of utility of the pathways' end products in the model. Two examples are cardiolipin metabolism and phosphoinositide metabolism. Cardiolipin plays a role in multiple mitochondrial processes such as transport of proteins into the mitochondria, mitochondrial energy production, regulation of apoptosis, and membrane fusion (Osman, Voelker et al. 2011). Phosphoinositides are involved in many functions such as signaling, recruitment of proteins to specific membranes, regulation of cell wall maintenance/synthesis, and vesicle-mediated membrane trafficking (Strahl and Thorner 2007). Although reactions for the synthesis of cardiolipin and phosphoinositides are contained in the model, their involvement in the aforementioned processes is beyond the scope of both the original and revised model. The comprehensive modeling of metabolism, regulation, signaling, and other cellular processes involves complex and, in many cases, unknown interactions; this presents a great challenge for modeling (Oberhardt, Palsson et al. 2009). As it stands, modeling only purely metabolic reactions for these metabolites yields blocked reactions in the pathways. Since the revised model expands upon the individual species of the lipid class, there is more of a penalty in the number of blocked reactions for the revised model compared to the original model which uses generic classes.

3.4.4 Predictions on Gene Essentiality

Analysis of gene essentiality has been one standard measure for the predictive ability of genome-scale metabolic models (Thiele and Palsson 2010). In this test, the effect of single gene deletions are assessed on the basis of whether removal of reactions associated with each gene in the model blocks the ability for biomass production. The results of the single gene deletions for both the original and modified Yeast v6.0 model using a simulated aerobic glucose-limited defined media are summarized in Table 3.3. Although there are individual differences in predictions of essential and lethal gene deletions, the aggregate statistics remain consistent between the original and modified model. This is to be expected since only a small portion of metabolism was focused upon for manual curation. Both models are able to match the *in vivo* phenotype of the single-gene deletion mutants for approximately 87% of the genes contained in the model. Both models have better prediction accuracies for viable gene deletions than lethal gene deletions, with approximately 98% of viable gene deletions correctly predicted versus 52% of lethal gene deletions correctly predicted.

Table 3.3: Accuracy of gene essentiality predictions for the Yeast v6.0 model before and after implementing the changes suggested through manual curation.

Yeast v6.0 Model	Before Changes	After Changes
True Positive (TP)	674	691
True Negative (TN)	114	114
False Positive (FP)	102	103
False Negative (FN)	11	8
Sensitivity	98.4%	98.9%
Specificity	52.8%	52.5%
Overall Accuracy	87.5%	87.9%

Positive = viable gene deletion. Negative = lethal gene deletion.
Sensitivity = TP/(TP + FN), Specificity = TN/(TN + FP)
Overall accuracy = (TP + TN)/(TP + TN + FP + FN)

While there is minimal change in relative accuracy between the original and modified model, focusing on individual differences in predictions of gene essentiality allows for assessment of how the changes made to the model affect simulation results and also provides further insight into the limitations of the models. The addition of 15 new genes to the model led to 14 new true positive predictions and 1 new false positive prediction. This false positive prediction was for the gene *FRQ1* whose gene product recruits *PIK1*-encoded phosphatidylinositol 4-kinase to the Golgi membrane and stimulates Pik1p activity (Strahl and Thorner 2007). The false positive prediction, i.e. deletion of *FRQ1* was predicted to be viable despite *in vivo* lethality, can be attributed to the lack of utility of phosphoinositides in the model (see Section 3.4.3). Thus, the synthesis of phosphoinositides is treated as inessential in the model. This leads to additional false positives in both the original and modified model for genes that catalyze the synthesis of phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate (i.e. *STT4*, *PIK1*, and *MSS4*). A simple workaround is to include these species in the pseudo-reaction representing biomass production, therefore requiring synthesis of these species to form biomass. For this approach, it should be noted that the separate pools of phosphatidylinositol 4-phosphate (PI 4-P) in the cell have their own distinct roles in cell functioning; absence of *STT4*, which generates PI 4-P at the plasma membrane, cannot be compensated by overproduction of *PIK1*, which generates PI 4-P at the Golgi membrane, and vice versa (Strahl and Thorner 2007).

Several of the genes involved in metabolism of very-long chain fatty acids had different predictions of gene essentiality between the original and modified model. Single-gene deletions of *FAT1*, *FEN1*, and *SUR4* were correctly predicted as being viable in the modified model, while incorrectly predicted as lethal in the original model. *FAT1*-encoded acyl-CoA synthetase

catalyzes the esterification of very long chain fatty acids with coenzyme A to form very long chain acyl-CoA's (Choi and Martin 1999). Another means of generating very long chain acyl-CoA's is through fatty acid elongation (see Section 3.4.1.2). *FAT1* is incorrectly predicted as essential in the original model since it depicts fatty acid elongation as producing free fatty acids instead of acyl-CoA's. Therefore, the very long chain acyl-CoA's needed to synthesize sphingolipids is only able to be produced through *FAT1*-encoded acyl-CoA synthetase in the original model.

The depiction of fatty acid elongation in the original model also leads to its false negative predictions for *FEN1* and *SUR4.* Both *FEN1* and *SUR4* encode for elongases which have partially overlapping ranges of acyl-CoA substrates and function; mutants with deletion of either *FEN1* or *SUR4* are viable while deletion of both genes is lethal (Oh, Toke et al. 1997). The original model contained two lumped reactions for fatty acid elongation which had gene associations of "*TSC13* and *FEN1* and *SUR4*". This logical relationship employs the AND condition which requires all the genes in the association to be present for fatty acid elongation to occur. In contrast, the modified model has individual reactions for the action of elongases on different acyl-CoA substrates and assigns the appropriate gene relationship based on substrate specificity of each elongase. As a result, the modified model accurately predicts the viability of single gene deletions of *FEN1* or *SUR4* and the lethality of simultaneous deletion of both genes.

For the fatty acid elongation system, one gene deletion that the original model is able to predict correctly as viable while the modified model predicts falsely as lethal is *IFA38.* This correct prediction in the original model however is not due to accuracy in its representation of the functioning of *IFA38*-encoded β-ketoacyl-CoA reductase. All of the reactions in the original model that are associated with *IFA38* are blocked reactions due to the lack of connection to the

rest of fatty acid elongation (see Section 3.4.1.2). Thus, all the reactions associated with *IFA38* are inconsequential to simulation results using the original model and therefore deletion of *IFA38* is predicted as viable. In comparison, the reactions associated with *IFA38* in the modified model can carry metabolic flux and are also needed for fatty acid elongation to occur. The synthesis of very long chain fatty acids is essential *in vivo* and the viability of *ifa38*Δ suggests capacity for residual β-ketoacyl-CoA reductase activity (Han, Gable et al. 2002). It has been hypothesized that Ayr1p is responsible for this residual activity based on its homology to *IFA38* and the synthetic lethality of *ifa38*Δ*ayr1*Δ (Han, Gable et al. 2002). However, overexpression of *AYR1* does not suppress the slow growth of *ifa38*Δ mutants (Han, Gable et al. 2002) and another study's *in vitro* data does not support any role of Ayr1p in fatty acid elongation in *ifa38*Δ mutants (Rössler, Rieck et al. 2003). Based on lack of direct evidence, the gene *AYR1* was not associated with fatty acid elongation in the modified model. This modeling decision meant that the only gene product capable of catalyzing the β-ketoacyl-CoA reductase activity for fatty acid elongation in the modified model was *IFA38's* and consequently the deletion of *IFA38* is falsely predicted as lethal for the modified model.

In comparing the original and modified model, there are also instances where the original model is able to correctly predict the single gene deletion as lethal which the modified model incorrectly predicts as viable. Nevertheless, the reason for why the gene is essential for simulation of biomass production in the original model does not match with the true role of the gene *in vivo*. This is the case for the genes *RIM2* and *PET8*. *RIM2* encodes for a transporter that imports (deoxy)pyrimidine nucleoside triphosphates into the mitochondria in exchange for intramitochondrially generated (deoxy)pyrimidine nucleoside monophosphates (Marobbio, Di noia et al. 2006). The imported (deoxy)pyrimidine nucleoside triphosphates are essential for

synthesis of mitochondrial DNA and RNA. Neither the original nor modified model captures the role of *RIM2* in providing precursors for mitochondrial DNA and RNA metabolism. Instead, the original model requires *RIM2* for exchange of the mitochondrial nucleotides CMP and CTP for synthesis of the glycerophospholipids phosphatidylinositol, phosphatidylserine, and CDP-diacylglycerol. The modified model does not require *RIM2* for this purpose because the reactions for *PIS1*-encoded phosphatidylinositol synthase and *CHO1*-encoded phosphatidylserine synthase are localized to the ER membrane while *CDS1*-encoded CDP-diacylglycerol synthase is localized to both the ER membrane and mitochondrial membrane. It should be noted that the original model does have these enzymes localized to other organelles besides the mitochondria; however, the reactions in these alternate compartments are blocked and flux can only be carried through the mitochondrial reactions in the original model.

The correct prediction of *PET8* as an essential gene using the original model and wrong prediction using the modified model can be attributed to similar circumstances described above for *RIM2*. The physiological role of Pet8p is to transport S-adenosylmethionine into the mitochondria, where it is utilized as a cofactor in biotin and lipoic acid synthesis and also as a methyl group donor for methylation of DNA, RNA, and protein (Marobbio, Agrimi et al. 2003). The involvement of S-adenosylmethionine in these processes is not accounted for in either the original or modified model. The reason for the essentiality of *PET8* in the original model is for the transport of S-adenosylmethionine into the mitochondria for the methylation of phosphatidylethanolamine to phosphatidylcholine. The modified model, in contrast, has these methylation reactions localized to the ER membrane and therefore does not require mitochondrial S-adenosylmethionine for this purpose.

A direct consequence of the expansion of glycerolipid and glycerophospholipid species in the modified model was the correct prediction that *ole1*Δ mutants require unsaturated fatty acids for growth (Stukey, McDonough et al. 1989), which the original model could not foresee. *OLE1* encodes for Δ9-desaturase which is required for production of unsaturated fatty acids (Stukey, McDonough et al. 1990). The inability of the original model to predict the unsaturated fatty acid auxotrophy of *ole1*Δ mutants stems from its usage of "isa" reactions which treats species as functionally equivalent. For instance, the "isa acyl-CoA" reaction categorizes various individual acyl-CoA species (e.g. palmitoyl-CoA, palmitoleoyl-CoA, etc.) into the generic term of acyl-CoA. Thus, any of the categorized acyl-CoA species can fulfill the requirement for the acyl-CoA used in lipid metabolism. In contrast, the modified model employs the assumption that the acyl chain found at the sn-2 position of the glycerol backbone of glycerolipids and glycerophospholipids is either C16:1 or C18:1. This assumption imposes a specific requirement for unsaturated acids that must be met either through synthesis with Ole1p or through supplementation of the media.

3.5 Conclusions

Through the process of manual curation, we have identified inconsistencies between information contained in the Yeast consensus model (Yeast v6.0) and literature knowledge for fatty acid, glycerolipid, and glycerophospholipid metabolism. These inconsistencies include instances of incorrect gene-reaction associations, improper definition of substrates/products in reactions, and inappropriate assignments of reaction directionality. In addition to correcting these inconsistencies, the addition of 15 new genes and the introduction of increased specificity in representation of glycerolipid and glycerophospholipid metabolism through the expansion of lipid classes and denotation of membrane reactions are proposed. A complete consideration of

triglyceride metabolism should include pathways for its synthesis and for its degradation. Through addition of new genes and reactions, the proposed modifications account for the synthesis of the key intermediate phosphatidate from dihydroxyacetone phosphate, for the conversion of diglyceride generated from hydrolysis of triglyceride into phosphatidate, and for the complete breakdown of triglyceride into glycerol and free fatty acids.

One consequence of the proposed changes to the Yeast v6.0 model is the newly acquired ability to simulate growth on fatty acids as sole carbon source. This functionality is enabled through the amendment of blocked reactions in β-oxidation. Analysis of the blocked reactions in the Yeast v6.0 model both before and after implementing the suggested changes revealed gaps reflecting the challenge in interlinking metabolic reactions to complex biological processes such as signaling and the unaccounted-for roles of metabolites such as cofactors. The presence of these blocked reactions ultimately leads to an incomplete picture that does not capture the function of genes associated with the blocked reactions. This is especially consequential for the predictive accuracy of the effect of gene deletions. In this respect, deletions of genes associated with blocked reactions are predicted as viable. The high percentage of blocked reactions is not unique to the Yeast consensus model; the iMM904 and iND750 models of yeast metabolism have respectively 31% and 41% of all reactions blocked (Heavner, Smallbone et al. 2012). Thus, with the various models, the accuracy of predictions for viability of gene deletions can be highly influenced by blocked reactions instead of a basis rooted in biological reality.

The suggested changes for Yeast v6.0 have been submitted to Dr. Kieran Smallbone, who maintains the Yeast metabolic network reconstruction, and has since been incorporated to generate Yeast v7.0. Prior to the publication of this paper, another genome-scale model of *S. cerevisiae* metabolism called iTO977 has been published (Osterlund, Nookaew et al. 2013).

iTO977 merges the earlier models of iIN800 (Nookaew, Jewett et al. 2008) and Yeast v1.0 (Herrgard, Swainston et al. 2008) and then improves and expands upon this through gap-filling methods and the introduction of additional genes, reactions, and metabolites based on literature evidence and database searches. The process presented in this paper for analysis of the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism was repeated for iTO977 and a discussion of the differences between Yeast v6.0 and iTO977 is provided in Appendix 1. Overall, there were instances where iTO977 had the same issues as the Yeast v6.0 model, where iTO977 was in agreement with the proposed modifications, and where iTO977 had its own unique issues.

The modeling of the entirety of an organism's metabolic network is highly complex. The databases upon which these networks are built from can contain errors, even in well-known pathways such as the TCA cycle (Stobbe, Houten et al. 2012). These inaccuracies can become embedded in the model unless efforts are taken to inspect the representation of metabolism. One approach to reach this goal is to manually curate sub-portions of the model defined by pathways of interest. Although this requires significant time and effort, the benefit of such manual curation is an increased confidence and also awareness of limitations in the examined portion of the model.

An accurate representation of the metabolic network is a key component to ensure that application of the model yields results reflective of actual biochemistry. Genome-scale metabolic models have been used to aid the interpretation of high-throughput data, to guide metabolic engineering, and to generate testable hypotheses on cellular behavior (Oberhardt, Palsson et al. 2009). Our own particular interest is in using this framework to explore the molecular organization of lipid biosynthesis and how this synthesis influences the metabolic fluxes and

pools of critical metabolites, such as PA, PI, DAG and TAG, which play important roles in signaling and responding to environmental stress. Yeast adapt to environmental stresses such as product inhibition, nutrient limitation, elevated temperatures, and high osmolarity through manipulation of lipid yield and membrane composition. A better understanding of key lipid biosynthetic pathways is important for industrial biotechnologists seeking to not only increase lipid yields but also to push the environmental limits of microbial conversion systems.

REFERENCES

Athenstaedt, K. and G. Daum (2000). "1-Acyldihydroxyacetone-phosphate Reductase (Ayr1p) of the Yeast Saccharomyces cerevisiae Encoded by the Open Reading Frame YIL124w Is a Major Component of Lipid Particles." Journal of Biological Chemistry **275**(1): 235-240.

Bornstein, B. J., S. M. Keating, A. Jouraku and M. Hucka (2008). "LibSBML: an API Library for SBML." Bioinformatics **24**(6): 880-881.

Carman, G. M. and G.-S. Han (2006). "Roles of phosphatidate phosphatase enzymes in lipid metabolism." Trends in Biochemical Sciences **31**(12): 694-699.

Choi, J.-Y. and C. E. Martin (1999). "The Saccharomyces cerevisiae FAT1 Gene Encodes an Acyl-CoA Synthetase That Is Required for Maintenance of Very Long Chain Fatty Acid Levels." Journal of Biological Chemistry **274**(8): 4671-4683.

Dahlqvist, A., U. Ståhl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne and S. Stymne (2000). "Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants." Proceedings of the National Academy of Sciences **97**(12): 6487-6492.

de Kroon, A. I. P. M. (2007). "Metabolism of phosphatidylcholine and its implications for lipid acyl chain composition in Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 343-352.

Debelyy, M. O., S. Thoms, M. Connerth, G. Daum and R. Erdmann (2011). "Involvement of the Saccharomyces cerevisiae Hydrolase Ldh1p in Lipid Homeostasis." Eukaryotic Cell **10**(6): 776-781.

Denic, V. and J. S. Weissman (2007). "A Molecular Caliper Mechanism for Determining Very Long-Chain Fatty Acid Length." Cell **130**(4): 663-677.

Duarte, N. C., M. J. Herrgård and B. Ø. Palsson (2004). "Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model." Genome Research **14**(7): 1298-1309.

Fakas, S., C. Konstantinou and G. M. Carman (2011). "DGK1-encoded Diacylglycerol Kinase Activity Is Required for Phospholipid Synthesis during Growth Resumption from Stationary Phase in Saccharomyces cerevisiae." Journal of Biological Chemistry **286**(2): 1464-1474.

Förster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen (2003). "Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network." Genome Research **13**(2): 244-253.

Grillitsch, K. and G. Daum (2011). "Triacylglycerol lipases of the yeast." Frontiers in Biology **6**(3): 219-230.

Ham, H. J., H. J. Rho, S. K. Shin and H.-J. Yoon (2010). "The TGL2 Gene of Saccharomyces cerevisiae Encodes an Active Acylglycerol Lipase Located in the Mitochondria." Journal of Biological Chemistry **285**(5): 3005-3013.

Han, G.-S., W.-I. Wu and G. M. Carman (2006). "The Saccharomyces cerevisiae Lipin Homolog Is a Mg2+-dependent Phosphatidate Phosphatase Enzyme." Journal of Biological Chemistry **281**(14): 9210-9218.

Han, G., K. Gable, S. D. Kohlwein, F. Beaudoin, J. A. Napier and T. M. Dunn (2002). "The Saccharomyces cerevisiae YBR159w Gene Encodes the 3-Ketoreductase of the Microsomal Fatty Acid Elongase." Journal of Biological Chemistry **277**(38): 35440-35449.

Heavner, B. D., S. A. Henry and L. P. Walker (2012). "Evaluating Sphingolipid Biochemistry in the Consensus Reconstruction of Yeast Metabolism." Industrial Biotechnology **8**(2): 72-78.

Heavner, B. D., K. Smallbone, B. Barker, P. Mendes and L. P. Walker (2012). "Yeast 5–an expanded reconstruction of the Saccharomyces cerevisiae metabolic network." BMC Systems Biology **6**(1): 55.

Heier, C., U. Taschler, S. Rengachari, M. Oberer, H. Wolinski, K. Natter, S. D. Kohlwein, R. Leber and R. Zimmermann (2010). "Identification of Yju3p as functional orthologue of mammalian monoglyceride lipase in the yeast Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1801**(9): 1063-1071.

Henry, S. A., S. D. Kohlwein and G. M. Carman (2012). "Metabolism and Regulation of Glycerolipids in the Yeast Saccharomyces cerevisiae." Genetics **190**(2): 317-349.

Herrgard, M. J., N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Büthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasie, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kurdar, M. Penttila, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen and D. B. Kell (2008). "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology." Nat Biotech **26**(10): 1155-1160.

Hiltunen, J. K., A. M. Mursula, H. Rottensteiner, R. K. Wierenga, A. J. Kastaniotis and A. Gurvitz (2003). "The biochemistry of peroxisomal β-oxidation in the yeast Saccharomyces cerevisiae." FEMS Microbiology Reviews **27**(1): 35-64.

Hiltunen, J. K., M. S. Schonauer, K. J. Autio, T. M. Mittelmeier, A. J. Kastaniotis and C. L. Dieckmann (2009). "Mitochondrial Fatty Acid Synthesis Type II: More than Just Fatty Acids." Journal of Biological Chemistry **284**(14): 9011-9015.

Keating, S. M., B. J. Bornstein, A. Finney and M. Hucka (2006). "SBMLToolbox: an SBML toolbox for MATLAB users." Bioinformatics **22**(10): 1275-1277.

Kohlwein, S. D., S. Eder, C.-S. Oh, C. E. Martin, K. Gable, D. Bacikova and T. Dunn (2001). "Tsc13p Is Required for Fatty Acid Elongation and Localizes to a Novel Structure at the Nuclear-Vacuolar Interface in Saccharomyces cerevisiae." Mol. Cell. Biol. **21**(1): 109-125.

Kosa, M. and A. J. Ragauskas (2011). "Lipids from heterotrophic microbes: advances in metabolism research." Trends in Biotechnology **29**(2): 53-61.

Kuepfer, L., U. Sauer and L. M. Blank (2005). "Metabolic functions of duplicate genes in Saccharomyces cerevisiae." Genome Research **15**(10): 1421-1430.

Kurat, C. F., K. Natter, J. Petschnigg, H. Wolinski, K. Scheuringer, H. Scholz, R. Zimmermann, R. Leber, R. Zechner and S. D. Kohlwein (2006). "Obese Yeast: Triglyceride Lipolysis Is Functionally Conserved from Mammals to Yeast." Journal of Biological Chemistry **281**(1): 491-500.

Lomakin, I. B., Y. Xiong and T. A. Steitz (2007). "The Crystal Structure of Yeast Fatty Acid Synthase, a Cellular Machine with Eight Active Sites Working Together." Cell **129**(2): 319-332.

Marobbio, C. M. T., G. Agrimi, F. M. Lasorsa and F. Palmieri (2003). "Identification and functional reconstitution of yeast mitochondrial carrier for S-adenosylmethionine." EMBO J **22**(22): 5975-5982.

Marobbio, C. M. T., M. A. Di noia and F. Palmieri (2006). "Identification of a mitochondrial transporter for pyrimidine nucleotides in Saccharomyces cerevisiae: bacterial expression, reconstitution and functional characterization." Biochem J **393**(2): 441-446.

Martin, C. E., C.-S. Oh and Y. Jiang (2007). "Regulation of long chain unsaturated fatty acid synthesis in yeast." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 271-285.

Mo, M., B. Palsson and M. Herrgard (2009). "Connecting extracellular metabolomic measurements to intracellular flux states in yeast." BMC Systems Biology **3**(1): 37.

Nookaew, I., M. Jewett, A. Meechai, C. Thammarongtham, K. Laoteng, S. Cheevadhanarak, J. Nielsen and S. Bhumiratana (2008). "The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism." BMC Systems Biology **2**(1): 71.

Oberhardt, M. A., B. Ø. Palsson and J. A. Papin (2009). "Applications of genome-scale metabolic reconstructions." Mol Syst Biol **5**.

Oelkers, P., D. Cromley, M. Padamsee, J. T. Billheimer and S. L. Sturley (2002). "The DGA1 Gene Determines a Second Triglyceride Synthetic Pathway in Yeast." Journal of Biological Chemistry **277**(11): 8877-8881.

Oelkers, P., A. Tinkelenberg, N. Erdeniz, D. Cromley, J. T. Billheimer and S. L. Sturley (2000). "A Lecithin Cholesterol Acyltransferase-like Gene Mediates Diacylglycerol Esterification in Yeast." Journal of Biological Chemistry **275**(21): 15609-15612.

Oh, C.-S., D. A. Toke, S. Mandala and C. E. Martin (1997). "ELO2 and ELO3, Homologues of the Saccharomyces cerevisiae ELO1 Gene, Function in Fatty Acid Elongation and Are Required for Sphingolipid Formation." Journal of Biological Chemistry **272**(28): 17376-17384.

Osman, C., D. R. Voelker and T. Langer (2011). "Making heads or tails of phospholipids in mitochondria." Journal of Cell Biology **192**(1): 7-16.

Osterlund, T., I. Nookaew, S. Bordel and J. Nielsen (2013). "Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling." BMC Systems Biology **7**(1): 36.

Paul, S., K. Gable and T. M. Dunn (2007). "A Six-membrane-spanning Topology for Yeast and Arabidopsis Tsc13p, the Enoyl Reductases of the Microsomal Fatty Acid Elongating System." Journal of Biological Chemistry **282**(26): 19237-19246.

Rodríguez-Vargas, S., A. Sánchez-García, J. M. Martínez-Rivas, J. A. Prieto and F. Randez-Gil (2007). "Fluidization of Membrane Lipids Enhances the Tolerance of Saccharomyces cerevisiae to Freezing and Salt Stress." Applied and Environmental Microbiology **73**(1): 110-116.

Rössler, H., C. Rieck, T. Delong, U. Hoja and E. Schweizer (2003). "Functional differentiation and selective inactivation of multiple *Saccharomyces cerevisiae* genes involved in very-long-chain fatty acid synthesis." Molecular Genetics and Genomics **269**(2): 290-298.

Sandager, L., M. H. Gustavsson, U. Ståhl, A. Dahlqvist, E. Wiberg, A. Banas, M. Lenman, H. Ronne and S. Stymne (2002). "Storage Lipid Synthesis Is Non-essential in Yeast." Journal of Biological Chemistry **277**(8): 6478-6482.

Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke and B. O. Palsson (2011). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0." Nature Protocols **6**(9): 1290-1307.

Schweizer, E. and J. Hofmann (2004). "Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems." Microbiol. Mol. Biol. Rev. **68**(3): 501-517.

Sorger, D. and G. Daum (2002). "Synthesis of Triacylglycerols by the Acyl-Coenzyme A:Diacyl-Glycerol Acyltransferase Dga1p in Lipid Particles of the Yeast Saccharomyces cerevisiae." Journal of Bacteriology **184**(2): 519-524.

Stobbe, M. D., S. M. Houten, A. H. C. van Kampen, R. J. A. Wanders and P. D. Moerland (2012). "Improving the description of metabolic networks: the TCA cycle as example." The FASEB Journal **26**(9): 3625-3636.

Strahl, T. and J. Thorner (2007). "Synthesis and function of membrane phosphoinositides in budding yeast, Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 353-404.

Stukey, J. E., V. M. McDonough and C. E. Martin (1989). "Isolation and characterization of OLE1, a gene affecting fatty acid desaturation from Saccharomyces cerevisiae." Journal of Biological Chemistry **264**(28): 16537-16544.

Stukey, J. E., V. M. McDonough and C. E. Martin (1990). "The OLE1 gene of Saccharomyces cerevisiae encodes the delta 9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene." Journal of Biological Chemistry **265**(33): 20144-20149.

Tehlivets, O., K. Scheuringer and S. D. Kohlwein (2007). "Fatty acid synthesis and elongation in yeast." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 255-270.

Thiele, I. and B. Ø. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature Protocols **5**(1): 93-121.

Toke, D. A. and C. E. Martin (1996). "Isolation and Characterization of a Gene Affecting Fatty Acid Elongation in Saccharomyces cerevisiae." Journal of Biological Chemistry **271**(31): 18413-18422.

U.S. Environmental Protection Agency (2011). "Biofuels and the Environment: First Triennial Report to Congress." Office of Research and Development, National Center for Environmental Assessment, Washington, DC. EPA/600/R-10/183F.

van Roermund, C. W., Y. Elgersma, N. Singh, R. J. Wanders and H. F. Tabak (1995). "The membrane of peroxisomes in Saccharomyces cerevisiae is impermeable to NAD(H) and acetyl-CoA under in vivo conditions." The EMBO journal **14**(14): 3480-3486.

Wagner, S. and F. Paltauf (1994). "Generation of glycerophospholipid molecular species in the yeast Saccharomyces cerevisiae. Fatty acid pattern of phospholipid classes and selective acyl turnover at sn-1 and sn-2 positions." Yeast **10**(11): 1429-1437.

You, K. M., C.-L. Rosenfield and D. C. Knipple (2003). "Ethanol Tolerance in the Yeast Saccharomyces cerevisiae Is Dependent on Cellular Oleic Acid Content." Applied and Environmental Microbiology **69**(3): 1499-1503.

Chapter 4: The Influence of Cofactors, Thermodynamics, and Compartmentalization on Calculated Pathway Yields throughout the Evolution of the Yeast Consensus Model

4.1 Abstract

4.1.1 Background

The choice of assumptions applied in genome-scale metabolic models can have a significant impact on simulated results. We examined the impact of changes made throughout the evolution of the genome-scale metabolic model for *Saccharomyces cerevisiae* on calculated maximum pathway yields of various industrially relevant biochemicals and of various lipids.

4.1.2 Results

The factors found to have affected the maximum pathway yield calculated by each of the metabolic models were: assumptions relevant to ATP and NADH/NADPH, assumed reversibility of reactions, and assumed transport reactions. The maximum pathway yields of 2-methylbutanol, 3-methylbutanol, succinate, lysine, citrate, and glutamate were found to be sensitive to the assumed proton stoichiometry for the electron transport chain and ATP synthase and to NADH/NADPH cofactor imbalance. The dominant factor that led to large differences in the calculated maximum pathway yields for lipids was whether or not acetyl-CoA hydrolase was assumed to be capable of catalyzing the reverse direction.

4.1.3 Conclusions

Genome-scale metabolic models condense our knowledge of the inventory of biochemical reactions occurring in cells and can be applied to identify limits on the maximum

yield of different metabolites based on the stoichiometry and reversibility of reactions in metabolic pathways. The maximum conversion yields that are calculated are sensitive to the modeling decisions employed as evidenced by differences in calculated pathway yields from four different versions of *S. cerevisiae* genome-scale metabolic models.

4.2 Background

The microbial conversion of renewable feedstocks to biofuels and high value chemicals represents a sustainable alternative to petroleum-derived chemicals. However, in order to be a viable alternative, the microbial conversion has to be competitive in product yield and in productivity. The use of genome-scale metabolic models to guide metabolic engineering can help achieve this goal. This may be accomplished through the use of these models to evaluate the maximum theoretical efficiencies of pathways, to select host strains based on predicted metabolic capabilities, to understand actual metabolic operation, and to predict the effect of genetic modifications (Durot, Bourguignon et al. 2009).

The steps involved in creating a genome-scale metabolic model are often iterative and ongoing as new research discovers previously uncharacterized reactions or elucidates mechanisms in metabolic pathways that expands or changes our current understanding of metabolism. For instance, the consensus model of *Saccharomyces cerevisiae* metabolism (Herrgard, Swainston et al. 2008) is currently on its seventh major version release. Changes made to this model throughout the course of development include the addition of new reactions; removal of dubious reactions; reassignment of reversibility/irreversibility to reactions; redefined enzyme specificity; and reclassified compartmentalization of reactions (Heavner, Henry et al. 2012; Heavner, Smallbone et al. 2012; Aung, Henry et al. 2013; Heavner, Smallbone et al. 2013). These changes have been found to have impacts on predictions of gene essentiality and

auxotrophy (Heavner, Smallbone et al. 2012; Aung, Henry et al. 2013; Heavner, Smallbone et al. 2013), simulated flux under aerobic and anaerobic conditions (Heavner, Smallbone et al. 2012), and ability to consume fatty acids as sole carbon source (Aung, Henry et al. 2013).

Here, we further evaluate the impact of the changes throughout the evolution of the consensus model of *Saccharomyces cerevisiae* metabolism by examining the calculated maximum pathway yields of various metabolites. This evolution is assessed in this paper through consideration of the models iMM904 (Mo, Palsson et al. 2009), a progenitor to the consensus models, and the three latest major version releases of the consensus model Yeast v5.01, v6.06, and v7.11 (http://yeast.sourceforge.net/). The maximum pathway yields calculated by the yeast models for eight different chemicals of interest in industrial biotechnology are compared to yields calculated by Dugar and Stephanopoulos without the use of genome-scale metabolic models (Dugar and Stephanopoulos 2011). Given our role in curating the lipid pathways in the consensus model (Aung, Henry et al. 2013), we also examined how changes to the yeast model affected the calculated maximum pathway yields of the major glycerophospholipids and triacylglycerol.

4.3 Results

4.3.1 Maximum Pathway Yield of Various Industrially Relevant Biochemicals

Table 4.1 shows the pathway yield of eight different chemicals of interest in industrial biotechnology as calculated by Dugar and Stephanopoulos (Dugar and Stephanopoulos 2011), the three latest major versions of the consensus model of *S. cerevisiae* metabolism (http://yeast.sourceforge.net/), and the iMM904 *S. cerevisiae* metabolic model (Mo, Palsson et al. 2009). Briefly, Dugar and Stephanopoulos' calculated yields are based on the stoichiometric pathway balance for product formation and the additional glucose consumption for production of

NADPH and ATP if needed by the pathway. Yield calculations using the genome-scale metabolic models utilize the same principle of balancing except applied to all metabolites contained within the model. This is accomplished using the method of flux balance analysis and optimizing for the exchange of each metabolite of interest.

Table 4.1: Maximum pathway yields calculated by Dugar & Stephanopoulos and the yeast models

	Maximum Pathway Yield (g Product/g Glucose)				
Metabolite	**Dugar & Stephanopoulos**	**Yeast v7.11**	**Yeast v6.06**	**Yeast v5.01**	**iMM904**
Ethanol	0.511	0.511	0.511	0.511	0.511
Isobutanol	0.411	0.411	0.370	0.408	0.411
2-Methylbutanol	0.380	0.364	0.383	0.381	0.370
3-Methylbutanol	0.326	0.351	0.345	0.367	0.360
Succinate	1.124	0.732	0.787	1.051	1.124
Lysine	0.695	0.546	0.512	0.558	0.564
Citrate	1.067	1.104	1.066	1.171	1.202
Glutamate	0.823	0.851	0.897	0.903	0.927

4.3.1.1 Calculating maximum yield of cytoplasmic ATP and NADPH from glucose

As an initial approach to understand the differences in pathway yields, we examined the theoretical maximum yield of ATP and NADPH from glucose for each of the genome-scale models of *S. cerevisiae* metabolism (Table 4.2). This was accomplished by adding a pseudo-reaction for ATP hydrolysis and for NADPH oxidation and conducting flux balance analysis maximizing for these pseudo-reactions.

Table 4.2: Maximum yield of cytoplasmic ATP and NADPH from glucose calculated by the yeast models

	Max Yield (mol/mol Glucose)			
	Yeast v7.11	**Yeast v6.06**	**Yeast v5.01**	**iMM904**
ATP	11.5	264.3	14.0	17.0
NADPH	11.1	12.0	11.3	11.4

For maximum ATP yield, the Yeast v6.06 model has an unrealistically high value of 264.3 mol ATP/mol glucose. For the Yeast v6.06 model, this stems from the operation of two different NADH shuttles running counter to each other in a way that generates a large proton gradient between the cytoplasm and mitochondria (Figure 4.1). Thus, this large proton gradient is simulated as generating large amounts of ATP through ATP synthase. This simulated proton gradient is constrained by the maximum flux bounds used in the model, which in this case was 1000 mmol/gDW/h. In contrast, these two NADH shuttles are prohibited from operating counter to each other in iMM904, Yeast v5.01, and Yeast v7.11 because of constraints in reaction directionality and/or missing transport reactions. For instance, by constraining the reaction for symport of aspartate and a proton (r_1117) to be unidirectional only from the cytoplasm to the mitochondria, the maximum ATP yield calculated by the Yeast v6.06 model becomes 14.75 mol ATP/mol glucose. Bakker et al. have estimated 16 mol ATP/mol glucose to be a reasonable yield for *S. cerevisiae*, which is admittedly much lower than the 'textbook value' of 36 mol ATP/mol glucose (Bakker, Overkamp et al. 2001).

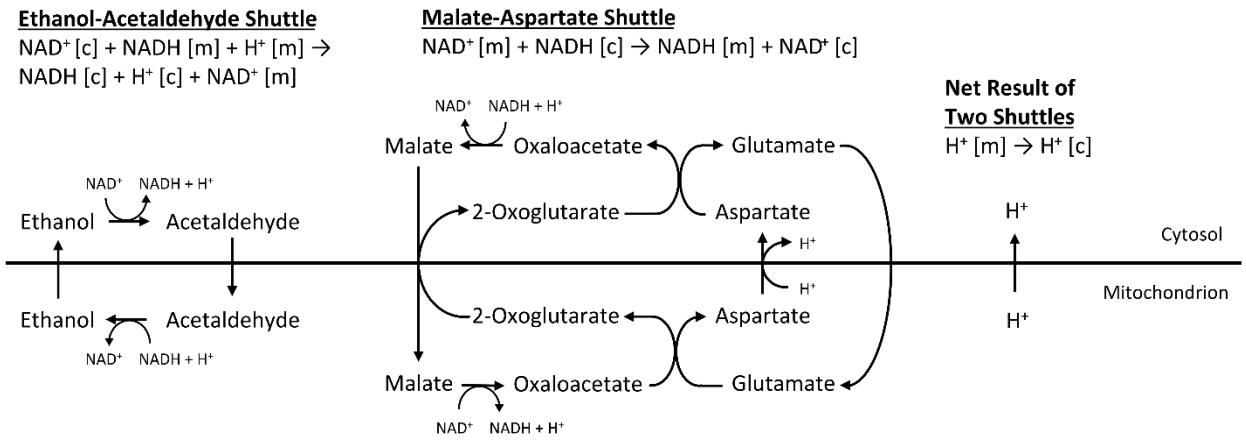


Figure 4.1: Simulation artifact from operation of two different NADH shuttles in Yeast v6.06. In Yeast v6.06, the simulated combination of the ethanol-acetaldehyde and malate-aspartate shuttles results in pumping of protons from the mitochondria to the cytoplasm, thus building up the proton gradient for ATP synthase. This mode of operation is inoperative in iMM904, Yeast v5.01, and Yeast v7.11 because of constrained reaction directionality and/or missing transport reactions.

Other sources of difference in estimated mechanistic ATP yields are the assumed stoichiometry for concomitant transport of protons between the cytoplasm and mitochondria with the electron transport chain and the required amount of protons needed to flow through ATP synthase to drive ATP synthesis. For instance, Yeast v7.11 assumes that NADH:ubiquinone oxidoreductase (r_0773) does not translocate any protons to the cytoplasm, instead of the 1.5 protons assumed in Yeast v5.01 and v6.06; that exchange of mitochondrial ATP for cytoplasmic ADP (r_1110) does not include transport of 1 proton from the cytoplasm to the mitochondria; and that ATP synthase (r_0226) requires 4, instead of 3, protons to drive ATP synthesis. The yield assumed by Dugar and Stephanopoulos is 29.85 mol ATP/mol glucose and is not specific to *S. cerevisiae*.

The maximal yield of NADPH is derived from the oxidative reactions of the pentose phosphate pathway. In this case, one glucose-6-phosphate can generate 12 NADPH molecules. However, the maximum total yield of NADPH from glucose is slightly less than 12 because of the ATP needed to phosphorylate glucose. For the Yeast v6.06 model, this ATP is able to be

freely generated by the proton gradient generated by the NADH shuttles, as illustrated in Figure 4.1. Consequently, Yeast v6.06 predicts a maximum yield of 12 mol NADPH/mol glucose, which is also the value assumed by Dugar and Stephanopoulos. iMM904, Yeast v5.01, and Yeast v7.11, in comparison, must divert some of the glucose to generate ATP which results in yields of 11.4, 11.3, and 11.1 mol NADPH/mol glucose, respectively; the difference in NADPH yield between these yeast models is only due to the stoichiometry associated for ATP generation.

4.3.1.2 Impact of assumed ATP yield on calculated pathway yields

Since there is not a universal value for the yield of ATP from glucose oxidation, we examined the sensitivity of the pathway yield of the eight different biochemicals listed in Table 4.1 to this parameter. The effective ATP yield assumed by the yeast models was adjusted by varying the stoichiometry of protons needed for ATP synthesis through ATP synthase (reaction r_0226 in the consensus models and ATPS3m in iMM904). Although the stoichiometry for proton translocation in the electron transport chain is also debatable, manipulation of the stoichiometry for just ATP synthase serves as a proxy for the net uncertainty in ATP yield. Figure 4.2 shows the variation in pathway yields for the 6 biochemicals that were sensitive to assumed ATP yield. For these 6 biochemicals, the yields varied from 3.6 to 8.7% for Yeast v7.11, were insensitive for Yeast v6.06, varied from 2.5 to 18.2% for Yeast v5.01, and varied from 0 to 15.3% for iMM904 over the range of ATP yields tested. The variation in pathway yields for the tested range of ATP yields exceeded 10% for lysine, citrate, and glutamate with iMM904 and Yeast v5.01. The pathway yields calculated by Yeast v6.06 were unresponsive to the tested range of ATP yields due to its unintended ability to freely generate a proton gradient for ATP synthesis (see previous section and Figure 4.1).

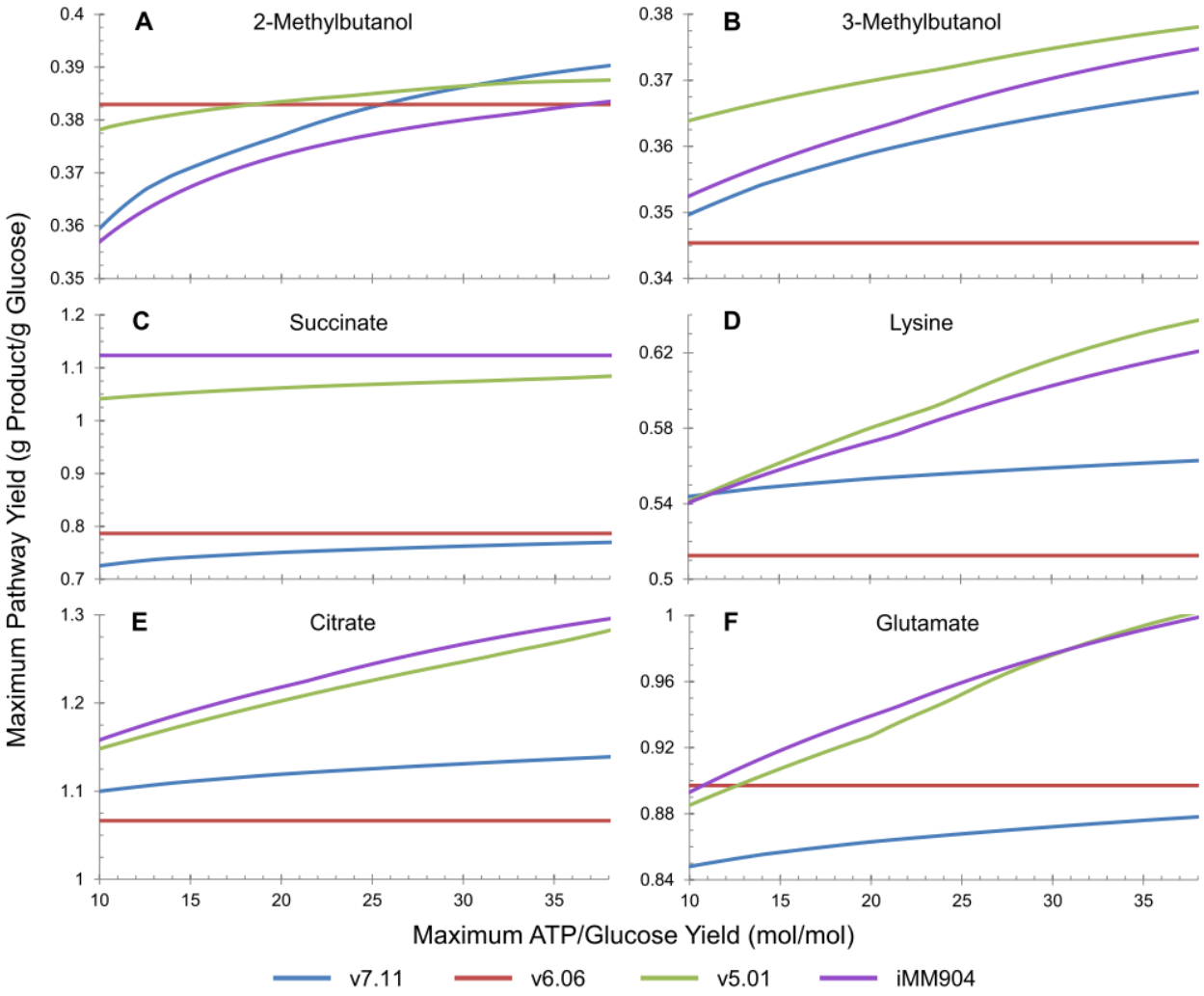


Figure 4.2: Sensitivity analysis of pathway yields with changing maximum ATP yield for the yeast models. Pathway yields for (A) 2-methylbutanol, (B) 3-methylbutanol, (C) succinate, (D) lysine, (E) citrate, and (F) glutamate were plotted as a function of maximum cytoplasmic ATP yields varying from 10 to 38 mol ATP/mol glucose. The maximum cytoplasmic ATP yield was altered in each model by adjusting the stoichiometry of protons needed for ATP synthase (reaction r_0226 in the consensus models and ATPS3m in iMM904).

4.3.1.3 Impact of NADH/NADPH cofactor balance on calculated pathway yields

NADH, in addition to NADPH, plays an important role in cellular redox state. The distinction between NADPH and NADH may not be of much concern in microorganisms with enzymes such as transhydrogenases (NADH + $NADP^+$ ⇔ NAD^+ + NADPH), but this distinction is significant in *S. cerevisiae* which lacks transhydrogenase activity (Bakker, Overkamp et al.

2001; Dugar and Stephanopoulos 2011). Further complicating the redox balance is the impermeability of the mitochondrial inner membrane to NADH and NADPH which creates distinct compartmentalized pools (Bakker, Overkamp et al. 2001). The yields for the various biochemicals reported in Table 4.1 for Dugar & Stephanopoulos assume interchangeability between the two cofactors, although these yields can also be corrected for cofactor imbalance through separate calculations. The yields calculated using the genome-scale metabolic models are sensitive to cofactor specificity and the mechanisms included in the metabolic reconstruction for conversion between NAD(H) and NADP(H). The differences in the *S. cerevisiae* genome-scale metabolic models that affect cofactor balance include: cofactor specificity of the malic enzyme (reactions r_0718/ME1m and r_0719/ME2m), reversibility of NAD^+-dependent glutamate dehydrogenase (r_0470/GLUDxi), inclusion of $NADP^+$-dependent glycerol dehydrogenase (r_0487/GLYCDy), and inclusion of $NADP^+$ phosphatase (r_1959/NADPPPS and r_1960/NADPPPSm).

In order to more aptly compare the pathway yields calculated by Dugar & Stephanopoulos and the *S. cerevisiae* genome-scale metabolic models, the pathway yields were recalculated after the addition of hypothetical reactions to the yeast models for cytoplasmic and mitochondrial transhydrogenases and for transport of the various NAD-species between the cytoplasm and mitochondria (Table 4.3). In general, the addition of these reactions increased the calculated pathway yields for all of the biochemicals except for ethanol. The metabolites with the highest increase in calculated pathway yields were succinate (54% for Yeast v7.11 and 43% for v6.06), lysine (27% for Yeast v7.11, 36% for Yeast v6.06, 25% for Yeast v5.01, and 23% for iMM904), citrate (29% for Yeast v7.11, 33% for Yeast v6.06, 21% for Yeast v5.01, and 18% for

iMM904), and glutamate (29% for Yeast v7.11, 22% for Yeast v6.06, 21% for Yeast v5.01, and 18% for iMM904).

Table 4.3: Maximum pathway yields calculated after treating NADH/NADPH as interchangeable and disregarding compartmentalization of NADH/NADPH in the yeast models. Highlighted cells show increases compared to Table 4.1.

	Maximum Pathway Yield (g Product/g Glucose)			
Metabolite	**Yeast v7.11**	**Yeast v6.06**	**Yeast v5.01**	**iMM904**
Ethanol	0.511	0.511	0.511	0.511
Isobutanol	0.411	0.411	0.411	0.411
2-Methylbutanol	0.391	0.391	0.391	0.391
3-Methylbutanol	0.391	0.391	0.391	0.391
Succinate	1.124	1.124	1.124	1.124
Lysine	0.696	0.696	0.696	0.696
Citrate	1.422	1.422	1.422	1.422
Glutamate	1.096	1.096	1.096	1.096

In the case of 3-methylbutanol, citrate, and glutamate, the pathway yields calculated by the yeast models assuming cofactor interconversion drastically exceeded the pathway yields calculated by Dugar & Stephanopoulos. A reoccurring pattern found in the individual simulated flux distributions for optimal production of these metabolites was the utilization of oxaloacetate for generation of acetyl-CoA and pyruvate through the intermediates of aspartate, threonine, and glycine. Blocking this mechanism by setting the flux of threonine aldolase (r_1040/THRAi: L-threonine [cytoplasm] → acetaldehyde [cytoplasm] + L-glycine [cytoplasm]) to 0 lowers the calculated pathway yields to values bounded by the yields calculated by Dugar & Stephanopoulos (Table 4.4). Additionally, the pathway yields calculated by Yeast v7.11 and v6.06, but not Yeast v5.01 nor iMM904, for succinate and by all the yeast models for lysine drop

22% below the yields calculated by Dugar & Stephanopoulos. The discrepancy for succinate stems from the ability of the model to produce succinate through reduction of fumarate (r_0455/ FRDcm) instead of alternate routes. This reaction is blocked in Yeast v7.11 and v6.06 due to its usage of cytoplasmic $FADH_2$, in contrast to cytoplasmic $FMNH_2$ in Yeast v5.01 and to mitochondrial $FADH_2$ in iMM904, as the reductant. Since there is no mechanism to regenerate cytoplasmic $FADH_2$ in any of the models, any reactions that use cytoplasmic $FADH_2$ cannot carry any flux. A recent review highlights how our understanding of FAD homeostasis is still limited (Gudipati, Koch et al. 2014). The discrepancy for lysine may be due to differences in the α-aminoadipate pathway found in yeast and fungi and the diaminopimelate pathway found in most bacteria and plants.

Table 4.4: Maximum pathway yields calculated after eliminating flux through threonine aldolase, treating NADH/NADPH as interchangeable, and disregarding compartmentalization of NADH/NADPH in the yeast models. Highlighted cells show decreases compared to Table 4.3.

	Maximum Pathway Yield (g Product/g Glucose)			
Metabolite	**Yeast v7.11**	**Yeast v6.06**	**Yeast v5.01**	**iMM904**
Ethanol	0.511	0.511	0.511	0.511
Isobutanol	0.411	0.411	0.411	0.411
2-Methylbutanol	0.391	0.391	0.391	0.391
3-Methylbutanol	0.326	0.326	0.326	0.326
Succinate	0.874	0.874	1.124	1.124
Lysine	0.541	0.541	0.541	0.541
Citrate	1.066	1.066	1.066	1.066
Glutamate	0.822	0.822	0.822	0.822

4.3.1.4 Assumed dissipation of excess NADH reducing equivalents

Excess reducing equivalents must be reoxidized in order to maintain redox neutrality. The pathway yields calculated by Dugar & Stephanopoulos assume that excess NADH is expended in futile pathways. In their paper, they also calculated an adjusted pathway yield (not shown) assuming glycerol as the predominant sink for NADH. Biochemical yield is not affected by excess NADH under the former assumption but decreases under the latter assumption as glycerol is produced as a by-product. For the *S. cerevisiae* genome-scale metabolic models under the assumption of aerobic respiration, oxidation of the excess NADH can be accomplished through the electron transport chain. The ATP produced through coupling with the electron transport chain can be utilized for additional ATP needs in the case of ATP-deficient pathways or can be expended through futile cycles in the case of ATP-excessive pathways. An example of a futile cycle is the combined action of glutamine synthetase and glutamate synthase in an antagonistic manner that synthesizes and breaks down glutamine with an overall consumption of ATP.

4.3.2 Maximum Pathway Yield of Various Lipids

Our contribution in developing version 7.11 of the Yeast consensus model was to critique and to update the reconstruction of fatty acid, neutral glycerolipid, and glycerophospholipid metabolism in the model (Aung, Henry et al. 2013). In order to evaluate the evolution of the yeast models for lipid metabolism, we extended the methods of the previous section to calculate maximum pathway yields of the major glycerophospholipids, i.e., phosphatidate (PA), phosphatidylinositol (PI), phosphatidylserine (PS), phosphatidylethanolamine (PE), and phosphatidylcholine (PC), and the neutral lipid triacylglycerol (TAG). These calculated yields for glucose minimal media with and without the supplementation of the lipid precursors inositol,

choline, and ethanolamine are shown in Table 4.5. For the purpose of comparison, we calculated maximum pathway yields for the species containing only C18:1 fatty acids. It should be noted that this modeling decision to focus on only C18:1 fatty acids, which is the longest chain length and highest degree of unsaturation typically found in neutral glycerolipids and glycerophospholipids of *S. cerevisiae*, was used as a simplification and to emphasize how differences in reconstruction of the lipid pathways affects calculated yield; the *in vivo* distribution of species consists of C16:0, C16:1, C18:0, and C18:1 fatty acids (Klose, Surma et al. 2012; Hofbauer, Schopf et al. 2014). Calculation of yields for C18:1-containing lipid species was implemented in different ways based on how the model represents lipids. The Yeast v5.01 and v6.06 models categorize individual specific species into general classes using "isa" reactions (e.g. "isa acyl-CoA": hexacosanoyl-CoA $\rightarrow$ acyl-CoA) and subsequently use the term for the general class throughout the reactions in the model. Yeast v7.11 expands the general classes into their corresponding individual species and indicates specific species in the reaction equations. iMM904 represents the general classes as defined composites of specific species. Thus, for Yeast v7.11, the objective function was set to demand reactions for the specific species. For Yeast v5.01 and 6.06, the objective function was set to demand reactions for the general class and all of the "isa acyl-CoA" and "isa fatty acid" reactions except for oleoyl-CoA and oleic acid were given flux bounds of 0, leaving oleoyl-CoA to be the only acyl-CoA and oleic acid the only fatty acid employed in the model. For iMM904, all of the reactions involving acylation were restricted to utilizing only oleoyl-CoA, instead of a mix of various acyl-CoA's. Additionally, to achieve a comparable energetic basis between the models, the ATP yields for glucose were set for all the models to match Yeast v7.11. This was accomplished in Yeast v5.01 and v6.06 by adjusting proton stoichiometry for ATP synthase (r_0226), NADH:ubiquinone oxidoreductase (r_0773),

and ADP/ATP transport (r_1110) and by restricting aspartate transport (r_1117) in only the direction of the cytoplasm to mitochondria. For iMM904, proton stoichiometry for ATP synthase (ATPS3m), ferrocytochrome-c:oxygen oxidoreductase (CYOOm), and ADP/ATP transport (ATPtm-H) was adjusted.

Table 4.5: Maximum pathway yield of C18:1-containing glycerophospholipids and triacylglycerol species calculated by the yeast models

		Maximum Pathway Yield (g Lipid/g Glucose)							
		Base Yeast Model							
Lipid	**Media Supplement**	**v7.11**	**v7.11***	**v6.06†**	**v6.06†‡**	**v5.01†**	**v5.01†‡**	**v5.01†‡§**	**iMM904†**
PA		0.310	0.315	0.364	0.312	0.379	0.323	0.310	0.303
PI		0.348	0.354	0.410	0.349	0.420	0.361	0.347	0.340
	+ Inositol	0.377	0.384	0.449	0.379	0.462	0.393	0.376	0.369
PS		0.332	0.338	0.391	0.334	0.401	0.344	0.330	0.325
PE		0.314	0.320	0.370	0.316	0.379	0.325	0.312	0.307
	+ Ethanolamine	0.324	0.330	0.387	0.326	0.398	0.338	0.323	0.316
PC		0.282	0.304	0.341	0.294	0.355	0.306	0.294	0.290
	+ Choline	0.342	0.349	0.341	0.294	0.355	0.306	0.294	0.334
	+ Ethanolamine	0.289	0.312	0.354	0.302	0.370	0.317	0.304	0.297
TAG		0.266	0.271	0.312	0.271	0.326	0.279	0.268	0.260

* Allowed recycling of glycine into serine by changing mitochondrial serine hydroxymethyltransferase (r_0503) and serine transport (r_2045) to be reversible

† Modified to have same ATP/glucose yield as Yeast v7.11 through changes in the reactions r_0226, r_0773, r_1110, and r_1117 for Yeast v6.06 and v5.01 or the reactions ATPS3m, ATPtm-H, and CYOOm for iMM904

‡ Creation of acetyl-CoA via reversed acetyl-CoA hydrolysis (r_0110) was disallowed

§ Blocked transport of acyl products of mitochondrial fatty acid synthesis to the cytoplasm (r_1780, r_1781, r_1782, r_1783, r_1784, r_1785, r_1786, and r_1787)

4.3.2.1 Impact of assumed reversibility for acetyl-CoA hydrolysis on calculated lipid pathway yields

For unsupplemented media, Yeast v5.01 predicts 20-26% higher lipid pathway yields than Yeast v7.11 and Yeast v6.06 predicts 17-21% higher yields than Yeast v7.11 while iMM904's yields are within 3% of Yeast v7.11 (Table 4.5). The major factor for this discrepancy is the reaction for cytoplasmic acetyl-CoA hydrolase (r_0110). In the Yeast v5.01 and v6.06 models, this reaction is represented as: acetate + coenzyme A + H^+ $\rightarrow$ acetyl-CoA + H_2O. This directionality goes counter to that of hydrolysis and also negates the necessity of ATP to produce acetyl-CoA from acetate. For iMM904 and Yeast v7.11, acetyl-CoA hydrolase operates only in the direction of hydrolysis and is localized only in the mitochondria based on the predominant mitochondrial localization of the enzyme Ach1p (Buu, Chen et al. 2003). If the cytoplasmic reaction r_0110 is removed from the Yeast v5.01 and v6.06 model, the calculated pathway yields for the different lipids decreases by 13-15% (Table 4.5). This large decrease is reasonable considering the ATP requirements needed to produce nine acetyl-CoA units for the creation of stearoyl-CoA (C18:0).

4.3.2.2 Impact of assumed reversibility for TAG hydrolysis on calculated lipid pathway yields

Similar to the case with acetyl-CoA hydrolase, the Yeast v5.01 and v6.06 models allow for the synthesis of triacylglycerol through the reverse direction of TAG lipase (r_1052 and r_1053), i.e., diglyceride + fatty acid $\Leftrightarrow$ triglyceride + H_2O. However, restricting the direction of r_1052 and r_1053 to only hydrolysis decreases the pathway yield for TAG only by 1% for Yeast v6.06 but completely blocks TAG production in Yeast v5.01 (not shown). With the reverse direction disabled in the model, the remaining source of TAG in the Yeast v5.01 model is

diacylglycerol acyltransferase in the lipid particle (r_0336) which catalyzes the reaction acyl-CoA + diglyceride ⇔ coenzyme A + triglyceride. In Yeast v5.01, coenzyme A [lipid particle] is isolated from the pool of coenzyme A found in other compartments. In contrast, Yeast v6.06 and Yeast v7.11 have a transport reaction for coenzyme A between the cytoplasm and lipid particle; iMM904 does not include the lipid particle as a separate compartment in the model so this reaction is assigned to the cytoplasm. Adding a coenzyme A transport reaction to Yeast v5.01 restores TAG production.

4.3.2.3 Impact of assumed function of mitochondrial versus cytoplasmic fatty acid synthesis on calculated lipid pathway yields

The remaining difference between pathway yields calculated by the Yeast v5.01 and v7.11 model can be attributed to the representation of mitochondrial fatty acid synthesis (Figure 4.3). A possible flux distribution in the Yeast v5.01 model is to produce acyl-[acyl carrier protein], i.e., acyl chains bound to acyl carrier protein (ACP), in the mitochondria and to transport and hydrolyze these acyl-ACP into free fatty acids in the cytoplasm. This route requires less ATP stoichiometrically than cytoplasmic fatty acid synthesis since mitochondrial pyruvate dehydrogenase does not hydrolyze ATP to produce acetyl-CoA from pyruvate. In contrast, the production of acetyl-CoA from pyruvate in the cytoplasm, which utilizes a different reaction sequence, requires two ATP equivalents for each acetyl-CoA since acetyl-CoA synthetase hydrolyzes ATP to AMP and pyrophosphate. However, the ability of mitochondrial pyruvate dehydrogenase to provide acetyl-CoA for lipid synthesis has been doubted (Pronk, Yde Steensma et al. 1996) and the simulated result could be considered not representative of actual *in vivo* metabolism.

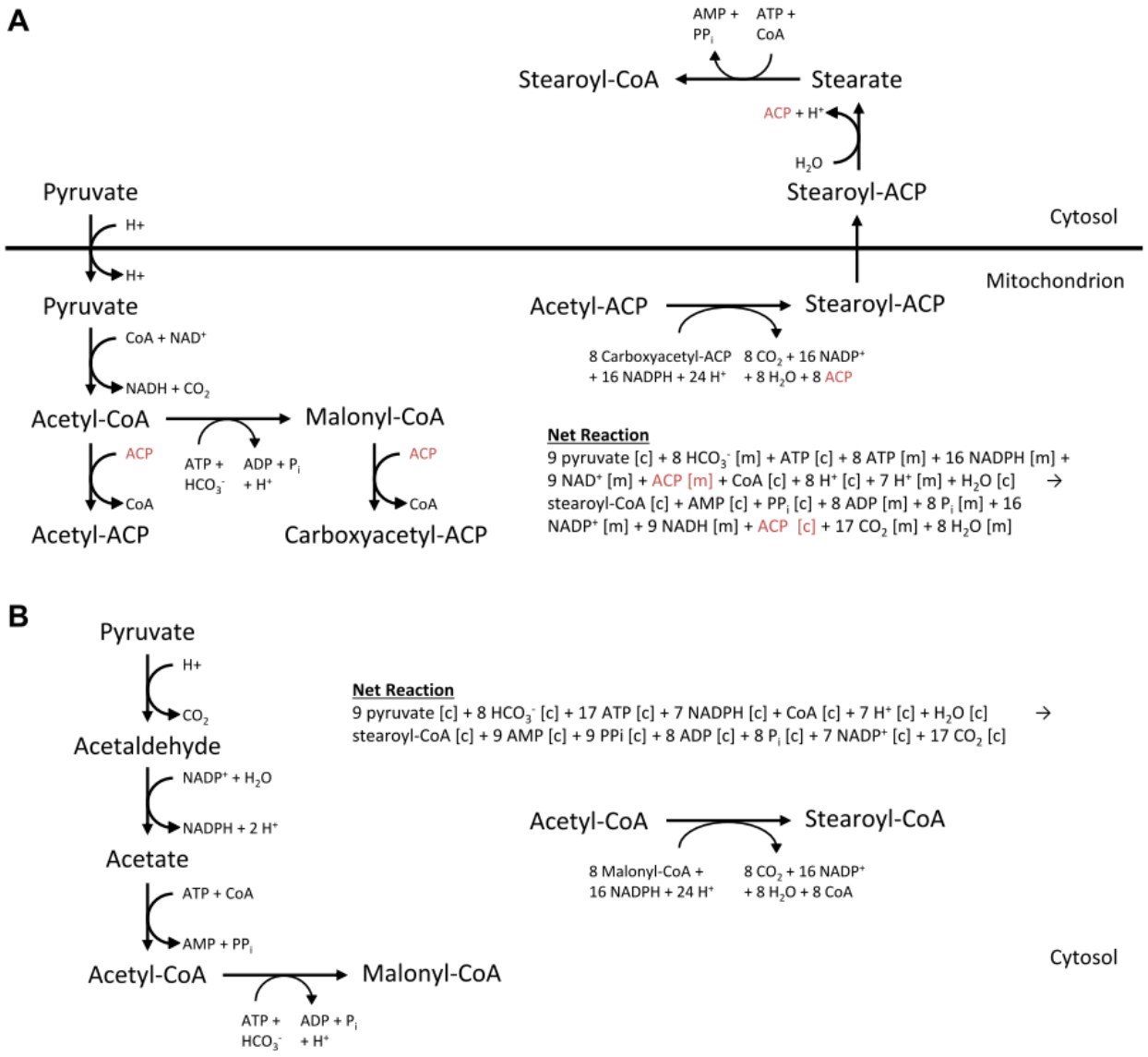


Figure 4.3: Differences between the representation of mitochondrial and cytoplasmic fatty acid synthesis in the yeast models. (A) iMM904, Yeast v5.01, and Yeast v6.06 includes reactions for the production of acyl-ACPs in the mitochondria and for their transport and hydrolysis in the cytoplasm. However, this route is blocked in iMM904 and Yeast v6.06 due to the inclusion of ACP as a balanced species. (B) The reactions for cytoplasmic fatty acid synthesis included in all the yeast models require more ATP in comparison to the counter mitochondrial reactions due to the differences in conversion of pyruvate to acetyl-CoA in the cytoplasm versus mitochondria.

While iMM904 and Yeast v6.06 also contain this same reaction mechanism as Yeast v5.01, the inclusion of acyl carrier protein (ACP) as a metabolite in these reactions leads to blocked mitochondrial fatty acid synthesis. For mitochondrial fatty acid synthesis in iMM904 and Yeast v6.06 model, free mitochondrial ACP are 'consumed' by being bound to acyl chains. When these resulting acyl-ACPs are transported to the cytoplasm and hydrolyzed into free fatty acids and free ACP, there is a deficit in mitochondrial ACP and the mitochondrial pool of ACP cannot be conserved. In contrast, Yeast v5.01 includes ACP only as a boundary metabolite which does not get accounted for in the system of mass balance equations. In contrast to iMM904, Yeast v5.01, and Yeast v6.06, Yeast v7.11 does not tie mitochondrial fatty acid synthesis to the cytoplasmic supply of fatty acids. This is based on evidence for the inability of mitochondrial fatty acid synthesis to compensate for loss of cytoplasmic fatty acid synthesis (Schweizer and Hofmann 2004). Thus, for iMM904, Yeast v6.06, and Yeast v7.11, mitochondrial fatty acid synthesis cannot be employed to save on ATP costs. Restricting transport of acyl products from mitochondrial fatty acid synthesis to the cytoplasm in Yeast v5.01 lowers the calculated pathway yield for all the lipids (Table 4.5).

4.3.2.4 Impact of cofactor specificity of fatty acid desaturase on calculated lipid pathway yields

The change made in Yeast v7.11 for the specification of NADH as the electron donor for fatty acid desaturase made no difference for the pathway yields when applied to Yeast v6.06 and increased yield only by approximately 1% when applied to Yeast v5.01 and iMM904 (not shown). In the reactions r_0837/DESAT16 and r_1013/DESAT18 for fatty acid desaturase, Yeast v6.06 did not include the NADH cofactor while Yeast v5.01 and iMM904 specified NADPH as the cofactor, which is more energetically costly than NADH. The creation of

stearoyl-CoA (C18:0) from 1 acetyl-CoA and 8 malonyl-CoA requires 16 NADPH. Thus, the insensitivity to cofactor specificity in this reaction is to be expected since the one reducing equivalent needed to desaturate C18:0 to C18:1 is miniscule in comparison to the NADPH needed for fatty acid synthesis.

4.3.2.5 Impact of assumed reversibility for one-carbon metabolism on calculated lipid pathway yields

Even after adjusting ATP yield, removing cytoplasmic acetyl-CoA hydrolase, and severing the connection between mitochondrial fatty acid synthesis and supply of cytoplasmic fatty acids, the pathway yield for phosphatidylcholine is still calculated to be 3 to 4% lower in Yeast v7.11 compared to the other yeast models. This can be attributed to the assumed irreversibility of several reactions involved in one-carbon metabolism for Yeast v7.11. S-adenosylmethionine (SAM) is used three times to methylate PE into PC. The one-carbon unit used for methylation can be derived from the amino acid serine. After the transfer of a one-carbon unit from serine to tetrahydrofolate (THF), glycine and methylene THF is produced and methylene THF can ultimately be used to make SAM. Under conditions of limiting serine, there is evidence that glycine can be used to regenerate serine via the mitochondrial glycine cleavage system, serine hydroxymethyltransferase, and transport of the mitochondrial serine to the cytoplasm (Kastanos, Woldman et al. 1997). On a basis of theoretical yield of PC, this recycling of glycine increases the yield since the diversion of glucose for serine synthesis is reduced. In contrast to the other yeast models, the flux distribution for Yeast v7.11 does not reutilize the produced glycine but instead excretes glycine to the extracellular environment. This is due to the assignment of irreversibility for mitochondrial serine hydroxymethyltransferase (r_0503) and serine transport (r_2045). These two processes are considered reversible according to literature

and the directionality is responsive to relative concentrations of glycine and serine (Kastanos, Woldman et al. 1997). Setting these two reactions to be reversible increases the yield of PC and surprisingly also the other lipids examined (Table 4.5). The increase in pathway yields for lipids that do not require SAM may be due to the role of serine-driven one carbon metabolism in production of reducing power. Fan et al. (2014) recently presented evidence for the significant role of serine/folate metabolism in NADPH production (Fan, Ye et al. 2014). However, since many of the reactions involved in one-carbon folate metabolism are reversible, the actual *in vivo* functional role and reaction directionality may vary under different conditions.

4.3.2.6 Calculated pathway yields for media supplemented with lipid precursors

In addition to unsupplemented glucose minimal media, we also examined how the lipid pathway yields change with supplementation of the lipid precursors inositol, choline, and ethanolamine (Table 4.5). It would be expected that the calculated pathway yields of PI would increase under inositol supplementation, PC and PE would increase under ethanolamine supplementation, and PC would increase under choline supplementation because the precursor molecule needed for lipid synthesis would not need to be endogenously produced. The simulated results agreed with these expectations for inositol and ethanolamine supplementation for all tested models. However, only iMM904 and Yeast v7.11 showed increase in yield for PC under choline supplementation. The discrepancy for Yeast v5.01 and Yeast v6.06 is due to the disconnect between the cytoplasmic localization of the water-soluble cytidine diphosphate-activated choline and the membrane localization of the hydrophobic lipid diglyceride, which does not allow these two metabolites to interact. Yeast v7.11 addresses this issue by assuming transport reactions of water soluble lipid precursors from the cytoplasm to the membrane while

iMM904 has the reaction localization for diacylglycerol cholinephosphotransferase simplified to the cytoplasm.

4.4 Discussion

4.4.1 Sensitivity to Assumed ATP Yields and Cofactor Imbalance

Given that the stoichiometry for proton transport in the electron transport chain and H^+/ATP ratio for ATP synthase is not certain (Brand 2005), we examined how the assumed maximum ATP yield affects the calculated maximum pathway yields for eight different chemicals of interest in industrial biotechnology. The pathway yields for six of these chemicals (i.e, 2-methylbutanol, 3-methylbutanol, succinate, lysine, citrate, and glutamate) were notably sensitive to the assumed maximum ATP yield (Figure 4.2). The sensitivity to ATP yield stems not only from the demand of ATP directly within the pathway for biosynthesis of the metabolite of interest but also for supply of NADPH. ATP can be utilized for NADPH production through the phosphorylation of glucose for the pentose phosphate pathway and through NADH kinase (Hou, Lages et al. 2009) or a transhydrogenase-like shunt (Suga, Matsuda et al. 2013) that generates NADPH at the expense of ATP and NADH. The addition of hypothetical transhydrogenase reactions (NADH + $NADP^+$ ⇔ NAD^+ + NADPH) to the genome-scale metabolic model, in general, increased the calculated pathway yields for all of the biochemicals except for ethanol (Table 4.3). The degree of response of each version of the *S. cerevisiae* metabolic model to the addition of the transhydrogenase reactions is partially influenced by factors such as what the model assumes for NADH/NADPH cofactor specificity, reversibility of various reactions, and stoichiometry associated with ATP production through oxidative phosphorylation.

4.4.2 Bypasses and Gain/Loss of Functionality with Changes to Reversibility of Reactions

Approaches to the assignment of reversibility include estimation of Gibbs free energy of reaction, use of heuristic rules, or use of reported reaction directionalities in literature and textbooks (Thiele and Palsson 2010). The assignment of (ir)reversibility to reactions in the metabolic model can have impacts on simulated flux distributions. An example discussed was the assumed direction of acetyl-CoA hydrolase (acetyl-CoA + H_2O ⇔ acetate + coenzyme A + H^+). The assignment of reversibility to this reaction allows for an *in silico* bypass to the *in vivo* requirement of ATP to produce acetyl-CoA from acetate through acetyl-CoA synthetase. This change alone lead to significantly higher calculated lipid pathway yields in Yeast v5.01 and v6.06 (Table 4.5).

For reactions accepted to be reversible, the *in vivo* reaction direction is sensitive to intracellular concentrations. The assignment of reversibility for general physiological conditions could, in certain scenarios, lead to discrepancies between *in silico* and *in vivo* utilization of reaction pathways. An example that highlights this uncertainty is one carbon metabolism in which most of the reactions are considered reversible (Kastanos, Woldman et al. 1997). The choice of reaction direction can determine how glycine produced from synthesis of S-adenosylmethionine or from threonine catabolism is utilized. This had impacts on calculated maximum pathway yields of the examined lipids (Table 4.5) and of 3-methylbutanol, succinate, lysine, citrate, and glutamate (Table 4.4). Therefore, it may be worthwhile to apply condition-specific constraints, but they must be applied to a model in such a way that specifies which conditions they are appropriate for or what justification is used for imposing those constraints.

4.4.3 Significance of Compartmentalization

Compartmentalization in eukaryotic cells allows for the separation of biochemical processes into different microenvironments. These different microenvironments may exhibit distinct functions or utilized substrates for the same corresponding enzymes (Hao, Ma et al. 2010). An example is the distinction between cytoplasmic and mitochondrial fatty acid synthesis. Cytoplasmic fatty acid synthesis is responsible for production of fatty acids of 16 or 18 carbon chain length while *in vivo* evidence for the contribution of mitochondrial fatty acid synthesis to long-chain fatty acid production is lacking (Schweizer and Hofmann 2004). The assumption of interchangeable function between these two systems would enable simulation of a less energetically costly route to fatty acid synthesis due to the ATP requirements for acetyl-CoA synthesis in the cytoplasm versus the mitochondria (Figure 4.3).

Many metabolic processes rely on coordinated interactions between different compartments where a product from one compartment feeds into another compartment. Transport reactions or shuttling mechanisms must be included in the reconstructed metabolic network to connect these two distinct compartments. Examples of dead-ends include coenzyme A in the lipid particle for Yeast v5.01 and CDP-choline in the endoplasmic reticulum for Yeast v5.01 and v6.06. In the case of iMM904, the generalized localization of many of the reactions involved in lipid metabolism to the cytosol obviates the need for many of the transport reactions.

Compartmentalization also allows for proton gradients that can drive ATP synthesis through ATP synthase. Accounting for this proton gradient can be problematic in that many transport reactions are in symport or antiport with protons (Thiele and Palsson 2010). Too many reversible transport reactions in the metabolic model may ultimately lead to simulated reaction cycling that allows for free transport of protons out of the mitochondria. This was the case for

Yeast v6.06 (Figure 4.1). This aberration in proton shuttling was signaled by a high maximum ATP yield. The other examined models of *S. cerevisiae* metabolism did not have this issue due to constrained reaction directionality and/or missing transport reactions.

4.5 Conclusions

Genome-scale metabolic models condense our knowledge of the inventory of biochemical reactions occurring in cells and can be applied to identify limits on the maximum yield of different metabolites based on the stoichiometry and reversibility of reactions in metabolic pathways. The maximum conversion yields that are calculated are sensitive to the modeling decisions employed as evidenced by differences in calculated pathway yields from four different versions of *S. cerevisiae* genome-scale metabolic models. These pathway yields represent a theoretical endeavor to question what the absolute maximum would be if the reconstructed metabolic network was solely used to produce a particular metabolite. Accurate prediction of actual *in vivo* product yields would require more complexity in modeling such as a biologically relevant objective function and incorporation of enzyme kinetics and regulation.

4.6 Methods

Versions 7.11, 6.06, and 5.01 of the consensus model of *S. cerevisiae* metabolism were downloaded from the yeastnet website (http://yeast.sourceforge.net/) and iMM904 was downloaded from the additional files linked to its online article (Mo, Palsson et al. 2009). These models were analyzed using the COBRA Toolbox 2.0.5 (Schellenberger, Que et al. 2011) in the MATLAB® R2013a environment (MathWorks, Natick, MA). Aerobic glucose-limited media was simulated by setting the exchange of ammonium, carbon dioxide, protons, oxygen, phosphate, sulfate, and water to be unconstrained and the glucose exchange to -1 mmol/gDW/h. The maximum pathway yield for each metabolite of interest was determined using the COBRA

function “optimizeCbModel” to perform flux balance analysis under steady-state conditions. The objective was set as maximization of the exchange of the metabolite of interest. This optimization problem was solved using Gurobi 5.6.0 (Gurobi Optimization, Houston, TX). The maximum pathway yield on a mole-basis (i.e., mol product/mol glucose) is simply the maximal flux value for the exchange of the metabolite of interest since the glucose exchange was set to unity.

REFERENCES

Aung, H. W., S. A. Henry and L. P. Walker (2013). "Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism." Industrial Biotechnology **9**(4): 215-228.

Bakker, B. M., K. M. Overkamp, A. J. A. van Maris, P. Kötter, M. A. H. Luttik, J. P. van Dijken and J. T. Pronk (2001). "Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae." FEMS Microbiology Reviews **25**(1): 15-37.

Brand, M. (2005). "The efficiency and plasticity of mitochondrial energy transduction." Biochemical Society Transactions **33**(5): 897-904.

Buu, L.-M., Y.-C. Chen and F.-J. S. Lee (2003). "Functional Characterization and Localization of Acetyl-CoA Hydrolase, Ach1p, in Saccharomyces cerevisiae." Journal of Biological Chemistry **278**(19): 17203-17209.

Dugar, D. and G. Stephanopoulos (2011). "Relative potential of biosynthetic pathways for biofuels and bio-based products." Nature biotechnology **29**(12): 1074-1078.

Durot, M., P.-Y. Bourguignon and V. Schachter (2009). "Genome-scale models of bacterial metabolism: reconstruction and applications." FEMS Microbiology Reviews **33**(1): 164-190.

Fan, J., J. Ye, J. J. Kamphorst, T. Shlomi, C. B. Thompson and J. D. Rabinowitz (2014). "Quantitative flux analysis reveals folate-dependent NADPH production." Nature **510**(7504): 298-302.

Gudipati, V., K. Koch, W.-D. Lienhart and P. Macheroux (2014). "The flavoproteome of the yeast Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1844**(3): 535-544.

Hao, T., H.-W. Ma, X.-M. Zhao and I. Goryanin (2010). "Compartmentalization of the Edinburgh Human Metabolic Network." BMC Bioinformatics **11**(1): 393.

Heavner, B. D., S. A. Henry and L. P. Walker (2012). "Evaluating sphingolipid biochemistry in the consensus reconstruction of yeast metabolism." Industrial Biotechnology **8**(2): 72-78.

Heavner, B. D., K. Smallbone, B. Barker, P. Mendes and L. P. Walker (2012). "Yeast 5–an expanded reconstruction of the Saccharomyces cerevisiae metabolic network." BMC Systems Biology **6**(1): 55.

Heavner, B. D., K. Smallbone, N. D. Price and L. P. Walker (2013). "Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance." Database **2013**: bat059.

Herrgard, M. J., N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Bluthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W.

Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasie, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kurdar, M. Penttila, E. Klipp, B. O. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen and D. B. Kell (2008). "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology." Nat Biotech **26**(10): 1155-1160.

Hofbauer, Harald F., Florian H. Schopf, H. Schleifer, Oskar L. Knittelfelder, B. Pieber, Gerald N. Rechberger, H. Wolinski, Maria L. Gaspar, C. O. Kappe, J. Stadlmann, K. Mechtler, A. Zenz, K. Lohner, O. Tehlivets, Susan A. Henry and Sepp D. Kohlwein (2014). "Regulation of Gene Expression through a Transcriptional Repressor that Senses Acyl-Chain Length in Membrane Phospholipids." Developmental Cell **29**(6): 729-739.

Hou, J., N. F. Lages, M. Oldiges and G. N. Vemuri (2009). "Metabolic impact of redox cofactor perturbations in Saccharomyces cerevisiae." Metabolic Engineering **11**(4–5): 253-261.

Kastanos, E. K., Y. Y. Woldman and D. R. Appling (1997). "Role of Mitochondrial and Cytoplasmic Serine Hydroxymethyltransferase Isozymes in de Novo Purine Synthesis in Saccharomyces cerevisiae." Biochemistry **36**(48): 14956-14964.

Klose, C., M. A. Surma, M. J. Gerl, F. Meyenhofer, A. Shevchenko and K. Simons (2012). "Flexibility of a Eukaryotic Lipidome – Insights from Yeast Lipidomics." PLoS ONE **7**(4): e35063.

Mo, M., B. Palsson and M. Herrgard (2009). "Connecting extracellular metabolomic measurements to intracellular flux states in yeast." BMC Systems Biology **3**(1): 37.

Pronk, J. T., H. Yde Steensma and J. P. Van Dijken (1996). "Pyruvate Metabolism in Saccharomyces cerevisiae." Yeast **12**(16): 1607-1633.

Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke and B. O. Palsson (2011). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0." Nature Protocols **6**(9): 1290-1307.

Schweizer, E. and J. Hofmann (2004). "Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems." Microbiology and Molecular Biology Reviews **68**(3): 501-517.

Suga, H., F. Matsuda, T. Hasunuma, J. Ishii and A. Kondo (2013). "Implementation of a transhydrogenase-like shunt to counter redox imbalance during xylose fermentation in Saccharomyces cerevisiae." Applied Microbiology and Biotechnology **97**(4): 1669-1678.

Thiele, I. and B. Ø. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature Protocols **5**(1): 93-121

Chapter 5: Training the Model: Using Experimental Data to Define Flux Constraints

5.1 Abstract

Growth of *S. cerevisiae* BY4742 in I+ glucose synthetic media was simulated using dynamic flux balance analysis with the Yeast v7.5 consensus model. Experimental data on dry cell weight, glucose, ethanol, glycerol, and acetate concentration was used to calibrate the Yeast v7.5 model. This was accomplished by using different functions to model the concentration data and by calculating specific rates for growth, glucose consumption, and product formation over time from these models. Two different functional forms were tested for fitting of dry cell weight concentration: (1) piece-wise defined function with a lag phase and logistic growth and (2) a generalized logistic equation. Two different functional forms were also tested for fitting of glucose concentration: (1) Monod equation for specific glucose consumption rate and (2) a generalized logistic equation for glucose concentration. The generalized logistic equation was used to fit ethanol, glycerol, and acetate concentration. Even though all the tested functions fit the concentration data extremely well ($R^2 > 0.99$), the different functions tested for cell dry weight and glucose concentration yielded distinctly different profiles for specific growth rate and specific glucose consumption rate.

The experimentally-derived specific rates were used as exact flux constraints in the Yeast v7.5 model. For certain time periods, the calculated specific rates were found to yield an infeasible solution based on flux balance analysis. This infeasibility arose because the set of specific rates at those times lead to violations of mass-balance in the flux balance analysis

framework. Since flux balance analysis requires mass-balance to always be satisfied, the carbon balance of products to glucose should not exceed 1.

5.2 Introduction

Genome-scale metabolic reconstructions encompass the entirety of an organism's metabolic network. However, not all of reactions in the metabolic network will be active in a given time and the level of activity of the reactions will vary depending on conditions. Accordingly, the metabolic reconstruction serves as a scaffold for incorporation of additional constraints to generate a model specific to a particular condition.

The approach of dynamic flux balance analysis has been developed to simulate cellular behavior in batch growth (Varma and Palsson 1994). In this approach, the time course of batch cultivation is discretized into small segments assumed to be at quasi-steady state. Flux balance analysis within each of these time steps is used to predict the flux distribution for that time step and the predicted exchange fluxes are used to calculate the extracellular metabolite concentration throughout the time step. The extracellular metabolite concentrations at the end of the time step then limit the flux for nutrient uptake in the next time step. The flux distribution within each time step can be further constrained by incorporating experimental data on consumption and production of extracellular metabolites. These constraints have previously been derived through a Monod equation for specific glucose consumption rate (Mahadevan, Edwards et al. 2002; Hjersted, Henson et al. 2007; Sánchez, Pérez-Correa et al. 2014), determination of upper limits to oxygen and glucose uptake (Varma and Palsson 1994), assuming specific production rate was proportional to the specific glucose consumption rate (Sánchez, Pérez-Correa et al. 2014), and differentiation of curve fits of metabolite and biomass concentrations (Zanghellini, Natter et al. 2008; Colombié, Nazaret et al. 2015).

In this study, experimental data on dry cell weight and extracellular metabolite concentrations over time was utilized to calibrate the Yeast v7.5 genome-scale metabolic model for simulation of batch cultivation of *S. cerevisiaie* BY4742 in I+ glucose synthetic media. This was accomplished by transforming the concentration data into time-dependent specific rates that could be used as flux constraints for growth and exchange reactions in the Yeast v7.5 model.

5.3 Methods

5.3.1 Strain, Media, and Growth Conditions

Saccharomyces cerevisiae wild-type strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was obtained from Dr. Susan Henry's lab and stored as 40% glycerol stocks at -80°C. All experiments were performed using glucose synthetic complete media containing inositol (I+). This I+ media contained (per liter): 20 g of glucose; 5 g of ammonium sulfate ($(NH_4)_2SO_4$); 0.982 g of potassium phosphate (KH_2PO_4); 0.491 of g magnesium sulfate ($MgSO_4$); 98.2 mg of sodium chloride (NaCl); 0.13 g of calcium chloride ($CaCl_2 \cdot 2H_2O$); 0.543 mg of boric acid (H_3BO_3); 0.0435 mg of cupric sulfate ($CuSO_4 \bullet 5\ H_2O$); 0.109 mg of potassium iodide (KI); 0.217 mg of ferric chloride ($FeCl_3 \bullet 6H_2O$); 0.435 mg of manganese sulfate ($MnSO_4 \bullet H_2O$); 0.217 mg of sodium molybdate ($MoNa_2O_4$); 0.435 mg of zinc sulfate ($ZnSO_4 \bullet 7H_2O$); 1.96 μg of biotin; 392 μg of calcium pantothenate; 1.96 μg of folic acid; 392 μg of niacin; 196 μg of *p*-aminobenzoic acid; 392 μg of pyridoxine hydrochloride; 196 μg of riboflavin; 392 μg of thiamine, HCl; 20 mg of adenine sulfate; 20 mg of L-arginine HCl; 20 mg of L-histidine HCl; 60 mg of L-leucine; 230 mg of L-lysine HCl; 20 mg of L-methionine; 300 mg of L-threonine; 20 mg of L-tryptophan; 40 mg of uracil; and 13.5 mg *myo*-inositol.

BY4742 cells from the -80°C frozen stock were streaked onto solid I+ medium containing 2% (w/v) agar and incubated at 30°C for 2.5 days. Five colonies from the I+ agar

plate were collected with an inoculation loop and transferred to 50 mL of I+ media in glass media bottles. This pre-inoculum culture was incubated at 30°C and 200 rpm in a shaking water bath for 22 hr. Five milliliters of the pre-inoculum was transferred to 250 mL of I+ media in glass media bottles and incubated at 30°C and 200 rpm in a shaking water bath for 16 hr. BioFlo 310 2.5 L benchtop fermentors (Eppendorf AG, Germany) containing 1.5 L of I+ media were inoculated with the appropriate volume of the inoculum to achieve an initial OD_{600} of 0.1. Benchtop fermenters were operated at 30°C, 200 rpm, and 1 L/min aeration.

5.3.2 Biomass Quantification

Cell growth was monitored by measuring the optical density at 600 nm using a BioPhotometer Plus (Eppendorf AG, Germany). Dry cell weight was determined by filtering 15 mL of fermentor samples through 0.4 µm Whatman nuclepore track-etched membranes (GE Healthcare Life Sciences, UK) and drying in an oven at 105°C for 1 day to get a constant weight. Dry cell weight was measured using a Mettler-Toledo XP1203S precision balance (Mettler-Toledo AG, Switzerland). The following relationship was determined for relating optical density at 600 nm to cell dry weight concentration:

$$CDW = 0.1726 * OD_{600} + 0.079 \qquad \text{Equation 5.1}$$

where:

CDW = cell dry weight concentration, ($\frac{gDW}{L}$)

OD_{600} = optical density at 600 nm

This correlation had $R^2 = 0.9891$.

5.3.3 Extracellular Metabolite Quantification

Glucose, ethanol, glycerol, and acetate concentration in the fermentors was determined by high performance liquid chromatography using an Aminex HPX-87H Organic Acid Column

(Bio-Rad, USA) with a SPD-20A UV-Vis detector (Shimadzu, Japan) at 210 nm and a RID-10A refractive index detector (Shimadzu, Japan). Samples were injected with a volume of 20 µL and eluted with 5 mM H_2SO_4 at a flow rate of 0.6 mL/min.

5.3.4 Modeling of Experimental Data

Modeling of experimental data for dry cell weight, glucose, ethanol, glycerol, and acetate concentration was performed using the Curve Fitting Toolbox in MATLAB® R2013a (MathWorks, Natick, MA). The equations used in these curve fits are described in the subsections below. The Curve Fitting Toolbox uses the method of least-squares in which the sum of squared residuals is minimized to estimate the parameters in the curve fits.

5.3.4.1 Modeling of dry cell weight

Two different models for fitting the data on dry cell weight concentration over time were examined. In the first model form, the rate of change in dry cell weight concentration was described using piece-wise defined derivatives (Equation 5.2).

$$\frac{dX}{dt} = \begin{cases} 0 \; for \; t \leq t_{lag} \\ \mu_m \left(1 - \frac{X}{X_m}\right) X \; for \; t > t_{lag} \end{cases} \qquad \text{Equation 5.2}$$

where:

$\frac{dX}{dt}$ = time derivative of dry cell weight concentration, ($\frac{gDW}{L-hr}$)

μ_m = maximum specific growth rate, ($\frac{1}{hr}$)

X = dry cell weight concentration at time t, ($\frac{gDW}{L}$)

X_m = maximum dry cell weight concentration, ($\frac{gDW}{L}$)

t_{lag} = lag phase duration, (hr)

It is assumed that during the lag phase ($t \leq t_{lag}$) the dry cell weight concentration remains constant. After the lag phase ($t > t_{lag}$), I assumed logistic growth where the rate of change in dry cell weight concentration is proportional to the current dry cell weight concentration and to the fraction of the remaining carrying capacity. This carrying capacity is represented by the maximum dry cell w eight concentration the system can support, X_m. The following is the integrated form of Equation 5.2 using X_o as the initial dry cell weight concentration:

$$X = \begin{cases} X_o \; for \; t \leq t_{lag} \\ \dfrac{1}{\frac{1}{X_m} + e^{-\mu_m (t - t_{lag})}(\frac{1}{X_o} - \frac{1}{X_m})} \; for \; t > t_{lag} \end{cases} \qquad \text{Equation 5.3}$$

The second model fitted to experimental data was a single generalized logistic function shown in Equation 5.4:

$$X = A + \frac{K - A}{(1 + Qe^{-B(t-M)})^{1/v}} \qquad \text{Equation 5.4}$$

where:

A = lower asymptote, ($\frac{gDW}{L}$)

K = upper asymptote, ($\frac{gDW}{L}$)

B = growth rate, ($\frac{1}{hr}$)

v = parameter that affects near which asymptote maximum growth occurs

Q = parameter that depends on the value X(0)

M = time of maximum growth if $Q = v$, (hr)

t = time, (hr)

Several of the parameters in Equation 5.4 cannot be uniquely estimated because the parameters appear in groups as products. The term $Qe^{-B(t-M)}$ in Equation 5.4 can be expanded as $Qe^{-Bt}e^{BM}$, which can have different sets of values for Q and e^{BM} that yield observationally

equivalent results. Thus, this product was lumped together to generate a new parameter Q' and the following equation was used for curve fits:

$$X = A + \frac{K - A}{(1 + Q'e^{-Bt})^{1/v}} \quad \text{Equation 5.5}$$

5.3.4.2 Modeling of glucose consumption

Two different models for fitting the glucose concentration data were examined. In the first model, a Monod expression for specific glucose consumption rate was used:

$$v_g = \frac{v_{g,max} G}{K_g + G} \quad \text{Equation 5.6}$$

where:

v_g = specific glucose consumption rate, ($\frac{g}{gDW-hr}$)

$v_{g,max}$ = maximum specific glucose consumption rate, ($\frac{g}{gDW-hr}$)

G = glucose concentration, ($\frac{g}{L}$)

K_g = glucose saturation constant, ($\frac{g}{L}$)

The rate of change of glucose concentration was determined by:

$$\frac{dG}{dt} = v_g X \quad \text{Equation 5.7}$$

The parameters for the Monod expression for glucose consumption was determined through least-squares minimization for the solution of Equation 5.7 obtained from the MATLAB solver *ode45*. Thus, Equation 5.7 and the MATLAB function *nlinfit* was used to determine the parameter values for Equation 5.6 that minimizes the sum of squared residuals. The MATLAB function *nlparci* was used to determine the 95% confidence intervals for the parameter estimates.

The second model explored is more straightforward and fits the generalized logistic function (with a lower asymptote of 0) to the glucose concentration directly:

$$G = \frac{K}{(1 + Q'e^{-Bt})^{1/v}}$$ Equation 5.8

5.3.4.3 Modeling of ethanol, glycerol, and acetate production

The generalized logistic function was used to fit the concentration of ethanol (Equation 5.9), glycerol (Equation 5.10), and acetate (Equation 5.11) over time.

$$E = \frac{K}{(1 + Q'e^{-Bt})^{1/v}}$$ Equation 5.9

$$Glyc = \frac{K}{(1 + Q'e^{-Bt})^{1/v}}$$ Equation 5.10

$$Ac = \frac{K}{(1 + Q'e^{-Bt})^{1/v}}$$ Equation 5.11

5.3.4.4 Derivation of fluxes from modeling of experimental data

Specific growth, substrate consumption, and product formation rates were calculated from the models for dry cell weight (Equation 5.3 or Equation 5.5), glucose (Equation 5.8), ethanol (Equation 5.9), glycerol (Equation 5.10), and acetate (Equation 5.11) concentration using Equation 5.12 to Equation 5.16, respectively.

$$\mu = \frac{\frac{dX}{dt}}{X}$$ Equation 5.12

$$v_g = \frac{\frac{dG}{dt}}{X}$$ Equation 5.13

$$v_e = \frac{\frac{dE}{dt}}{X}$$ Equation 5.14

$$v_{glyc} = \frac{\frac{dGlyc}{dt}}{X}$$ Equation 5.15

$$v_{ac} = \frac{\frac{dAc}{dt}}{X}$$ Equation 5.16

5.3.5 Genome-Scale Metabolic Model

Version 7.5 of the consensus model of *S. cerevisiae* metabolism (Yeast v7.5) was used in this study and was downloaded from the yeastnet website (http://yeast.sourceforge.net/). This current version is the outcome of curation efforts documented in Chapter 4. This model was utilized in the MATLAB® R2013a environment (MathWorks, Natick, MA) using the COBRA Toolbox 2.0.5 (Schellenberger, Que et al. 2011). The BY4742 strain background (*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was simulated by setting the flux bounds of the following reactions to zero: 3-isopropylmalate dehydrogenase (r_0061), imidazoleglycerol-phosphate dehydratase (r_0564), L-aminoadipate-semialdehyde dehydrogenase (NADPH) (r_0678), and orotidine-5'-phosphate decarboxylase (r_0821). I+ glucose synthetic media was simulated by allowing for uptake of media components through negative flux bounds for the following exchange reactions: adenine (r_1639), ammonium (r_1654), biotin (r_1671), D-glucose (r_1714), folate (r_1792), H^+ (r_1832), iron (r_1861), L-arginine (r_1879), L-histidine (r_1893), L-leucine (r_1899), L-lysine (r_1900), L-methionine (r_1902), L-threonine (r_1911), L-tryptophan (r_1912), myo-inositol (r_1947), nicotinate (r_1967), oxygen (r_1992), phosphate (r_2005), potassium (r_2020), pyridoxine (r_2028), riboflavin (r_2038), sodium (r_2049), sulphate (r_2060), thiamine (r_2067), uracil (r_2090), and H_2O (r_2100).

To simulate batch growth using the Yeast v7.5 model, the time course of the fermentation was discretized into intervals defined by a uniform time step Δt. Each interval was assumed to be in quasi-steady state such that there is no accumulation or depletion of intracellular metabolites and the fluxes are constant over the small time interval. The assumption of constant fluxes and

intracellular metabolite pools is based upon "the typical time constant of metabolic transients [being] relatively faster than the simulation time step" (Min Lee, Gianchandani et al. 2008), thus rapidly achieving new steady states throughout the course of growth.

For each time interval, flux balance analysis (FBA) was conducted at the beginning of the interval to calculate a feasible flux distribution (Equation 5.17). This flux distribution was constrained by the specific rates calculated from models of experimental data. The calculated specific growth rate was set as the flux value for the growth reaction (r_2111); the calculated specific glucose consumption rate as the flux value for the glucose exchange reaction (r_1714); the calculated specific glycerol production rate as the flux value for the glycerol exchange reaction (r_1808); the calculated specific ethanol production rate as the flux value for the ethanol exchange reaction (r_1761); and the calculated specific acetate production rate as the flux value for the acetate exchange reaction (r_1634). Despite constraining these fluxes, multiple flux distributions that satisfy all the constraints exist. Thus, a particular solution was found by minimizing the L1-norm of the flux distribution to generate a minimal flux distribution that satisfies the imposed flux constraints. This is akin to assuming parsimonious enzyme use by the cell if flux is related to enzyme concentration (Smallbone and Simeonidis 2009).

$$\begin{aligned} &\min|v_i| \\ &s.t.\ \ S \cdot v_i = 0 \\ &\qquad v_{min,i} \leq v_i \leq v_{max,i} \end{aligned} \qquad \text{Equation 5.17}$$

where:

S = stoichiometric matrix of m metabolites and n reactions

v_i = flux vector at the ith time point, $(\frac{mmol}{g\ DW-hr})$

$v_{min,i}$ = lower bound flux vector at the ith time point, $(\frac{mmol}{g\ DW-hr})$

$v_{max,i}$ = upper bound flux vector at the *i*th time point, ($\frac{mmol}{g\ DW-hr}$)

The concentration of biomass (Equation 5.18) and extracellular metabolites (Equation 5.19) at the end of the interval was determined based on forward difference approximations using the resulting fluxes from FBA. The concentration of biomass and extracellular metabolites were then used to update the lower bound constraints for the fluxes of exchange reactions in the next time interval (Equation 5.20). The process of solving for a particular flux distribution in the time interval, determining concentration of biomass and extracellular metabolites at the end of the time interval, and updating lower bound flux constraints for exchange reactions in the next time interval was repeated for each time interval.

$$X_{i+1} = X_i(1 + \mu_i \Delta t) \quad \text{Equation 5.18}$$

$$[ext.mets]_{i+1} = [ext.mets]_i + v_{exchange,i} X_i \Delta t \quad \text{Equation 5.19}$$

$$v_{exchange,i+1} \geq -\frac{[ext.mets]_{i+1}}{X_{i+1}\Delta t} \quad \text{Equation 5.20}$$

where:

X_i = dry cell weight concentration at the *i*th time point, ($\frac{gDW}{L}$)

X_{i+1} = dry cell weight concentration at the *(i+1)*th time point, ($\frac{gDW}{L}$)

μ_i = specific growth rate at the *i*th time point, ($\frac{1}{hr}$)

Δt = time step of intervals, (*hr*)

$[ext.mets]_{i+1}$ = extracellular metabolite concentrations at the *(i+1)*th time point, (*mM*)

$[ext.mets]_i$ = extracellular metabolite concentrations at the *i*th time point, (*mM*)

$v_{exchange,i}$ = flux vector for exchange reactions at the *i*th time point, ($\frac{mmol}{g\ DW-hr}$)

$v_{exchange,i+1}$ = flux vector for exchange reactions at the *(i+1)*th time point, ($\frac{mmol}{g\ DW-hr}$)

5.4 Results

The approach used to incorporate experimental data as a constraint on the genome-scale metabolic model is summarized in Figure 5.1. In order to utilize the experimental data on cell dry weight, glucose, ethanol, glycerol, and acetate concentration, different mathematical functions were used to model the data and were evaluated. These models allow for estimation of concentrations between measured points. Specific growth, substrate consumption, and product formation rates were calculated from these models based on Equation 5.12 to Equation 5.16, respectively. The specific rates were then used as flux constraints within the genome-scale metabolic model, which further limits the possible metabolic behavior of the simulated cell culture. This constraining effectively "trains" the model (Zanghellini, Natter et al. 2008) and contextualizes the external response of the cell culture to the media environment.

Within the dynamic flux balance analysis framework, the fermentation time course is discretized into small time intervals. The flux within each of the time intervals is determined by flux balance analysis under a steady-state assumption and by flux constraints (Equation 5.17). These fluxes then determine the concentration of biomass (Equation 5.18) and extracellular metabolites (Equation 5.19) within the time interval. These new concentrations are used to update the bounds on flux values for the next time step's FBA (Equation 5.20). This cycle of FBA and calculation of new concentrations and flux bounds is iterated for the complete fermentation time course.

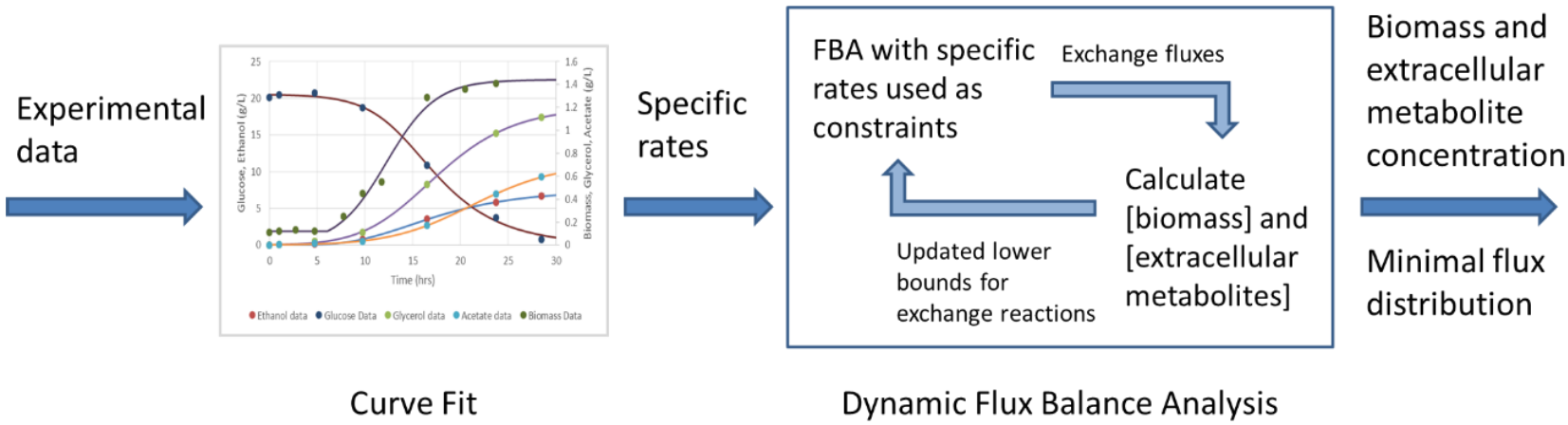


Figure 5.1: Overview of the approach used to incorporate experimental data as a constraint on the genome-scale metabolic model. The outputs of the dynamic flux balance analysis framework are profiles of biomass and extracellular metabolite concentration over time and the minimal flux distribution for each time interval that satisfies the imposed flux constraints.

5.4.1 Kinetic Analysis of Batch Growth of *S. cerevisiae* in I+ Glucose Synthetic Media

5.4.1.1 Specific growth rate

Experimental data on dry cell weight concentration over time is shown in Figure 5.2. A plateau in dry cell weight concentration of approximately 1.4 gDW/L appears to be achieved by BY4742 growing in I+ glucose synthetic media. This is comparable to the value of 1.349 gDW/L determined at the end of fermentation by Ng, Jung et al. (2012) for BY4742 growing in glucose synthetic complete dropout media.

The pattern of growth is sigmoidal in shape and thus functions capable of describing this shape were explored. The estimated parameters, confidence regions, and R^2 value for the models of experimental data on dry cell weight concentration over time are listed in Table 5.1 and the plotted fits are shown in Figure 5.2. Although there are slight differences in how the models describe the profile of dry cell weight concentration, both models show an excellent fit to the experimental data ($R^2 > 0.99$).

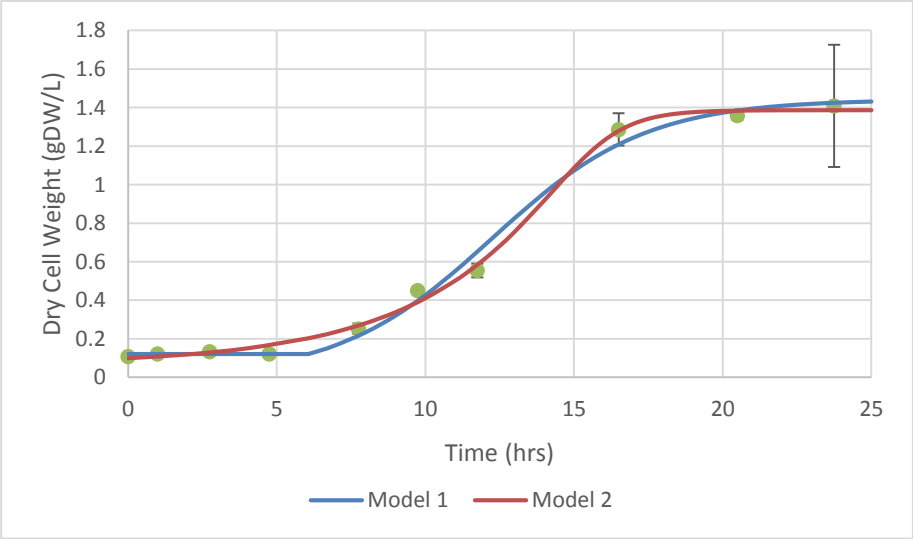


Figure 5.2: Dry cell weight concentration for batch growth of wild-type *S. cerevisiae* in I+ synthetic media as predicted by Model 1 (piece-wise function consisting of lag phase and logistic growth) and by Model 2 (generalized logistic equation). Data points are the mean of three reactors and error bars show the standard deviation of these three reactors.

Table 5.1: Curve fit of dry cell weight concentration data[a]

Model 1: Lag phase followed by logistic growth (Equation 5.3)	$\mathbf{X_o}$ (gDW/L)	$\mathbf{X_m}$ (gDW/L)	$\mathbf{t_{lag}}$ (hr)	$\mathbf{\mu_m}$ (1/hr)	$\mathbf{R^2}$	
	0.1212 (0.106, 0.136)	1.442 (1.194, 1.689)	6.06 (2.71, 9.42)	0.3867 (0.136, 0.638)	0.9924	
Model 2: Generalized logistic equation (Equation 5.5)	**A** (gDW/L)	**B** (1/hr)	**K** (gDW/L)	**Q'**	**ν**	$\mathbf{R^2}$
	0.06605 (-0.068, 0.2)	1.016 (-0.517, 2.55)	1.387 (1.311, 1.464)	8.42E6 (-2.2E8, 2.37E8)	4.301 (-5.386, 13.99)	0.997

[a] 95% confidence interval for parameters shown in parentheses.

The advantage of Model 1 is that its parameters have intuitive meaning. The estimated values for X_o of 0.1212 gDW/L and for X_m of 1.442 gDW/L are consistent with expectations for initial and maximum dry cell weight concentration, respectively, based on visual inspection of Figure 5.2. The estimated value for the lag phase duration, t_{lag}, is 6.06 hrs but this has a relatively wide 95% confidence interval with a margin equal to 55% of the estimated value. The estimated value for the maximum specific growth rate, μ_m, is 0.3867 hr^{-1}. This parameter also has a relatively wide 95% confidence interval with a margin equal to 65% of the estimated value. Thus, the parameters of lag phase duration and maximum specific growth rate are not as tightly determined.

In contrast to Model 1, the parameters estimated for the generalized logistic equation do not have an intuitive basis. For example, the expected values for Q' and ν are not obvious. Another challenge in the use of this model for our experimental data on dry cell weight concentration are the relatively wide 95% confidence intervals for all of the parameters except for the upper asymptote, K. The most egregious of these is the eight orders of magnitude covered in the 95% confidence interval for Q'. Therefore, although this model fits the data extremely well ($R^2 = 0.997$), the certainty of the estimated parameter values is low.

The profile of specific growth rate (Equation 5.12) derived from the two models differs greatly (see Figure 5.3). For the first model form, the specific growth rate is equal to 0 throughout the lag phase ($t \leq 6.06$ hr) and starts the growth phase at its highest value of 0.35 hr^{-1} which tapers off as time progresses. This discontinuity is inherent to the modeling choice of using piece-wise defined functions to define a lag and growth phase. The concept of a lag phase duration is a modeling construct in that biologically the "transition from the 'no-growth phase' to the 'exponential' growth regime is continuous rather than abrupt [and] the 'lag time' has no

clearly or uniquely defined duration" (Peleg and Corradini 2011). In the second model form, the specific growth rate starts at 0.08 hr^{-1} which increases to a maximum value of 0.21 hr^{-1} at 12.1 hr and subsequently continues to decrease after this point. The BY4742 wild-type strain has been found to have a doubling time of 2.5 - 3 hr in I+ synthetic media which corresponds to a specific growth rate of 0.23 - 0.28 hr^{-1} (Gaspar, Hofbauer et al. 2011). For two different synthetic media formulations similar to that used in this study, a specific growth of 0.40 and 0.47 hr^{-1} was observed for BY4742 during exponential phase (Hanscho, Ruckerbauer et al. 2012). These two media differed from this study's recipe by the supplementation levels of amino acids and inositol (55 µM in their study versus 75 µM in our study). Based on comparison with these reported literature values, we opted to use the first model form for subsequent calculations because of its higher specific growth rates during the period of fast growth.

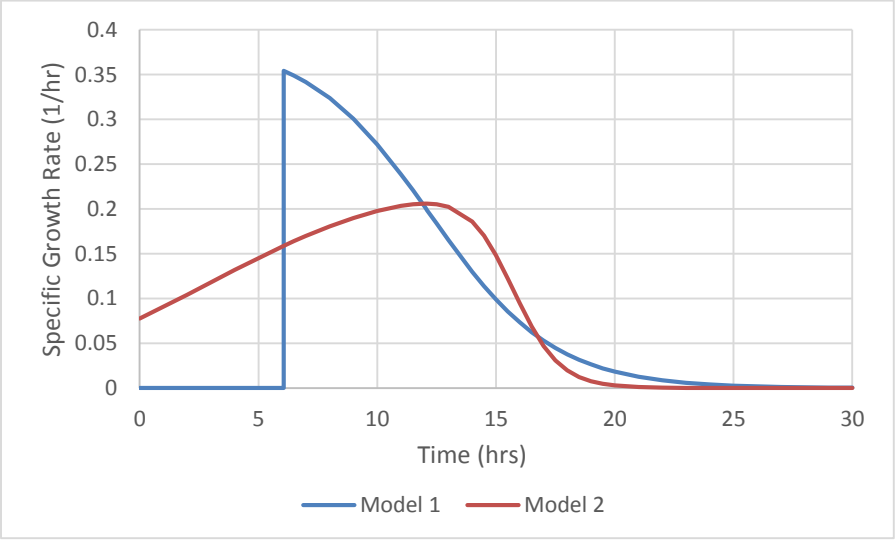


Figure 5.3: Specific growth rate of wild-type *S. cerevisiae* in I+ synthetic media over time as calculated by Model 1 (piece-wise function consisting of lag phase and logistic growth) and by Model 2 (generalized logistic equation). Specific growth rate calculated as dX/dt/X.

5.4.1.2 Specific glucose consumption rate

Experimental data on glucose concentration over time is shown in Figure 5.4. Over 96% of the starting glucose concentration is consumed at the measured time point of 28.5 hrs. Comparison to published data for BY4742 in glucose synthetic media shows variability in the rate of glucose consumption with Cao, Yue et al. (2011) observing glucose exhaustion at 16 hrs and with Ng, Jung et al. (2012) observing glucose exhaustion at 40 hrs. This discrepancy could be explained by the differences in factors such as cultivation conditions and formulation of the glucose synthetic media. It has been found that some standard recipes used in literature can be further supplemented with components such as inositol or amino acids to ensure optimal growth of BY4742 (Hanscho, Ruckerbauer et al. 2012).

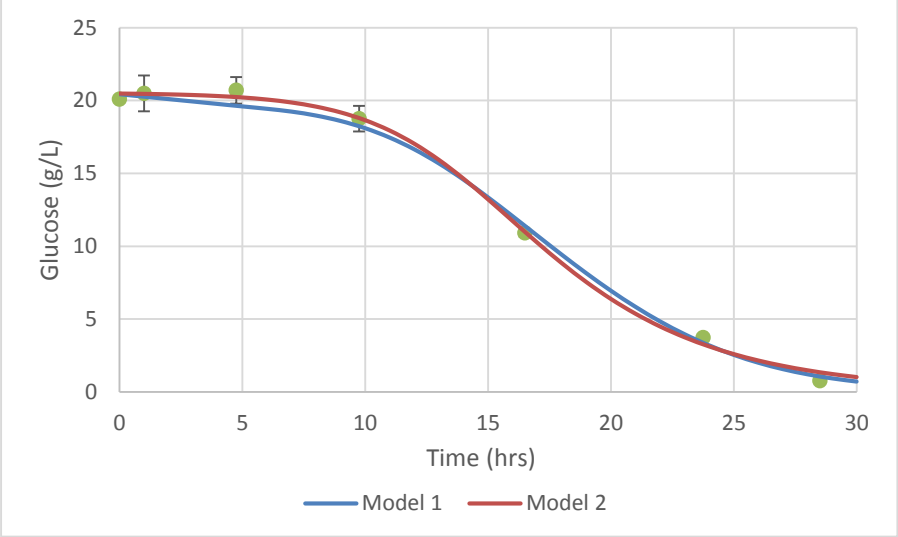


Figure 5.4: Glucose concentration for batch growth of wild-type *S. cerevisiae* in I+ synthetic media as predicted by Model 1 (Monod expression for specific glucose consumption rate) and by Model 2 (generalized logistic equation). Data points are the mean of three reactors and error bars show the standard deviation of these three reactors.

The estimated parameters, confidence regions, and R^2 value for models of experimental data on glucose concentration over time are listed in Table 5.2 and the plotted fits are shown in Figure 5.4.

Table 5.2: Curve fit of glucose concentration data[a]

Model 1: Monod expression (Equation 5.6)	$v_{g,max}$ (g/gDW/hr)	K_g (g/L)	R^2		
	-2.08 (-3.729, -0.432)	10.20 (-6.444, 26.852)	0.995		
Model 2: Generalized logistic equation (Equation 5.8)	**K** (g/L)	**B** (1/hr)	**Q'**	**ν**	R^2
	20.54 (19.28, 21.81)	-0.381 (-0.724, -0.037)	0.005 (-0.012, 0.021)	2.022 (-0.769, 4.813)	0.998

[a] 95% confidence interval for parameters shown in parentheses.

Using a Monod expression for specific glucose consumption rate (Equation 5.6), the values of -2.08 g/gDW/hr for $v_{g,max}$ and 10.20 g/L for K_g were determined to best fit the experimental data on glucose concentration ($R^2 = 0.995$). The 95% confidence interval for both these parameters are relatively wide with a margin equal to 79% of the estimated value for $v_{g,max}$ and 163% for K_g, which indicates the uncertainty in estimating the precise parameter values. Nevertheless, these values are comparable to the estimated parameters determined for the low-affinity hexose transporter Hxt3p ($v_{g,max}$ = 2.70 ± 0.54 g/gDW/hr, K_g = 11.71 ± 2.70 g/L) (Reifenberger, Boles et al. 1997) and estimated parameters for the enological diploid strain V8-6 grown anaerobically in synthetic grape must ($v_{g,max}$ = 2.00 g/gDW/hr, K_g = 8.11 g/L) (Salmon, Vincent et al. 1993).

In contrast to the Monod expression, the parameters estimated for the generalized logistic equation (Equation 5.8) do not correspond to conventional kinetic properties. Unlike the use of the generalized logistic equation for modeling the experimental data on cell dry weight concentration over time, the 95% confidence intervals for this equation applied to glucose concentration over time is much narrower.

The specific glucose consumption rate determined through the Monod expression (Equation 5.6) exhibits the largest magnitude of 1.39 g/gDW/hr at t = 0 hr and decreases in magnitude as the glucose concentration decreases (Figure 5.5). In contrast, the specific glucose consumption rate determined through the time derivative of the generalized logistic equation for glucose concentration (Equation 5.8) divided by the dry cell weight concentration (Equation 5.3) starts at a magnitude of 0.15 g/gDW/hr at t = 0 hr that ramps up to 1.42 g/gDW/hr at the start of logistic growth (Figure 5.5). The specific glucose consumption rate slightly increases to a maximum magnitude of 1.46 g/gDW/hr at t = 12 hr and subsequently decreases after that time.

The difference in the behavior of the two models is evident in the more rapid consumption of glucose initially predicted by the Monod expression which is later surpassed due to the larger consumption rate predicted by the logistic equation (Figure 5.4). Based on how the model predictions corresponded with experimental data for the early stages of fermentation, the generalized logistic equation was favored for further use.

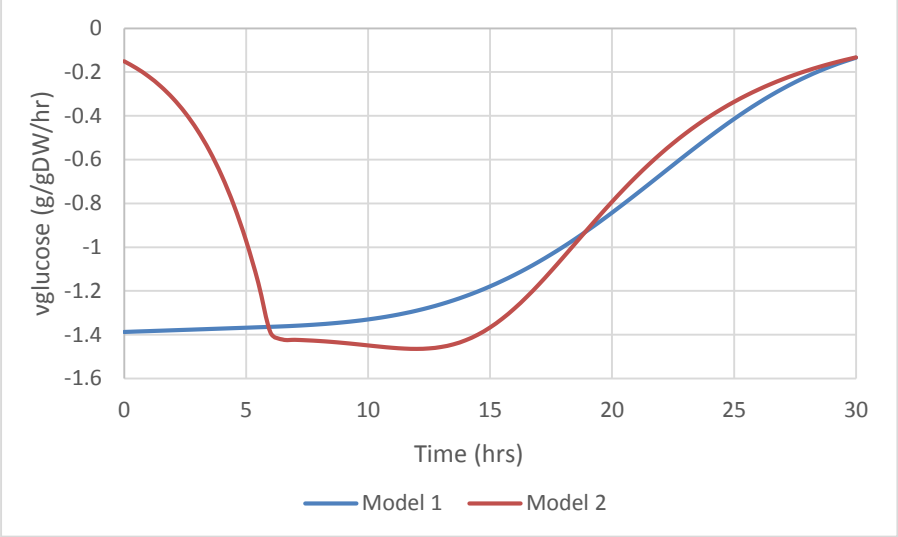


Figure 5.5: Specific glucose consumption rate of wild-type *S. cerevisiae* in I+ synthetic media over time as calculated by Model 1 (Monod expression for specific glucose consumption rate) and by Model 2 (generalized logistic equation). The dry cell weight used in calculating specific glucose consumption rate for Model 2 was determined through a piece-wise function consisting of lag phase and logistic growth (Equation 5.3).

5.4.1.3 Specific product formation rate

Experimental data on ethanol, glycerol, and acetate concentration over time is shown in Figure 5.6. At 28.5 hr, at which point most of the glucose has been consumed, concentrations of 6.68 ± 0.30, 1.12 ± 0.07, 0.59 ± 0.04 g/L for ethanol, glycerol, and acetate, respectively, are observed. Considering other reported data for growth of BY4742 in glucose synthetic media, Cao, Yue et al. (2011) observed final concentrations of 7.19 ± 0.31, 2.01 ± 0.01, and 0.14 ± 0.01 g/L for ethanol, glycerol, and acetate, respectively, while Ng, Jung et al. (2012) observed final

concentrations of 8.34, 0.64, and 0.34 g/L for ethanol, glycerol, and acetate, respectively. Thus, our experimental data for final product concentrations seems reasonable.

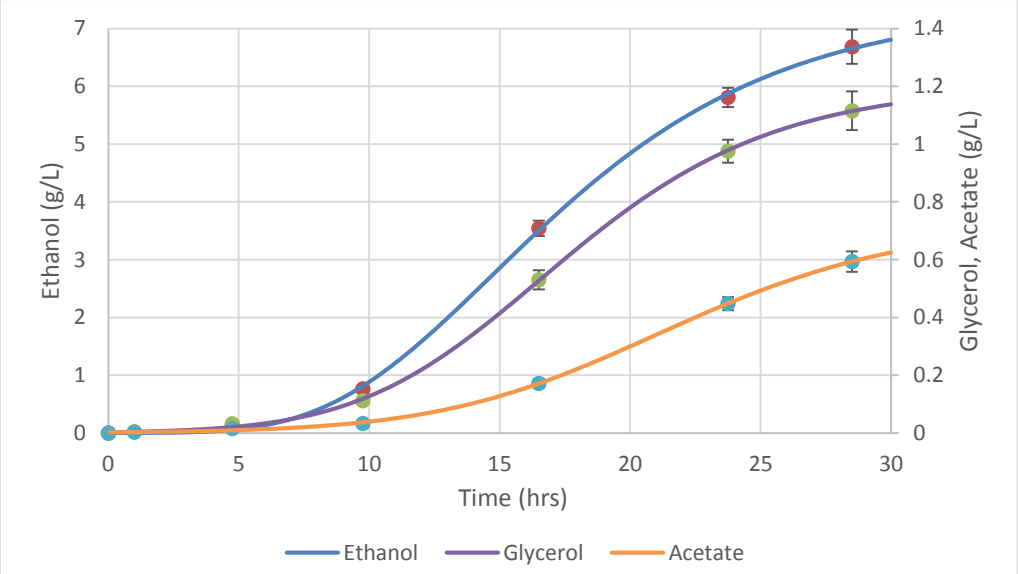


Figure 5.6: Ethanol, glycerol, and acetate concentration for batch growth of wild-type *S. cerevisiae* in I+ synthetic media as predicted by generalized logistic equations. Data points are the mean of three reactors and error bars show the standard deviation of these three reactors.

For fitting ethanol, glycerol, and acetate concentration, the same functional form of the generalized logistic equation used to model glucose concentration was assumed. The results of the curve fitting are summarized in Table 5.3 and Figure 5.6. All of the models fit the data well ($R^2 > 0.999$).

Table 5.3: Curve fit of ethanol, glycerol, and acetate concentration data using generalized logistic equation[a]

	K (g/L)	**B** (1/hr)	**Q'**	**ν**	$\mathbf{R^2}$
Ethanol	7.41 (6.353, 8.468)	0.161 (0.064, 0.259)	0.074 (-8.177, 8.325)	0.0071 (-0.748, 0.761)	0.9997
Glycerol	1.205 (1.116, 1.294)	0.211 (0.14, 0.282)	15.58 (-21, 52.17)	0.472 (-0.026, 0.97)	0.9999
Acetate	0.7292 (0.623, 0.835)	0.201 (0.115, 0.287)	46.05 (-80.5, 172.6)	0.6777 (0.051, 1.305)	0.9999

[a] 95% confidence interval for parameters shown in parentheses.

Like the previous application of the generalized logistic equation in this study, the estimate for the upper asymptote, K, is able to be tightly determined while the estimates for Q' and ν have notably wide 95% confidence intervals. All three products have similar estimates for the rate constant parameter B.

The profile of specific production rates over time as determined through the time derivative of the generalized logistic equation for product concentration (Equation 5.9 to Equation 5.11) divided by the dry cell weight concentration (Equation 5.3) is shown in Figure 5.7. The peak specific production rate of glycerol and acetate is attained at the start of logistic growth at 6.1 hr with a value of 0.10 g glycerol/gDW/hr and 0.03 g acetate/gDW/hr, respectively. The peak specific production rate of ethanol is attained at a delayed time of 7.8 hr with a value of 0.79 g ethanol/gDW/hr. The specific production rate for ethanol and glycerol continues to decline after this peak. The specific production rate for acetate, however, stays within 10% of this value from t = 6.1 to 7.3 hr and from t = 18.6 to 22.0 hr.

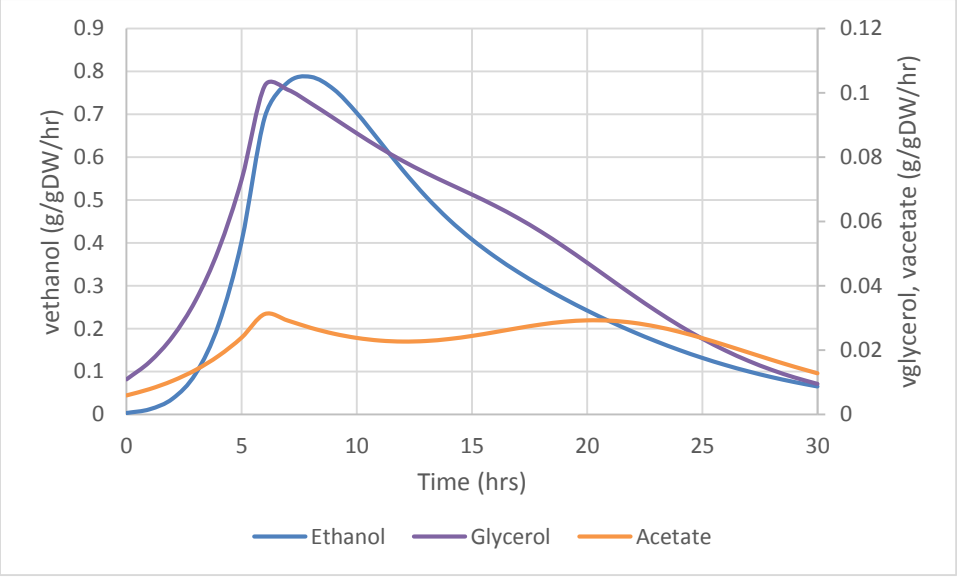


Figure 5.7: Specific ethanol, glycerol, and acetate production rate of wild-type *S. cerevisiae* in I+ synthetic media over time as calculated by the time derivative of the generalized logistic equation for product concentration (Equation 5.9 to Equation 5.11) divided by the dry cell weight concentration (Equation 5.3).

5.4.2 Constraining the Yeast Genome-Scale Metabolic Model with Experimentally Derived Fluxes

The experimentally-derived specific rates for growth, glucose consumption, and production of ethanol, glycerol, and acetate were used as exact constraints to fluxes within the Yeast v7.5 model. Imposing all five constraints on the model yielded infeasible solutions for different regions of the fermentation time-course as highlighted in Figure 5.8. The term "infeasibility" refers to the inability of the linear optimization algorithm to find a solution to the steady-state mass balance given the flux constraints (Equation 5.17). For time points that have infeasible specific rates, the specific consumption rate of glucose was determined to not be sufficient to drive the specific production rate of biomass, ethanol, glycerol, and acetate. For instance, the specific ethanol production rate at 8 hr is infeasible since it is more than twice the specific rate of glucose consumption and the highest molar yield of ethanol from glucose is 2 based on the overall equation for ethanol fermentation $C_6H_{12}O_6 \rightarrow 2\ C_2H_5OH + 2\ CO_2$. The region of infeasible experimentally-derived specific rates goes from 5.4 to 12.3 hr and from 26.7 hr and onwards, which encompasses 30.5% of the measured fermentation time.

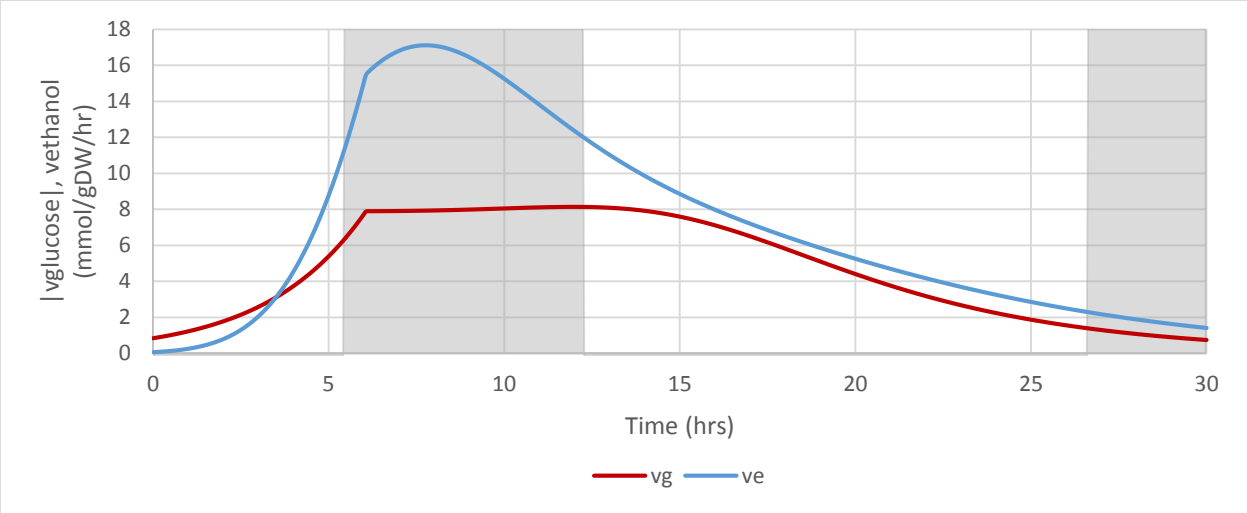


Figure 5.8: Profile of the infeasibility of the experimentally-derived specific rates used as constraints for modeling batch growth of wild-type *S. cerevisiae* in I+ synthetic media. The highlighted area indicates that the combination of specific rates at that time point cannot occur. The specific rates were calculated from curve fits using Equation 5.3 and 5.8 to 5.11. For sake of visual clarity, only the specific ethanol production rate and the absolute value of the specific glucose consumption rate are shown; the assessment of infeasibility was based on use of all of the experimentally-derived specific rates.

The presence of infeasible experimentally-derived specific rates suggests errors in modeling the data and/or measurement of concentrations. The feasibility of the specific rates calculated using all combinations of the tested dry cell weight (Equation 5.3 and Equation 5.5) and glucose (Equation 5.6 and Equation 5.8) models were tested but still did not resolve the issue of infeasible specific rates at certain time points (data not shown). The possibility of errors in the measurements of concentrations was examined through calculation of product yields and comparison with published data for the BY4742 strain (Table 5.4). The yields calculated using data at 16.5 and 23.75 hr is comparable to the yields calculated by other sources. However, the yields for dry cell weight and ethanol calculated using data at 9.75 hr greatly exceeds the yields calculated by other sources. Also, the calculated yield for ethanol at this time point exceeds the theoretical limit of 0.511 g ethanol/g glucose.

Table 5.4: Product yields for *S. cerevisiae* BY4742 in glucose synthetic media

	Yield (g/g glucose)[a]				
Time Used for Calculation (hr)	**Dry Cell Weight**	**Ethanol**	**Glycerol**	**Acetate**	**Source**
9.75	0.254	0.567	0.084	0.025	This study
16.5	0.128	0.386	0.058	0.019	This study
23.75	0.079	0.355	0.060	0.027	This study
16	Not specified	0.399	0.111	0.008	(Cao, Yue et al. 2011)
40	0.067	0.417	0.032	0.017	(Ng, Jung et al. 2012)
Exponential phase	0.12	0.385	0.031	0.021	(Hanscho, Ruckerbauer et al. 2012)

[a]Yield is calculated as g product produced/g glucose consumed using experimental data from the start of fermentation to the specified time.

Based on the dubiously large yields calculated at 9.75 hr, the prospect that the glucose consumption rate could have been underestimated was explored. For the time points with infeasible specific rates, the minimum glucose consumption rates needed to meet the constrained specific production rates were determined (Figure 5.9). The period from 6.1 to 10.8 hr required a 20-54% increase in the constrained specific glucose consumption rate in order to meet the set specific production rate of dry cell weight, ethanol, glycerol, and acetate.

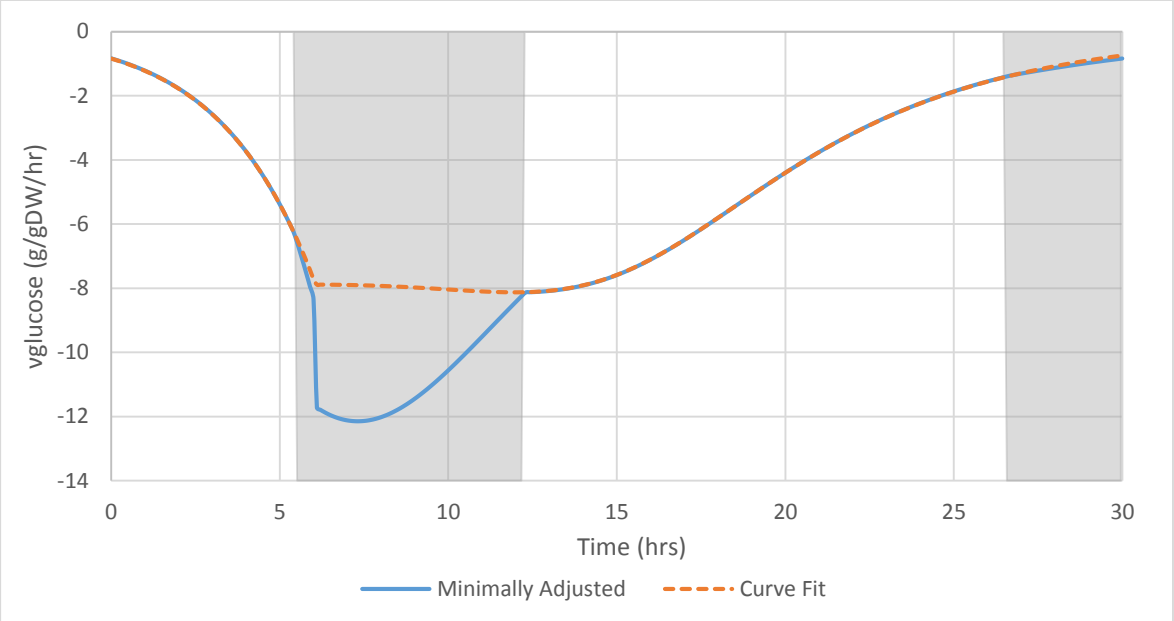


Figure 5.9: Minimum adjustment of the experimentally-derived specific glucose consumption rate needed to satisfy the constrained specific production rates of dry cell weight, ethanol, glycerol, and acetate. The highlighted area indicates that the combination of all experimentally-derived specific rates at that time point cannot occur unless constraints are adjusted or relaxed.

Adjustment of the assumed specific glucose consumption rate to the minimally required level would alter the yields at 9.75 hr to values more comparable to yields reported in other sources (Table 5.4, Table 5.5). To understand how the concentration profile from the experimental data and the flux constraints calculated by the model of the experimental data and the minimally adjusted model of the experimental data influenced the outcome of the linear programming solution, a carbon balance at 9.75 hr was calculated (Table 5.5). The carbon balance for the experimental data and model of the experimental data both have carbon balances greater than 1, which is impossible. Thus, it is understandable why the set of specific rates calculated at 9.75 hr using the model of the experimental data would yield an infeasible solution in the flux balance analysis framework. One source of difference between the yield from the experimental data and from the model of the experimental data is that the yield determined from the experimental data is based solely on two time points while the yield calculated from the

model is also influenced by data at other time points, which were used to define the parameters of the model. In contrast to the experimental data and the model of the experimental data, the minimally adjusted model of the experimental data has a carbon balance of 1, which would be expected for a system with mass-balance enforced.

Table 5.5: Carbon balance results for 9.75 hr using different data sources

	Products (g/g glucose)*					
Source	**Ethanol**	**Biomass**	**Glycerol**	**Acetate**	**CO_2†**	**Carbon Balance‡**
Experimental Data	0.567	0.254	0.084	0.025	0.560	1.55
Model of Experimental Data	0.497	0.194	0.061	0.017	0.487	1.30
Minimally Adjusted Model of Experimental Data	0.369	0.144	0.046	0.012	0.409	1.00

* For experimental data, yield is calculated as g product produced/g glucose consumed using concentration data at t = 0 and t = 9.75 hr. For the models, yield is calculated as specific production rate/specific glucose consumption rate.
† Since CO_2 was not experimentally measured, the CO_2 yield for the experimental data and model of experimental data is based on the assumption that 1 mole of CO_2 is produced for every mole of ethanol and for every mole of acetate. The CO_2 yield for the minimally adjusted model of experimental data is calculated from the flux balance analysis solution: -CO_2 exchange flux/glucose exchange flux.
‡ Assumes an average formula of $CH_{1.596}O_{0.396}N_{0.216}S_{0.0024}P_{0.017}$ for *S. cerevisiae* biomass (Villadsen, Nielsen et al. 2011).

The updated profile of concentrations simulated by dynamic flux balance analysis using the minimally adjusted specific glucose consumption rates and the experimentally-derived specific production rates is shown in Figure 5.10. Despite the large adjustments needed for the specific glucose consumption rate at certain time points, the difference in glucose concentration predicted by modeling of the data and by dynamic flux balance analysis using the adjusted specific glucose consumption rate is at most 0.75 g/L throughout the entire analyzed period. This maximum deviation occurs at 12 hr and represents a difference of -4.5% from the original glucose concentration predicted at 12 hr.

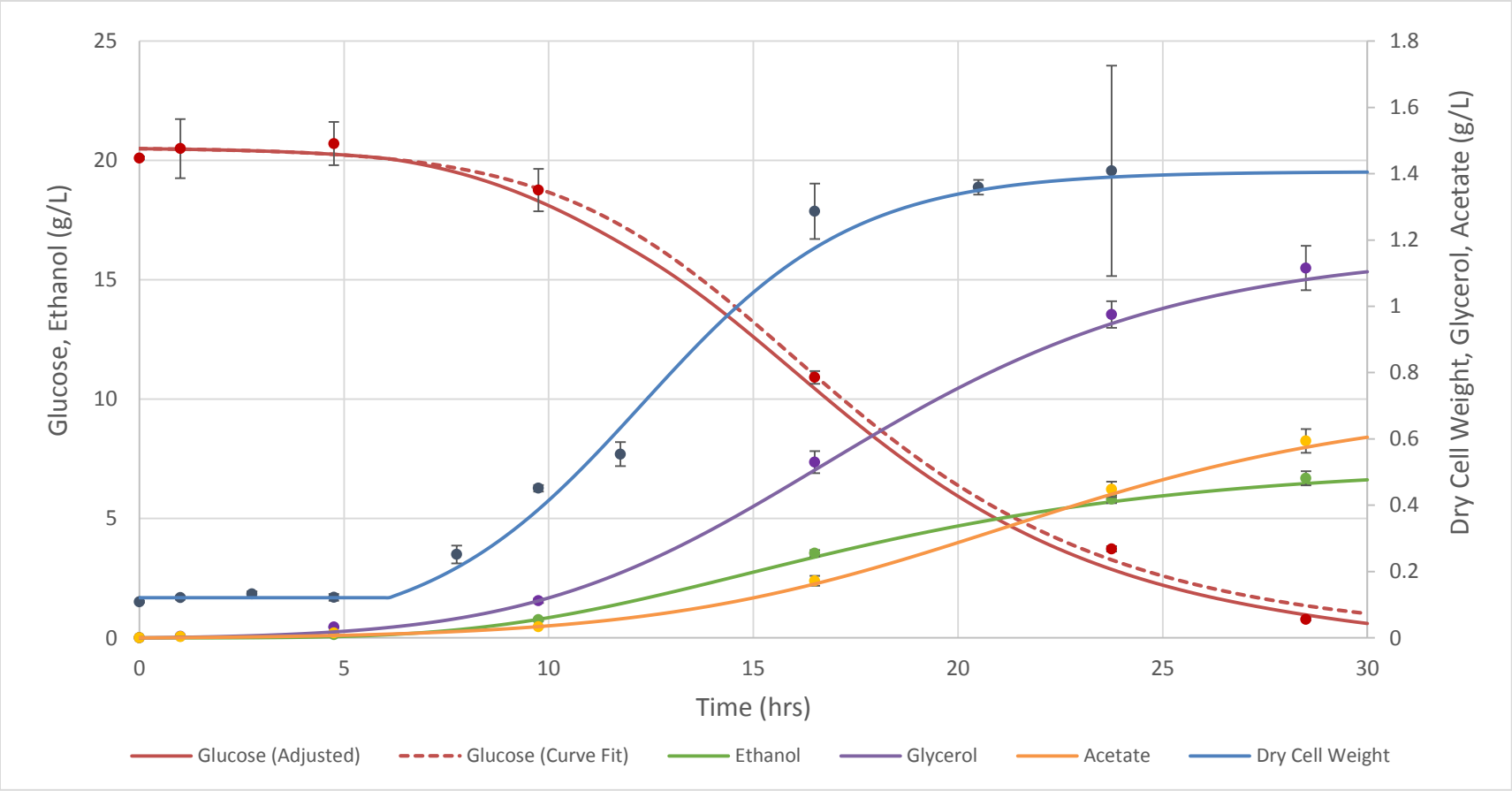


Figure 5.10: Simulation of batch growth of wild-type *S. cerevisiae* in I+ synthetic media using dynamic flux balance analysis constrained by experimentally-derived specific production rates and adjusted specific glucose consumption rates. Time step used in dynamic flux balance analysis was 0.1 hr. Data points are the mean of three reactors and error bars show the standard deviation of these three reactors.

5.5 Conclusions

The profile of calculated specific rates for batch yeast cultivation was found to be highly sensitive to the choice of function used to model the concentration data. Since the R^2 value, a statistical measure of goodness of fit, was greater than 0.99 for all the tested models, other metrics were considered for evaluation of the different models. Thus, the decision to choose one model over the other was based on comparison with specific rates reported by other sources and personal opinion on the underlying shape of the concentration profiles. The confidence in the appropriateness of the chosen model could be strengthened by an increased number of data points. Use of an appropriate model allows for estimation of concentrations between measured points.

Use of all the experimentally-derived specific rates for growth, glucose consumption, and product formation as exact flux constraints in the Yeast v7.5 model yielded infeasible solutions for certain time periods throughout the course of fermentation. This infeasibility arose because the set of specific rates at those times lead to violations of mass-balance in the flux balance analysis framework. Lindfors, Jouhten et al. (2014) have also reported that the use of specific rates determined from chemostat cultivations as exact flux constraints resulted in an infeasible solution. They resolved this infeasibility by relaxing the flux constraints through expanded lower and upper bounds. Since flux balance analysis requires mass-balance to always be satisfied, it is recommended that the carbon balance of the experimental data be checked to insure that it does not exceed 1. In cases where some of the time points have carbon balances greater than 1, the use of models to first capture the behavior of the concentration data could be helpful to smooth out noise and measurement errors present in the data. Otherwise, the use of lower and upper bounds to the specific rates would be needed to avoid over-constraining the mass-balance with

incompatible specific rates and to also represent the uncertainty in determining accurate specific rates.

REFERENCES

Cao, H., M. Yue, S. Li, X. Bai, X. Zhao and Y. Du (2011). "The impact of MIG1 and/or MIG2 disruption on aerobic metabolism of succinate dehydrogenase negative Saccharomyces cerevisiae." Applied Microbiology and Biotechnology **89**(3): 733-738.

Colombié, S., C. Nazaret, C. Bénard, B. Biais, V. Mengin, M. Solé, L. Fouillen, M. Dieuaide-Noubhani, J.-P. Mazat, B. Beauvoit and Y. Gibon (2015). "Modelling central metabolic fluxes by constraint-based optimization reveals metabolic reprogramming of developing Solanum lycopersicum (tomato) fruit." The Plant Journal **81**(1): 24-39.

Gaspar, M. L., H. F. Hofbauer, S. D. Kohlwein and S. A. Henry (2011). "Coordination of Storage Lipid Synthesis and Membrane Biogenesis." Journal of Biological Chemistry **286**(3): 1696-1708.

Hanscho, M., D. E. Ruckerbauer, N. Chauhan, H. F. Hofbauer, S. Krahulec, B. Nidetzky, S. D. Kohlwein, J. Zanghellini and K. Natter (2012). "Nutritional requirements of the BY series of Saccharomyces cerevisiae strains for optimum growth." FEMS Yeast Research **12**(7): 796-808.

Hjersted, J. L., M. A. Henson and R. Mahadevan (2007). "Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture." Biotechnology and Bioengineering **97**(5): 1190-1204.

Lindfors, E., P. Jouhten, M. Oja, E. Rintala, M. Oresic and M. Penttila (2014). "Integration of transcription and flux data reveals molecular paths associated with differences in oxygen-dependent phenotypes of Saccharomyces cerevisiae." BMC Systems Biology **8**(1): 16.

Mahadevan, R., J. S. Edwards and F. J. Doyle III (2002). "Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli." Biophysical Journal **83**(3): 1331-1340.

Min Lee, J., E. P. Gianchandani, J. A. Eddy and J. A. Papin (2008). "Dynamic Analysis of Integrated Signaling, Metabolic, and Regulatory Networks." PLoS Comput Biol **4**(5): e1000086.

Ng, C. Y., M.-y. Jung, J. Lee and M.-K. Oh (2012). "Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering." Microbial Cell Factories **11**(1): 68.

Peleg, M. and M. G. Corradini (2011). "Microbial Growth Curves: What the Models Tell Us and What They Cannot." Critical Reviews in Food Science and Nutrition **51**(10): 917-945.

Reifenberger, E., E. Boles and M. Ciriacy (1997). "Kinetic Characterization of Individual Hexose Transporters of Saccharomyces Cerevisiae and their Relation to the Triggering Mechanisms of Glucose Repression." European Journal of Biochemistry **245**(2): 324-333.

Salmon, J. M., O. Vincent, J. C. Mauricio, M. Bely and P. Barre (1993). "Sugar Transport Inhibition and Apparent Loss of Activity in Saccharomyces cerevisiae as a Major Limiting Factor of Enological Fermentations." American Journal of Enology and Viticulture **44**(1): 56-64.

Sánchez, B. J., J. R. Pérez-Correa and E. Agosin (2014). "Construction of robust dynamic genome-scale metabolic model structures of Saccharomyces cerevisiae through iterative re-parameterization." Metabolic Engineering **25**(0): 159-173.

Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke and B. O. Palsson (2011). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0." Nature Protocols **6**(9): 1290-1307.

Smallbone, K. and E. Simeonidis (2009). "Flux balance analysis: A geometric perspective." Journal of Theoretical Biology **258**(2): 311-315.

Varma, A. and B. O. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." Applied and Environmental Microbiology **60**(10): 3724-3731.

Villadsen, J., J. Nielsen and G. Lidén (2011). Bioreaction engineering principles, Springer Science & Business Media.

Zanghellini, J., K. Natter, C. Jungreuthmayer, A. Thalhammer, C. F. Kurat, G. Gogg-Fassolter, S. D. Kohlwein and H.-H. von Grünberg (2008). "Quantitative modeling of triacylglycerol homeostasis in yeast – metabolic requirement for lipolysis to promote membrane lipid synthesis and cellular growth." FEBS Journal **275**(22): 5552-5563.

Chapter 6: Conclusions

This dissertation documents the intensive manual curation of the consensus model of yeast metabolism. This curation effort was motivated by our desire to use the model to suggest genetic engineering strategies to increase triglyceride productivity. Manual curation of the reconstruction of fatty acid, glycerophospholipid, and neutral glycerolipid metabolism was accomplished through comparison of information contained in the Yeast v6.0 model to that found in scientific literature. Three different MATLAB functions were developed to probe the information contained within the model with regards to a list of all the reactions within the model that a specified metabolite participates in; the assigned flux bounds, associated gene(s), and stoichiometry of a specified reaction; and a list of all the reactions within the model that are catalyzed by the product of a specified gene. By examining the information contained within the model on metabolites, reactions, and genes involved in fatty acid, glycerophospholipid, and neutral glycerolipid metabolism, inconsistencies with literature knowledge were identified. These inconsistencies included incorrect gene-reaction associations, improper definition of substrates/products in reactions, missing connections between metabolites, inappropriate assignments of reaction directionality, nonfunctional β-oxidation pathways, and missing reactions relevant to the synthesis and degradation of triglycerides. Changes to address these issues were implemented to generate Yeast v7.0.

The Yeast v7.0 model contains 15 additional genes of which 11 introduce new catalytic activity. This model captures a more complete representation of the reactions for synthesis and degradation of triglyceride. New reactions added in this version allow for the synthesis of the key intermediate phosphatidate from dihydroxyacetone phosphate, the conversion of diglyceride

generated from hydrolysis of triglyceride into phosphatidate, and the complete breakdown of triglyceride into glycerol and free fatty acids. This version of the model is now capable of simulating growth on fatty acids as the sole carbon source due to amendments of blocked reactions in β-oxidation.

Yeast v7.0 shows minimal change from v6.0 in terms of accuracy of gene essentiality predictions. However, attention to individual predictions of gene essentiality provided insight into the limitations of both models. One cause for false positive predictions, i.e., gene deletion is predicted to be viable despite *in vivo* lethality, was the lack of utility of metabolites within the model. An example of this are phosphoinositides, which are involved in many functions such as signaling, recruitment of proteins to specific membranes, regulation of cell wall maintenance/synthesis, and vesicle-mediated membrane trafficking (Strahl and Thorner 2007). Such processes are beyond the current scope of the Yeast model and, as such, the essential role that these metabolites play is not able to be captured by the model. The linking of metabolism to complex biological processes, such as those aforementioned, represents a major challenge for modeling efforts.

The Yeast v7.0 model represents one version in the long history of the Yeast consensus model, whose publication record first began in 2008 (Herrgard, Swainston et al. 2008). To understand the impact of the changes made throughout the evolution of the Yeast consensus model, I examined differences in the maximum pathway yield of various metabolites calculated by the iMM904 model (Mo, Palsson et al. 2009), a progenitor to the consensus models, and the three latest major version releases of the consensus model at that time: Yeast v5.01, v6.06, and v7.11 (http://yeast.sourceforge.net/). The metric of maximum pathway yield allows for comprehension of the inherent capacity limit of a cell's metabolic network to transform substrate

into the desired product. Thus, if metabolic models are to be used to guide metabolic engineering strategies, verification of the reasonableness of the maximum pathway yield calculated by the model is crucial. The comparison of maximum pathway yields calculated by the different models of *S. cerevisiae* metabolism illustrated how seemingly small differences in model assumptions and representations led to tangible changes in simulated results.

Each of the models had different ATP yields from glucose due to variations in the assumed proton stoichiometry for the electron transport chain and ATP synthase. In the case of Yeast v6.06, an unreasonably large ATP yield was calculated due to a simulated reaction cycling that allowed for free generation of a mitochondrial proton gradient to drive ATP synthesis. The differences in ATP yield between the models has an impact not only on the supply of ATP used directly within the pathway for synthesis of the metabolite of interest but also for production of NADPH. Both NADPH and its dephosphorylated form NADH play an important role in cellular redox. The differences in the models that affect the balance and demand for these two distinct forms include: cofactor specificity of the malic enzyme, reversibility of NAD+-dependent glutamate dehydrogenase, inclusion of NADP+-dependent glycerol dehydrogenase, and inclusion of NADP+ phosphatase.

The changes made in v7.0 of the consensus model made the calculated maximum pathway yield of different lipid species more reasonable. The dominant factor that led to this improvement was setting the reaction for acetyl-CoA hydrolase to be irreversible: acetyl-CoA + $H_2O \rightarrow$ acetate + coenzyme A + H^+. This distinction is important because otherwise this reaction could serve as a route to produce acetyl-CoA from acetate without the input of ATP needed in the acetyl-CoA synthetase reaction. This correction reduces the calculated maximum pathway

yield for lipids by 13-15% due to the large amount of ATP needed to produce the acetyl-CoA units used for fatty acid synthesis.

In addition to the examination of how differences in the reconstruction of the metabolic network affect simulation results, I explored how experimental data on concentration of biomass and extracellular metabolites over time for batch cultivation can be used to further constrain simulated metabolic behavior. The intermediaries between this data and the metabolic model were specific rates for growth, glucose consumption, and product formation that could be input into the metabolic model as flux constraints. These specific rates are defined by the time derivative of concentration normalized to the present biomass concentration. The calculation of specific rates was facilitated by modeling of experimental data on dry cell weight, glucose, ethanol, glycerol, and acetate concentration over time. These fitted models allowed for estimation of concentrations between measured points, simple calculation of derivatives, and capture of the trend over the entire measured range.

The profile of calculated specific rates for batch yeast cultivation was found to be highly sensitive to the choice of function used to model the concentration data. Despite slight differences in how the tested models described the concentration profile, all the models tested had $R^2 > 0.99$. Thus, the decision to choose one model over the other was based on other metrics: comparison with specific rates reported by other sources and personal opinion on the underlying shape of the concentration profiles.

Inputting all of the experimentally-derived specific rates for growth, glucose consumption, and product formation as exact flux constraints in the Yeast v7.5 model yielded infeasible solutions for certain time periods throughout the course of fermentation. This infeasibility arose because the set of specific rates at those times lead to violations of mass-

balance in the flux balance analysis framework. Since flux balance analysis requires mass-balance to always be satisfied, it is recommended that the carbon balance of products to glucose determined from the data not exceed 1. In cases where some of the time points have carbon balances greater than 1, the use of models to first capture the behavior of the concentration data could be helpful to smooth out noise and measurement errors present in the data. For data sets with infeasible exact flux constraints, the use of lower and upper bounds to the fluxes could be used to avoid over-constraining the mass-balance with incompatible specific rates and to also represent the uncertainty in determining accurate specific rates.

REFERENCES

Herrgard, M. J., N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Büthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasie, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kurdar, M. Penttila, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen and D. B. Kell (2008). "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology." Nat Biotech **26**(10): 1155-1160.

Mo, M., B. Palsson and M. Herrgard (2009). "Connecting extracellular metabolomic measurements to intracellular flux states in yeast." BMC Systems Biology **3**(1): 37.

Strahl, T. and J. Thorner (2007). "Synthesis and function of membrane phosphoinositides in budding yeast, Saccharomyces cerevisiae." Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1771**(3): 353-404.

Chapter 7: Suggestions for Future Research

7.1 Modeling Changes in Lipid Content

The triacylglycerol (TAG) content of yeast cells is dynamic and varies throughout the growth phase. When stationary phase cells are transferred to fresh media, their TAG content is low during early exponential growth as TAG is quickly broken down to supply lipid precursors for rapid initial growth (Taylor and Parks 1979; Kurat, Natter et al. 2006; Zanghellini, Natter et al. 2008; Markgraf, Klemm et al. 2014). The pool of TAG begins to replenish as the cells enter further into exponential growth and reaches its highest level during stationary phase (Taylor and Parks 1979; Kurat, Natter et al. 2006; Czabany, Wagner et al. 2008; Zanghellini, Natter et al. 2008; Markgraf, Klemm et al. 2014).

This variation in TAG content over time is not able to be captured with models employing the balanced-growth assumption that treats the biomass composition as unchanging. This is the case in the usage of genome-scale metabolic models that have a singular definition of fixed proportions of biomass components. Therefore, it would be beneficial to develop a framework that can be incorporated into the genome-scale metabolic model to capture all of the changes in lipid content from the initial breakdown of TAG to its replenishment and accumulation in later periods.

7.2 Trade-Off between Accumulation of Storage Lipids and Production of Active Biomass

The maximum theoretical yield for conversion of glucose to TAG is 33% (w/w), assuming no production of any other cell material (Ratledge and Cohen 2008). Furthermore, Ratledge and Cohen (2008) estimate that the practical yield ceiling is around 22% since the

glucose must also be used to generate the rest of the cell. The use of optimization frameworks that formulate genetic engineering strategies based on determining mutants that maximize production of the desired metabolite while achieving its optimal growth is not as straight-forward in the application of increasing TAG content. If TAG is treated as a separate entity from the rest of the biomass, no TAG would be expected to be produced under the assumption of optimal growth of the TAG-free biomass since this would represent a diversion of cellular resources. Thus, attempts to find genetic engineering strategies should consider the trade-off between accumulation of TAG and production of active biomass.

7.3 Metrics and Tools to Evaluate Genome-Scale Metabolic Models

As revisions of current genome-scale metabolic models are being published, it would be useful to establish standardized test suites to gauge how changes affect the quality of the model. One of the metrics used to evaluate genome-scale metabolic models is the accuracy of predicting viable and inviable gene deletions. However, as discussed in Chapter 3, there can be instances where the model is able to correctly predict whether a gene is essential or not but does so for the wrong reason. Also, as evidenced in Chapter 4, the implemented assignment of (ir)reversibility to each reaction can affect simulation results by restricting or permitting the use of reactions in certain directions. Except for reactions undoubtedly occurring unidirectionally under physiological conditions, the direction of the remaining reactions is condition-specific. Thus, compilation of information on known reaction directions under specific conditions could provide more appropriate constraints than generic default reversibilities.

REFERENCES

Czabany, T., A. Wagner, D. Zweytick, K. Lohner, E. Leitner, E. Ingolic and G. n. Daum (2008). "Structural and Biochemical Properties of Lipid Particles from the Yeast Saccharomyces cerevisiae." Journal of Biological Chemistry **283**(25): 17065-17074.

Kurat, C. F., K. Natter, J. Petschnigg, H. Wolinski, K. Scheuringer, H. Scholz, R. Zimmermann, R. Leber, R. Zechner and S. D. Kohlwein (2006). "Obese Yeast: Triglyceride Lipolysis Is Functionally Conserved from Mammals to Yeast." Journal of Biological Chemistry **281**(1): 491-500.

Markgraf, Daniel F., Robin W. Klemm, M. Junker, Hans K. Hannibal-Bach, Christer S. Ejsing and Tom A. Rapoport (2014). "An ER Protein Functionally Couples Neutral Lipid Metabolism on Lipid Droplets to Membrane Lipid Synthesis in the ER." Cell Reports **6**(1): 44-55.

Ratledge, C. and Z. Cohen (2008). "Microbial and algal oils: Do they have a future for biodiesel or as commodity oils?" Lipid Technology **20**(7): 155-160.

Taylor, F. R. and L. W. Parks (1979). "Triacylglycerol metabolism in Saccharomyces cerevisiae relation to phospholipid synthesis." Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **575**(2): 204-214.

Zanghellini, J., K. Natter, C. Jungreuthmayer, A. Thalhammer, C. F. Kurat, G. Gogg-Fassolter, S. D. Kohlwein and H.-H. von Grünberg (2008). "Quantitative modeling of triacylglycerol homeostasis in yeast – metabolic requirement for lipolysis to promote membrane lipid synthesis and cellular growth." FEBS Journal **275**(22): 5552-5563.

Appendix A: Supplementary Information for Chapter 3

A.1 Differences in Representation of Fatty Acid Metabolism between Yeast v6.0 and iTO977

A.1.1 Fatty Acid Synthesis

In iTO977, cytoplasmic fatty acid synthesis (FAS) is represented as capable of generating C10:0, C12:0, C14:0, C16:0, and C18:0 free fatty acids and acyl-CoAs. This representation is misleading since the reaction intermediates are shuttled within the FAS enzyme complex and only the terminal acyl-CoA products of C16:0 and C18:0 (with minor amounts of C14:0) are released into the cytosol *in vivo*. In addition, the products of cytoplasmic FAS are acyl-CoAs, not free fatty acids. These two issues were present in the original Yeast v6.0 model as well and were addressed in the changes to the Yeast v6.0 model.

iTO977 depicts mitochondrial FAS in the same manner as cytoplasmic FAS (i.e. individual steps for extension of acetyl-[acp] all the way to stearoyl-[acp]). Mitochondrial FAS and cytoplasmic FAS are connected in iTO977 through transport of C4:0 to C18:0 acyl-ACP between the cytoplasm and mitochondria. This representation in iTO977 is similar to that found in the original Yeast v6.0 model. In contrast, the modified Yeast v6.0 model removes the production of mitochondrial FAS products longer than C8:0 acyl-ACP due to lack of experimental evidence and also removes the transport of acyl-ACP between the cytoplasm and mitochondria based on the observation that mitochondrial FAS is unable to compensate for cytoplasmic FAS in *fas1* or *fas2* mutants.

A.1.2 Fatty Acid Elongation

iTO977's representation of fatty acid elongation is in accord with the modified Yeast v6.0 model.

A.1.3 Fatty Acid Desaturation

Like the original Yeast v6.0 model, iTO977 is missing NADH/NAD+ in the reaction equations for the fatty acid desaturation. The modified Yeast v6.0 model includes these cofactors in the equations.

A.1.4 β-Oxidation

iTO977 did not have the issue of directionality of the *MDH3* reaction which regenerates NAD+ needed for continued β-oxidation. Nor did it have the issue of directionality of the mitochondrial carnitine acetyl-CoA transferase which is used for transfer of acetyl units into the mitochondria.

However, iTO977 has its own unique issues for β-oxidation. iTO9077's representation of β-oxidation has multiple blocked reactions since it had reactions for acyl-CoA oxidase in the cytoplasm while the reactions for the other enzymes of β-oxidation are in the peroxisome; this led to discontinuity between reactions. Also, the reactions for 2-enoyl-CoA hydratase should be associated with the gene *FOX2*, not *POX1*. iTO977's representation of β-oxidation of unsaturated fatty acids is not in accord with the mechanism presented in literature (DOI:10.1016/S0168-6445(03)00017-2) in terms of how the reactions for *ECI1*, *DCI1*, and *SPS19* enter in β-oxidation. In addition, iTO977 has reactions for β-oxidation of even chain lengths from 4 to 18 carbons, but not for very long chains. In contrast, the edited Yeast 6.0 model includes β-oxidation up to 26 carbons.

A.2 Differences in Representation of Glycerolipid and Glycerophospholipid Metabolism between Yeast v6.0 and iTO977

A.2.1 New Genes Added

Of the 15 additional genes added to the original Yeast v6.0 model, iTO977 had 4 of these genes (*AYR1*, *GCY1*, *IDP2*, *LRO1*) and was missing 11 of these genes.

A.2.2 Expansion of Species

iTO977 uses an approach more like that of iFF708, iLL672, iMM904, and iND750 in the sense that defined composites of specific species are used. iTO977 has model-specific reactions that pool various species together as shown below:

Pool_Acyl1 0.03 decanoyl-CoA [c] + 0.03 lauroyl-CoA [c] + 0.05 myristoyl-CoA [c] + 0.02 tetradecanoyl-9-ene-CoA [c] + 0.19 palmitoyl-CoA [c] + 0.48 hexadecanoyl-9-ene-CoA [c] + 0.08 octadecanoyl-CoA [c] + 0.12 octadecanoyl-9-ene-CoA [c] → acyl acids [c] + coenzyme A [c]

Pool_Acyl2 0.01 decanoyl-CoA [c] + 0.01 lauroyl-CoA [c] + 0.02 myristoyl-CoA [c] + 0.01 tetradecanoyl-9-ene-CoA [c] + 0.15 palmitoyl-CoA [c] + 0.60 hexadecanoyl-9-ene-CoA [c] + 0.04 octadecanoyl-CoA [c] + 0.16 octadecanoyl-9-ene-CoA [c] → acyl CoAs [c]

The successive acylation of glycerol-3-phosphate is then described in iTO977 as:

acyl CoAs [c] + sn-glycerol 3-phosphate [c] → acyl-sn-glycerol 3-phosphates [c] + coenzyme A [c]

acyl CoAs [c] + acyl-sn-glycerol 3-phosphates [c] → coenzyme A [c] + phosphatidate [c]

As a side-note, iTO977 also has this reaction utilizing acyl CoAs which does not make sense on a mass-balance basis.

acyl CoAs [c] + acyl-sn-glycerol 3-phosphates [c] → coenzyme A [c] + sn-glycerol 3-phosphate [c]

The hydrolysis of lipids such as triglycerides or steryl esters yields ‘acyl acids’. However, as is, iTO977 does not have reactions for the activation of these ‘acyl acids’ back to ‘acyl CoAs’ so that they may be re-utilized for acylation reactions or broken down in β-oxidation.

A.2.3 Compartmentalization

From the paper presenting iTO977 (Osterlund, Nookaew et al. 2013): “The compartments in iTO977 are Cytoplasm, Mitochondria, Peroxisome and Extracellular… Some of the reactions in the consensus network that takes place in other compartments, such as ER or nucleus, were included in the iTO977 model but localized to the cytoplasm, while some reactions in other compartments were out of scope of the model and discarded.”

REFERENCES

Osterlund, T., I. Nookaew, S. Bordel and J. Nielsen (2013). "Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling." BMC Systems Biology **7**(1): 36.