

Mapping and Analysis of the Complexity of Lipid Metabolism for Applications in Health and Sustainability

Présentée le 11 septembre 2020

à la Faculté des sciences de base
Laboratoire de biotechnologie computationnelle des systèmes
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Sofia TSOUKA

Acceptée sur proposition du jury

Prof. K. Sivula, président du jury
Prof. V. Hatzimanikatis, directeur de thèse
Dr I. Borodina, rapporteuse
Dr L. Wu, rapporteur
Prof. G. van der Goot, rapporteuse

*To Mom,
who shows me the meaning of strength and endurance every single day.*

Acknowledgements

First and foremost, I would like to express my gratitude to my thesis supervisor, Professor Vassily Hatzimanikatis, who gave me this wonderful opportunity and guided me through all of the difficulties and challenges of these past years. Vassily, thank you for sharing your valuable advice, scientific or otherwise, and for teaching me how to be a strong and independent researcher (and how to drink tea!).

I would also like to thank the members of my thesis committee, Dr. Irina Borodina, Professor Gisou van der Goot and Dr. Liang Wu, and my jury president, Professor Kevin Sivula, for their challenging questions, stimulating discussion and insightful comments on my work.

I am also grateful for the immense support I received from Christine Kupper in all administrative matters as well as in personal ones.

I would like to say a big thank you to the LCSB family, past and present: Firstly, to Misko, for our long discussions involving science, culture, language, food and anything you can imagine, thank you for being there when I needed help, advice or just someone to listen. Secondly, to my team, Maria, thank you for being my best friend since my(our) first days in the lab and helping me feel that I am not alone throughout all these years. I will always have your back no matter where we end up. To Tuure and Milenko, thank you for being my concert buddies, for the epic nights we've had together, and for the mental support you gave me in difficult times. To the rest of the LCSB-Epic members, Pierre, Robin, and Daniel, thank you for providing me with a daily dose of laughter and for all the beers! To Joana, Anush, and Noushin, thank you for our long talks and advice. I will always cherish our little brunches! I also want to thank Homa, my roommate, for the long talks, laughs, and hugs, Georgios F. for being himself (no further explanation needed), Zhaleh and Jasmin for always bringing a smile to my face, and Georgios S. for introducing me to the world of lipids! To Meriç., Tiziano, Yves, Stepan, Vikash, Daniel H., Anastasia, Omid, Evangelia, Asli, Remy, Thomas, Aarti, Cecilia, Maxime, and Yiannis, thank you for everything we've shared together, from scientific knowledge to absurd

Acknowledgements

discussions and from coffee breaks to shots, I am forever thankful to call you my colleagues and friends.

To my friends in Greece, you are too many to name, but thank you for supporting me even from afar. Eleni and Maria, you mean the world to me and I couldn't have done this without you. I think we made phone companies very happy! To all my friends in Lausanne, thank you for filling these years with wonderful experiences. Angeliki, thank you for some of the most fabulous months of my life! I am also very grateful to have met some amazing people all over Europe these past years: Chrisa, Fanny, and Sandy, thank you for making every show a unique experience!

My deepest gratitude goes to my family, my mother Evi and my father Andreas, who have given to me everything I would ever need in life, even if it meant they would be deprived of it. I know it hurts not having me around and I miss you every single day. To the rest of my family, Dimitris, Liana, Sotiris, Yiannis and Hara, thank you for the constant encouragement. I am also thankful to the Monnard family for welcoming me into their lives and embracing me as one of their own.

Finally, I want to thank Julien, my biggest supporter, my other half, my everything. Thank you for going through this journey with me, for picking me up whenever I fell down, for bringing me food and for loving me with all my flaws. I love you more and more each day.

Lausanne, February 20th, 2020

P.S.: Warrel, thank you for inspiring me to take on this big step in my life. You are forever...

Abstract

Metabolism is a network of biotransformations that sustain life in cells. Over the past decades, metabolism has been studied in multiple contexts, ranging from medicine to industrial manufacturing. In particular, lipid metabolism has been linked to various physiopathologies, and its study could lead to new diagnostics and treatments. Metabolic networks are extremely complex and a systems approach is essential in deciphering the different levels of cellular organization and regulation that an organism possesses. Genome-scale metabolic models (GEMs) encompass all the available information for an organism, though their large size and complexity introduce a lot of uncertainty in the accuracy of predictions. GEM reductions are usually done in an *ad hoc* manner to produce context-specific models, and cannot serve multiple studies.

The study of regulatory mechanisms requires the development of kinetic models that can accurately capture dynamic responses to perturbations. To this end, constraint-based models can be enhanced through the integration of kinetic information. Constructing consistent large-scale kinetic models is a challenging endeavor, since detailed knowledge about regulatory mechanisms and kinetic parameter values is scarce.

Examining the mechanisms of enzymatic and metabolic restructuring in states of mutation could increase our fundamental understanding of how diseases evolve, and facilitate phenotype mapping. Metabolic Control Analysis (MCA) has been a well-established tool for the prediction and evaluation of genetic modification strategies. However, it fails to account for the physiological limitations of an organism and can often lead to unrealistic predictions.

In this thesis, we developed computational models, tools, and methodologies to facilitate the study of lipids and their regulatory mechanisms. Firstly, we constructed and curated a metabolic model that focuses on lipid metabolism, through the integration of detailed lipid pathways into a GEM of *S. cerevisiae*. This model was then systematically reduced around these pathways to provide a more manageable model size for complex studies. We show that this model is as consistent and inclusive as other yeast GEMs, and can be used as a scaffold for integrating lipidomics data to improve predictions in studies of lipid-related biological functions. Secondly, we used this model as a basis to build a large-scale kinetic model. Enzymes in the lipid metabolism are typically promiscuous and multifunctional, giving rise to enzymatic coupling. To accurately capture these properties, we assigned suitable kinetic rate expressions to the reactions in the network. We generated populations of kinetic parameter sets through a sampling-based workflow and we demonstrate how the consideration of enzymatic coupling is essential for factual predictions. Thirdly, we developed a constraint-based formulation that utilizes MCA-based control coefficients for the consistent derivation of metabolic engineering strategies. We show how the parametrization and introduction of biological constraints enhances this formulation in comparison to the classical MCA approach, and we highlight its ability to generate alternative optimal strategies. Fourthly, we defined and introduced additional mathematical objectives and constraints to this formulation to enable the mapping of enzymes that are responsible for different phenotypes. Concluding, we discuss the contribution and potential applications of this thesis.

Keywords: metabolism, lipids, *S. cerevisiae*, pathway integration, model reduction, thermodynamics, kinetic models, enzyme promiscuity, metabolic engineering strategies, mutant identification

Résumé

Le métabolisme est le réseau de réactions biochimiques qui permet la vie cellulaire. Au cours des dernières décennies, le métabolisme a été étudié dans plusieurs domaines, allant de la médecine à la production chimique industrielle. En particulier, le métabolisme des lipides est lié à plusieurs types de physiopathologies, et son étude promet de nouvelles pistes pour le développement de diagnostics et traitements. Les réseaux métaboliques sont extrêmement complexes, et une approche systémique est essentielle pour déchiffrer les différents niveaux d'organisation cellulaire et de régulation que possède un organisme. Les modèles métaboliques à échelle du génome (GEMs) contiennent toute l'information génétique d'un organisme ; toutefois, leur grande taille et complexité sont la source d'incertitudes dans leur capacité prédictive. Les GEMs sont souvent réduits de façon *ad hoc*, afin de produire des modèles spécifiques à un contexte d'étude et ces modèles ne peuvent donc pas être utilisés de façon indépendante.

L'examen des mécanismes régulateurs requiert le développement de modèles cinétiques ayant la capacité de capturer des réponses dynamiques à des perturbations. A cet effet, les modèles à base de contraintes peuvent être améliorés grâce à l'intégration d'informations cinétiques. La construction consistante de modèles cinétiques de grande échelle reste cependant un défi, à cause du peu d'informations disponibles concernant les mécanismes de régulation et les valeurs des paramètres cinétiques.

L'étude des mécanismes de restructuration enzymatique et métabolique lors de la mutation promet d'étendre notre compréhension de l'évolution des maladies, et de faciliter la cartographie de phénotypes. L'analyse du Contrôle Métabolique (MCA) est

un outil bien établi pour la prédiction et l'évaluation de la réponse d'un organisme à des modifications génétiques. Toutefois, cet outil ne considère pas les limitations physiologiques d'un organisme, et peut rendre des prédictions irréalistes.

Dans cette thèse, nous avons développé des modèles, des outils et des méthodologies informatiques qui facilitent l'étude des lipides et de leurs mécanismes régulateurs. Tout d'abord, nous avons construit et édité un modèle métabolique du métabolisme des lipides, à travers l'intégration de voies de biosynthèse détaillées des lipides dans un GEM de *S. cerevisiae*. Ce modèle a ensuite été systématiquement réduit autour de ces voies afin de produire un modèle de taille plus adaptée à des études complexes. Nous montrons ensuite que ce modèle est aussi consistant et inclusif que d'autres GEMs de levures, et qu'il peut être utilisé comme échafaudage pour intégrer des données lipidiques et ainsi améliorer les prédictions dans les études des fonctions biologiques reliées aux lipides. De plus, nous avons utilisé ce modèle comme base pour la construction d'un modèle cinétique de grande échelle. Les enzymes du métabolisme des lipides sont typiquement peu spécifiques et multifonctionnelles, provoquant un couplage enzymatique. Pour capturer précisément ces propriétés, nous avons assigné à chaque réaction du réseau métabolique une expression cinétique appropriée. Nous avons ensuite généré des populations de paramètres cinétiques à travers une méthode d'échantillonnage aléatoire, et nous montrons que la prise en compte du couplage enzymatique est essentielle à des prédictions précises. Ensuite, nous avons développé une formulation à base de contraintes qui utilise les coefficients de contrôle obtenus par MCA, afin de concevoir des stratégies d'ingénierie métaboliques cohérentes. Nous montrons aussi que le paramétrage et l'introduction de contraintes biologiques améliorent la formulation en comparaison à l'approche MCA classique, et nous démontrons la capacité de notre modèle à générer des stratégies optimales alternatives. Enfin, nous avons défini et introduit des objectifs et contraintes supplémentaires dans cette formulation afin d'établir une correspondance causale entre enzymes et phénotypes. En conclusion, nous discutons de la contribution et des applications potentielles de cette thèse.

Mots-clés: métabolisme, lipides, *S. cerevisiae*, Intégration de voies métaboliques, réduction de modèles, thermodynamique, modèles cinétiques, enzymes non-spécifiques, stratégies d'ingénierie métabolique, identification de mutants

Table of Contents

ACKNOWLEDGEMENTS	I
ABSTRACT	III
RESUME	V
TABLE OF CONTENTS	VII
LIST OF TABLES	XI
LIST OF FIGURES	XIII
LIST OF ABBREVIATIONS	XV
CHAPTER 1 – INTRODUCTION	1
1.1 <i>Background</i>	1
1.2 <i>Aim and Scope</i>	6
1.2.1. Lipid metabolism	6
1.2.2. Yeast as a model organism.....	8
1.2.3. Kinetic modeling of lipids.....	8
1.2.4. Metabolic engineering strategies in sustainability and health	9
1.3 <i>Thesis Overview</i>	10
<i>Articles included in this thesis</i>	12
CHAPTER 2 – REDLIPS: A COMPREHENSIVE MECHANISTIC MODEL OF THE LIPID METABOLIC NETWORK OF YEAST	13
2.1 <i>Introduction</i>	13
2.2 <i>Materials and Methods</i>	16
2.2.1 Starting reaction network.....	16
2.2.2 Consistent reduction of models.....	16
2.2.3 Genome Scale Model (before integration)	18
2.2.4 Gathered Lipid Reactions Network (before integration)	18
2.2.5 Genome Scale Model (after integration).....	19

2.2.6 Lipidomics – biosynthetic fluxes.....	19
2.2.7 Thermodynamics	20
2.2.8 Fatty Acid chain lengths.....	22
2.2.9 Lipidomics – concentrations.....	23
2.2.10 Media	23
2.3 Results and Discussion.....	23
2.3.1 redGEM output model.....	23
2.3.2 Overview of the reactions and metabolites in each expansion step.....	24
2.3.3 Generated lumped reactions.....	28
2.3.4 Thermodynamics	31
2.3.5 Gene Essentiality Analysis and Comparison	33
2.3.6 Comparison with yeast GEMs.....	37
2.4 Conclusions.....	41
CHAPTER 3 – EXPLORING THE EFFECT OF ENZYMATIC COUPLING ON KINETIC MODELING AND METABOLIC CONTROL ACROSS THE LIPID PATHWAYS OF YEAST	43
3.1 Introduction	43
3.1.1 Stoichiometric Modeling of Metabolism	43
3.1.2 Kinetic Modeling of Metabolism.....	44
3.1.3 Lipid Regulation	45
3.1.4 Metabolic Control Analysis.....	46
3.1.5 Aim and Contribution	46
3.2 Materials and Methods.....	47
3.2.1 Kinetic modeling techniques and metabolic control.....	47
3.2.2 The effect of enzymatic coupling on Control Coefficients.....	49
3.2.3 The updated redLips network.....	49
3.2.4 Experimental data integration.....	50
3.2.5 Sampling the solution space.....	52
3.3 Results and Discussion.....	52
3.3.1 Bidirectional Reactions and Flux Directionality Profiles.....	52
3.3.2 Conserved moieties.....	54
3.3.3 Kinetic mechanisms – Apparent inhibition through competing substrates	55
3.3.4 Kinetic mechanisms – Rest of the network	59
3.3.5 The importance of accounting for the enzymatic coupling effect.....	61
3.3.6 Validation of the model’s predictive capabilities.....	66
3.4 Conclusions.....	70

CHAPTER 4 – RESPONSE BALANCE ANALYSIS: INVESTIGATING METABOLIC CONTROL IN A CONSTRAINT-BASED FORMULATION FOR THE DESIGN OF STRAIN ENGINEERING STRATEGIES	73
4.1 <i>Introduction</i>	73
4.2 <i>Materials and Methods</i>	75
4.2.1 Metabolic Control Analysis notions.....	75
4.2.2 Kinetic model description.....	76
4.3 <i>Results and Discussion</i>	77
4.3.1 Response Balance Analysis formulation.....	77
4.3.2 Navigating the degrees of freedom.....	79
4.3.2.1 Selecting a set of kinetic parameters.....	80
4.3.2.2 Selecting the objective function(s).....	81
4.3.2.3 Defining the bounds of variables.....	81
4.3.2.4 Defining the bounds of parameters.....	83
4.3.2.5 Fixing the number of parameter manipulations.....	84
4.3.3 Manipulation-limiting concentrations.....	85
4.3.4 Selecting top candidate strategies.....	88
4.4 <i>Conclusions</i>	91
CHAPTER 5 – INVERSE RESPONSE BALANCE ANALYSIS: A REVERSE ENGINEERING FORMULATION FOR THE MAPPING OF ENZYMES RESPONSIBLE FOR DIVERSE PHENOTYPES	93
5.1 <i>Introduction</i>	93
5.1.1 Evolutionary biology and regulatory responses.....	93
5.1.2 Mutations in health and disease.....	94
5.1.3 Formulations for the study of metabolic control.....	95
5.1.4 Aim and scope.....	95
5.2 <i>Materials and Methods</i>	96
5.2.1 Response Balance Analysis.....	96
5.2.2 Models used in this study.....	96
5.2.3 Integrating experimental data of mutant states.....	97
5.3 <i>Results and Discussion</i>	99
5.3.1 Part 1 – Reaching optimal growth states.....	99
5.3.1.1 Definition of the problem: evolution for optimal growth.....	99
5.3.1.2 Quadratic objective problem formulation.....	100
5.3.1.3 Choosing a kinetic parameter set.....	101
5.3.1.4 Reaching optimal growth – Number of enzymatic adjustments.....	102
5.3.1.5 Reaching optimal growth – Metabolic and genetic restructuring.....	103

5.3.1.6 Reaching other phenotypes.....	108
5.3.2 Part 2 – Mutant identification and mapping	110
5.3.2.1 Definition of the problem: identification of mutants	110
5.3.2.2 Experimental data for Δ lcb3 mutant strains.....	110
5.3.2.3 Selection of variable bounds and objective function.....	111
5.3.2.4 Mapping the mutant phenotype to enzymatic changes	111
5.4 Conclusions	114
CHAPTER 6 – CONCLUSIONS AND FUTURE PERSPECTIVES.....	115
APPENDIX A - KINETIC MECHANISM EXPRESSIONS.....	119
A.1 Michaelis-Menten kinetics	119
A.1.1 Uni-Uni	119
A.1.2 Irreversible (Michaelis-Menten like kinetics).....	120
A.2 Generalized Reversible Hill kinetics with Hill coefficient $h = 1$	120
A.2.1 Bi-Bi	120
A.2.2 Uni-Bi	120
A.2.3 Bi-Uni	121
A.3 Generalized Reversible Hill kinetics with Hill coefficient $h = 4$	121
A.3.1 PFK specific mechanism	121
A.4 Convenience kinetics	122
A.4.1 Bi-Ter	122
A.4.2 Ter-Bi	122
A.4.3 Stoichiometric coefficient larger than unity	123
A.4.4 Containing apparent inhibition terms.....	123
A.5 Mass-action kinetics	123
A.5.1 Diffusion of species or chemical reaction	124
Note: Case of participating compensating small molecules.....	124
BIBLIOGRAPHY	125
CURRICULUM VITAE.....	134

List of Tables

Table 2.1. Values for pH and ionic strength (in M) for each model compartment, and cross-membrane potentials (in mV) for each set of these compartments (where applicable) – opposite arrow direction will correspond to the same value with opposite sign. 21

Table 2.2. List of subsystems included in redLips, and the corresponding number of reactions that were added in each step of the reduction process. Total number of reactions per subsystem and the percentage coverage of the corresponding integrated GEM subsystem is also reported. Boldface denotes the lipid pathway subsystems. FGS: Final Graph Search, PP: Post-Processing. (*)The biomass reaction representing cell growth is not part of either the starting network or any expansion step. (**)The lumped reactions are not part of the expansion steps, and they are computed and added to the model after D3 and before FGS. 26

Table 2.3. Biomass building blocks for iMM904, the size of the subnetworks generated by lumpGEM, and the corresponding number of lumped reactions. (*)Produced by the core network. 32

Table 2.4. Detailed comparison between redLips and other yeast GEMs in terms of reactions, species, and cellular compartments. The curated numbers of lipid reactions and species are given in parentheses next to the non-curated numbers if these numbers differ. 40

Table 3.1. Promiscuous enzymes in the redLips network, identified by their standard gene name, along with the corresponding case, defined reaction groups and number of catalyzed reactions. 57

Table 3.2. Overall assigned kinetic mechanisms in the redLips network according to each enzymatic function and stoichiometry. 60

Table 4.1. Violating metabolite species for the case of three metabolite concentration violations, for a range of parameter manipulations. Each set of three species was unique for each number of parameter manipulations. 87

Table 4.2. Top ranked parameters based on their control over glucose uptake flux GLCptspp. Ranking was computed based on the mean values of 50'000 sets of Control Coefficients. 88

Table 4.3. Computed alternative strategies for the increase of glucose uptake. The number of manipulations is set to five, the manipulation magnitude bounds are set to 2-fold, the flux deviation bounds are set to 100-fold and the concentration deviation bounds are set to TFA+2-fold. These calculations were made using the reference kinetic parameter set. 90

Table 5.1. Computed enzymatic adjustment statistics for the 19 selected models: Number of models for which each enzyme is predicted to have altered activity, and the type of enzymatic activity regulation. Boldface denotes enzymes that are part of the predictions of the representative model. 107

List of Figures

Figure 2.1. The L-methionine minimal subnetwork (in blue). The purple box highlights the textbook methionine biosynthetic route starting from aspartate. The orange box highlights the sulfate assimilation pathway for methionine biosynthesis. Reactions in red are part of the core network and part of the biosynthetic routes. Reactions in green are part of the core network but not part of the biosynthetic routes and serve the mass balancing of the subnetwork. 29

Figure 2.2. The ergosterol minimal subnetwork (in blue). Reactions in red are part of the core network and part of the biosynthetic route. Reactions in green are part of the core network but not part of the biosynthetic route, serving instead as the mass balance for the subnetwork..... 30

Figure 2.3. The phosphatidylcholine (PC) minimal subnetwork (in blue). Reactions in red are part of the core network and part of the biosynthetic route. Reactions in green are part of the core network but not part of the biosynthetic route, serving instead as the mass balance for the subnetwork..... 31

Figure 2.4. (a) Venn diagram of the genes included in redLips and iMM904. (b) Gene essentiality analysis in redLips and comparison with experimental evidence. (c) Improvements that can be made to the predictions by enforcing lipid biosynthetic requirements. (d) Gene essentiality analysis in both redLips and iMM904 for the enzymes they have in common; and comparison with experimental evidence. (e) The nine genes that correspond to false positive predictions of redLips and explanations of the occurrence. (e) The five true positive predicted genes of redLips that iMM904 predicts falsely negative. The Matthews Correlation Coefficient (MCC)(Matthews 1975) is also reported for each case..... 36

Figure 3.1. Top 20 ranked Flux Control Coefficients corresponding to the reaction PIS1, for the cases of (a) no enzymatic coupling and without restructured Π matrix, (b) no enzymatic coupling with restructured Π matrix, (c) enzymatic coupling without restructured Π matrix, (d) enzymatic coupling with restructured Π matrix. The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively. Colors denote parameters of interest and are discussed in the text..... 63

Figure 3.2. Flux Control Coefficients corresponding to the reactions of the Kennedy pathway for the biosynthesis of (a) PE (in purple) and (b) PC (in green), with respect to the enzymes catalyzing the CDP-DAG biosynthetic pathway (in red). The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively. (c) The CDP-DAG and Kennedy biosynthetic pathways, with corresponding colors..... 68

Figure 3.3. (a) Top 7 ranked Concentration Control Coefficients (top) corresponding to the enzyme activity TAZ1 along with the lower ranked MLCL and CL species (bottom), and (b) zoom in on the bar corresponding to CL. The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively..... 69

List of Figures

Figure 4.1. Schematic representation of the RBA formulation. Diagram providing information about the various steps necessary for constructing an RBA model from kinetic models for studying system response to perturbations. 79

Figure 4.2. Effect of the metabolite concentration bounds on the achievable fold change in glucose uptake for various numbers of genetic manipulations. The flux bounds were maintained flexible at 100-fold in order to study the impact of the concentration bounds on flux predictions. The maximal allowed enzymatic perturbation was 2-fold in magnitude. The whiskers and the diamonds indicate the interquartile ranges and the means of 19 models considered, respectively. 82

Figure 4.3. Effect of the enzymatic perturbation magnitude bounds on the achievable fold change in glucose uptake for various numbers of genetic manipulations. The flux bounds were maintained flexible at 100-fold in order to study the impact of the enzymatic perturbation magnitude bounds on flux predictions. The concentration bounds were fixed to the thermodynamically feasible calculated values. The whiskers and the diamonds indicate the interquartile ranges and the means of 19 models considered, respectively. 84

Figure 4.4. Maximal attainable fold increase for glucose uptake with respect to the number of metabolite concentration violations, for a various numbers of parameter manipulations. The flux bounds were maintained flexible at 100-fold, the non-violated concentration bounds were fixed to the thermodynamically permissible plus physiologically feasible bounds (TFA+2-fold), and the parameter manipulation magnitude bounds were set to 5-fold. For all cases only the reference model was used. 86

Figure 5.1. Euclidean distance from target deviation vector, for various numbers of enzymatic manipulations across the 19 selected sets of kinetic parameters. For each number of manipulations, the red lines denote the median solution value, the edges of the blue boxes the upper and lower quantiles, the black whiskers extend to the full range of solution values excluding outliers and the red crosses denote the outliers. The bounds for flux, concentration and enzymatic activity deviations are all set to 100-fold. .103

Figure 5.2. Euclidean distance of each reaction from target deviation vector (optimal growth). The bounds for flux, concentration and enzymatic activity deviations are set to 100-fold, TFA&2-fold and 5-fold, respectively. 105

List of Abbreviations

BBB	Biomass Building Block
BDR	Bidirectional Reaction
CC	Control Coefficient
CCC	Concentration Control Coefficient
<i>E. coli</i>	<i>Escherichia coli</i>
FBA	Flux Balance Analysis
FCC	Flux Control Coefficient
FDP	Flux Directionality Profile
GEM	Genome-scale Metabolic Model
GLRN	Gathered Lipids Reaction Network
GPI	Glycosyl-Phosphatidylinositol
GPR	Gene-Protein-Reaction
iRBA	inverse Response Balance Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
MCA	Metabolic Control Analysis
MCC	Matthews Correlation Coefficient
MILP	Mixed-Integer Linear Programming
MIQP	Mixed-Integer Quadratic Programming
ODE	Ordinary Differential Equation
ORACLE	Optimization and Risk Analysis of Complex Living Entities
PCA	Principal Component Analysis
RBA	Response Balance Analysis
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SGD	Saccharomyces Genome Database
TFA	Thermodynamics-based Flux Analysis
TVA	Thermodynamic Variability Analysis
$\Delta_r G'^o$	Standard Gibbs Free Energy of Reaction

Chapter 1 – Introduction

1.1 Background

Metabolism is the sum of all life-sustaining biochemical transformations that occur in organisms. Through metabolism, fuel is generated and provided to perform cellular processes such as the synthesis of proteins, lipids, nucleic acids, and some carbohydrates. These sets of enzyme-catalyzed reactions allow cells to grow, sustain their structures, and respond to environmental changes. Metabolism and metabolic pathways have been the focus of numerous studies over the last decades. These studies focused on a wide variety of scientific areas from medicine to industrial manufacturing. While many of the underlying mechanisms have been revealed, others remain uncovered. For example, metabolic syndrome has been associated with cancer, obesity, inflammation, insulin resistance and fully developed diabetes, and elevated risk of cardiovascular disease (DeBerardinis and Thompson 2012, Despres and Lemieux 2006, Hotamisligil 2017). Moreover, metabolic synergy in the gut microbiota, an assortment of organisms that operate in unison with the host, has been identified as a prominent area of study in the promotion of health but also the initiation of disease (Marchesi et al. 2016).

Metabolic engineering is the design and development of microbial strains that can aptly produce valuable compounds, and is instrumental in industrial biotechnology. Today more than ever, climate change and threatened energy security have rendered the production of sustainable alternatives to fossil fuels a necessity (Ragauskas et al. 2006, Rodionova et al. 2017). The conversion of renewable biomass to valuable fuels and products through metabolic engineering has enabled the development of large-scale biorefineries (Khan, Shin and Kim 2018) as well as advances in

1.1 Background

biomanufacturing through the use of cell-free systems (Zhang 2010). Throughout the years, metabolism has been established as a prosperous area of research and has maintained its importance in the ever-shifting biology field.

Metabolism is extremely complex and thus requires a systems approach towards deciphering the fundamental mechanisms and cellular processes that will lead to solutions for metabolic engineering applications. Metabolic networks and their mathematical representations are indispensable the study of all the different levels of cellular organization and regulation. These networks are built through gene-protein-reaction (GPR) associations (Francke, Siezen and Teusink 2005), and are the backbone of metabolic models (Christensen and Nielsen 2000). Stoichiometric metabolic models are based on a stoichiometric matrix, which stores the information of all the metabolites and their stoichiometric participation in each reaction of the network. Due to the advent and advancement of genome sequencing techniques over the years, these models systematically evolved into genome-scale metabolic networks (GEMs). GEMs contain the entirety of available information for an organism and are built using a bottom-up approach, which includes draft reconstruction, manual curation, conversion to a mathematical model, and validation (Price et al. 2003, Thiele and Palsson 2010). Swift progress in genome sequencing techniques are making available the complete genome sequence for an array of organisms. The genomic data can thus be incorporated into metabolic models, making them more comprehensive, and enhancing their predictive capabilities.

The ever-increasing availability of omics data is leading to a rapid increase of model size and complexity. As such, metabolic models cannot be intuitively understood or analyzed. To this end, a variety of computational tools have been developed for the analysis of these models (Tomar and De 2013). The metabolite mass balances in a metabolic network form a set of linear equations. Accompanied by the quasi-steady state assumption, which assumes that there is no accumulation of any metabolite in the system, and the definition of an appropriate cellular objective, this set of equations defines a constraint-based optimization problem. In a properly constructed metabolic network, the number of equations (mass balances) is smaller than the number of variables (reaction fluxes), meaning that the system is underdetermined and that the

optimization problem formulated for metabolic networks does not have a unique solution (Bonarius, Schmid and Tramper 1997).

Flux Balance Analysis (FBA) introduces linear constraints to the mathematical problem in the form of inequalities, leading to a feasible solution space, which is delimited by imposed bounds on the system variables (Orth, Thiele and Palsson 2010, Varma and Palsson 1994). However, since this space is convex, there is only one global optimum value for a specified objective function. Biologically, this means that there is an infinite number of flux distributions throughout the network that satisfy the same cellular objective. Many different objectives have been proposed and tested throughout the evolution of systems biology, in an effort to efficiently navigate the solution space and capture the cell's physiology. The most predominantly used objective functions are the maximization of biomass production and ATP formation (Schuetz, Kuepfer and Sauer 2007). Another popular objective is the minimization of the sum of all fluxes in the network, a formulation which is called parsimonious FBA (Lewis et al. 2010). Very recently, multi-objective methods employing up to four simultaneous cellular objectives have been proposed (Dai et al. 2019, Gardner, Hodge and Boyle 2019). The existence of multiple alternative solutions results in an associated uncertainty in terms of metabolic fluxes, that increases with the size of the model.

Researchers have proposed numerous methods to characterize and ultimately reduce this uncertainty (Mahadevan and Schilling 2003). The first method to this end is to identify the reaction directionalities in the network. If an enzyme is catalytically reversible, then the reaction directionality depends on the reaction's displacement from thermodynamic equilibrium. In a metabolic network, kinetically reversible reactions can be either bidirectional, which means that they can operate in both the forward and reverse directions, or unidirectional, which means that they can operate only in one direction. Few enzymes have been classified as kinetically irreversible, and the vast majority catalyzes reactions that can operate in both directions. Therefore, the integration of thermodynamic feasibility constraints is an invaluable tool in reducing the uncertainty stemming from bidirectionality, as they can reveal the thermodynamically permissible directionalities of reactions (Soh and Hatzimanikatis 2014). Using the change in Gibbs Free Energy of reactions,

1.1 Background

thermodynamics can also give estimate each reaction's displacement from equilibrium which, apart from participating in the estimation of directionalities, is an important information for constructing dynamic models. Over the years, multiple approaches have been proposed for the formulation of thermodynamic constraints in metabolic networks (Ataman and Hatzimanikatis 2015). A prominent approach in this field is Thermodynamics-based Flux Analysis (TFA) (Henry, Broadbelt and Hatzimanikatis 2007), which employs a mixed-integer linear programming (MILP) formulation, uses Gibbs free energy constraints as a function of logarithmic concentrations, and computes the reaction directionality based on their thermodynamically feasible ranges, thus introducing a minimum possible bias in the calculations. TFA also accommodates the integration of metabolomics data in the network, since the thermodynamic feasibility constraints introduce metabolite concentrations as variables in the formulation.

In order to improve the accuracy of model predictions, approaches that aim to incorporate various types of experimental data through network constraints have been proposed. Efforts have been made in relating gene expression levels with metabolic fluxes through genomics (Lerman et al. 2012, O'Brien et al. 2013, Pandey, Hadadi and Hatzimanikatis 2019) and proteomics (Liu et al. 2014, Yang et al. 2016). Recently, resource allocation limitations along with enzyme kinetic properties have been added to these models (Sánchez et al. 2017), along with thermodynamic feasibility constraints (Salvy and Hatzimanikatis 2019).

With GEMs being the most detailed representations of metabolism, but of large complexity, and data-driven context-specific models often being constructed in an *ad hoc* manner, multiple frameworks for a systematic way to produce consistent context-specific models have been proposed to bridge this gap (Singh and Lercher 2019). One of them is redGEM (Ataman et al. 2017), a framework that employs a graph-based search algorithm to construct reduced metabolic models of variable size from GEMs by focusing on selected metabolic subsystems. The generated models are consistent with their parent model in terms of flux and concentration variability and essential genes/reactions (Hameri, Fengos and Hatzimanikatis 2019). The companion algorithm to this framework is called lumpGEM (Ataman and Hatzimanikatis 2017), and further reduces the complexity of GEMs in terms of network topology. lumpGEM

uses an optimization formulation to identify and enumerate the possible reaction subnetworks that are required for the production of each biomass building block (BBB) from core carbon metabolites. Moreover, each of these subnetworks is then lumped into a single reaction, forming reduced models that are not dependent on the biosynthetic networks of BBBs.

Although stoichiometric models of metabolism are great tools for the analysis of steady state flux and concentration distributions, they cannot capture dynamic responses to perturbations and transient behaviors of cells. Over the past years, constraint-based modeling has been enhanced through formulations that aim to capture these properties by integrating kinetic information (Saa and Nielsen 2017, Strutz et al. 2019). Since detailed knowledge about regulatory mechanisms and kinetic parameter values is still limited, constructing consistent large-scale kinetic models is a difficult task (Hameri et al. 2019). The uncertainty stemming from the deficiency of knowledge in combination with model complexity can lead to the existence of multiple kinetic parametrizations of models that describe the same observed physiology (Link, Christodoulou and Sauer 2014). In early efforts to generate kinetic models of metabolism, researchers proposed various simplified kinetic formulations in order to accommodate parameter fitting using *in vivo* data (Hatzimanikatis and Bailey 1997, Savageau 1969). However, these methods do not account for thermodynamic principles, and their predictions are limited to the vicinity of the chosen operational parameters.

Advances in sampling-based strategies for kinetic model construction and analysis have bridged this gap by generating populations of kinetic parameters that satisfy thermodynamic constraints. Both the Ensemble Modeling (EM) (Tran, Rizk and Liao 2008) and Optimization and Risk Analysis of Complex Living Entities (ORACLE) (Miskovic and Hatzimanikatis 2010) workflows efficiently sample the permissible parameter space and produce a multitude of kinetic models around a selected metabolic steady state. EM then prunes the generated models through comparison with experimental data to select a unique set of parameters that describes the desired physiology. On the other hand, ORACLE considers all of the generated models as distributions of kinetic parameters, and performs statistical quantification and analysis, thus introducing a lesser amount of bias. Smart Monte-Carlo sampling can

1.2 Aim and Scope

also be applied to reduce uncertainty in kinetic parameters and allow for the definition of more complex kinetic mechanisms (Miskovic et al. 2019b). Recently, a method that combines parameter sampling methods and machine learning techniques was proposed, using integrated datasets to prune the a priori generated kinetic parameters (Andreozzi, Miskovic and Hatzimanikatis 2016, Miskovic et al. 2019a).

When studying dynamic properties, one important aspect is the identification of the enzyme (or enzymes) which controls the amount of flux through a metabolic pathway. For a long time, the belief that a single enzyme controls the flux rate was predominant, and the reaction being catalyzed by said enzyme was the rate-limiting step of the pathway. With the introduction of Metabolic Control Analysis (MCA) (Kacser and Burns 1973b), it became evident that these relationships are not intuitive, and metabolic control is distributed among multiple enzymes in the network (Schuster 1999). MCA uses Control Coefficients (CCs) to quantitatively capture a network's response to perturbations on systemic parameters. Since its conception, the MCA framework has been extensively developed (Fell and Sauro 1985, Hatzimanikatis and Bailey 1996, Hatzimanikatis and Bailey 1997, Heinrich and Rapoport 1974) and has been established as a powerful tool in multiple scientific areas, such as biotechnology and medicine (Bowden 1999, Cascante et al. 2002, Nishiguchi et al. 2019, Ward, Chatzivasileiou and Stephanopoulos 2019). However, MCA is a local sensitivity method and is thus limited to the study of small perturbations around the reference steady state.

1.2 Aim and Scope

1.2.1. Lipid metabolism

Lipids are a diverse group of hydrophobic organic compounds that includes fats, oils, waxes, and hormones. They also include phospholipids, sphingolipids, and steroids, which comprise the major components of cellular membranes. Due to their structure they serve multiple purposes in an organism, such as acting as energy storage sources

and participating in signaling cascades. Lipids also regulate a range of biological processes spanning from cell proliferation and apoptosis to angiogenesis and immunity.

The metabolism of lipids and its regulation has been closely associated with numerous physiopathologies. Lipid metabolism provides the cell membranes with the required lipids in the correct proportions to ensure their proper function and is responsible for the appropriate adjustment of these proportions in response to external stimuli. Impaired lipid homeostasis can lead to acute metabolic disorders such as Parkinson's and Alzheimer's disease, and diabetes (Kosicek and Hecimovic 2013, Markgraf, Al-Hasani and Lehr 2016, Santiago and Potashkin 2013). An atypically active lipid metabolism has also been observed in cancer cells, allowing them to rapidly proliferate (Alves et al. 2016, Ogretmen and Hannun 2004, Vriens et al. 2019). Although lipid pathways are in principle well characterized, the intricacies of their networks are yet to be completely understood.

Lipidome is the term used to describe the totality of lipid species present in an organism. Lipidomics is the large-scale study of lipid networks and pathways in biological systems, and its aim is to interpret and characterize the lipidome (Dennis 2009). Similar to genomics, advancements in mass spectrometry (MS)-based methodologies also enable the detailed analysis and mapping of the lipidome. The identification and quantification of thousands of lipid species can greatly facilitate our basic understanding of cellular metabolism and regulation (Han 2016, Wenk 2005), and aid in the development of medical diagnostics and treatments (Lydic and Goo 2018).

Lipidome characterization is accompanied by the need of appropriate tools for the analysis of the data and statistical quantification thereof. Computational metabolic models of various lipid pathways have emerged in an effort to evaluate the vast omics data available. Top-down approaches have been used to construct context-specific models of metabolism from condition-specific metabolome data, and process this data in a mathematical formulation to extract information content so as the active network is deduced (Cakir et al. 2009). These models, however, are usually limited to specific data sets and cannot serve multiple studies (Cook and Nielsen 2017).

1.2.2. Yeast as a model organism

Yeast has been established as a prominent model organism for the study of lipid biochemistry (Klug and Daum 2014). Apart from a very well documented genome sequence (Goffeau et al. 1996, Mewes et al. 1997), yeast is easy and inexpensive to cultivate and genetically modify (Santos and Riezman 2012). Even shortly after the whole genome sequence of *S. cerevisiae* was published, wide mutant strain collections were made available and their number has only increased since (Goffeau 2000, Norman and Kumar 2016). *S. cerevisiae* can synthesize most of the lipid classes currently identified, and possesses a high degree of homology with the human genome. Moreover, the majority of regulatory mechanisms are preserved between the two species, thus yeast could perform brilliantly as a platform to study lipid dysregulation in human cells. (Natter and Kohlwein 2013, Nielsen 2009, Petranovic et al. 2010).

The first GEM of yeast was published in 2003 for the metabolic network of *S. cerevisiae* (Forster et al. 2003). Since then, multiple yeast GEMs have been updated and published by several research groups (Lopes and Rocha 2017). One of these models, namely iIN800 (Nookaew et al. 2008), especially focused in the accurate representation of lipid pathways, although it only considered two cellular compartments. In an effort to enforce community standards and resolve inconsistencies in annotation, a consensus network reconstruction has been developed, with its latest version being Yeast 8 (Lu et al. 2019). Yeast 8 is the largest and most detailed yeast GEM to date; it encompasses almost 4000 reactions involving more than 2500 metabolites. Although GEMs are a useful resource in studying lipids in the context of cell metabolism, their large size introduces a lot of mathematical complexity, and it can be difficult to handle.

1.2.3. Kinetic modeling of lipids

For the study of lipid metabolism and its regulation, kinetic models are essential. Lipid regulatory mechanisms are imperative to maintain membrane permeability and cell homeostasis. Multiple efforts to capture the complexities of these mechanisms have been made the past decades using both deterministic and stochastic approaches

(Alvarez-Vasquez et al. 2011, Alvarez-Vasquez et al. 2004, Gupta et al. 2011, Knoblauch et al. 2000, Schutzhold et al. 2016). However, all of these studies focused on smaller parts of the lipid network by examining few pathways at a time, and the reasons are manifold. As mentioned above, large-scale kinetic modeling is very demanding and requires an extensive parametrization with little available knowledge. Furthermore, enzymes associated with lipid pathways are highly promiscuous, which means that they do not have an affinity for a particular lipid species but for whole lipid classes (Carbonell, Lecointre and Faulon 2011). This effect has been not taken into account in the vast majority of existing studies, although it has been shown to be crucial in predicting accurately the network's dynamic responses (Savoglidis et al. 2016).

1.2.4. Metabolic engineering strategies in sustainability and health

Industrial bioprocesses exploit microorganisms by using them as cell factories that convert sugars into valuable chemicals. Mathematical modeling is vital tool in the constant development of such processes and kinetic models of metabolism are especially utilized in synthetic biology and metabolic design (Almquist et al. 2014, Sarkar and Maranas 2019). Metabolic engineering aims to increase the production of valuable compounds through genetic modifications. Synthetic biology is taking this process one step further by creating genetic material that is not native to an organism, and incorporating it in its genetic makeup. Kinetic models can be used to predict and evaluate the effects of modifying various components of a cell factory and assist in the design of modification strategies. The key aspect of manipulating an organism's capabilities is the identification of possible gene modification targets, which ultimately translates in changes in enzymatic expression levels. MCA has also been widely used to extract vital information concerning metabolic engineering approaches (Fell 1998, He, Murabito and Westerhoff 2016, Moreno-Sanchez et al. 2008, Volke et al. 2019). CCs express on a basic level the influence that each genetic modification will have in flux rates and concentration levels throughout the network. However, MCA fails to account for cascade-like effects that these modifications may cause due to fact that CCs only consider the changes that a single parameter will

instigate in the network, uncoupled from the rest. Additionally, MCA does not include any form of physiological constraints, often leading to unrealistic predictions.

Stepping back from cell factories, humans have been selecting desirable traits in one species and transferring them to another for centuries (i.e. cross-breeding in plants and animals). Although the genetic engineering tools to allow the transfer of genetic material and the traits associated with it are widely available, the mechanisms of evolutionary selection are still not completely understood. Over the past years, studies have focused on the development of computational tools for the prediction of mutant fitness and the pruning of mutant libraries (Heckmann, Zielinski and Palsson 2018, Zaugg et al. 2014). Even though the genetic restructuring taking place in each mutant generation can be studied *a posteriori*, the progression of metabolic restructuring remains largely unexplored. Analogously, the study of enzymatic and metabolic restructuring could immensely increase our fundamental understanding of how diseases evolve. To date, sequencing the genome of the patient has been the first step to diagnosis and treatment. However, even by selectively resequencing parts of the genome, the cost can amount to thousands of dollars (Benner et al. 2008). As computational biomedicine and personalized medicine are still in their infancy, the search for efficient tools aiding in phenotyping and the identification of relevant biomarkers is still ongoing (Ho and Chen 2017, Jain 2013).

1.3 Thesis Overview

In **Chapter 2** we present a novel metabolic model that focuses on the lipid metabolism of *S. cerevisiae*. This model was assembled by gathering available detailed pathway knowledge of the lipid network and integrating it into an already existing GEM. This integrated model was then systematically reduced around the subsystems defined by these pathways using the redGEM and lumpGEM framework, to provide a comprehensive network of manageable size for complex studies of metabolism. We additionally curated this model with GPR annotations and thermodynamic feasibility

constraints, and show it is as consistent and inclusive as other yeast GEMs regarding the focus and detail on the lipid metabolism.

In **Chapter 3** we employ the ORACLE framework to construct a kinetic model of the lipid metabolism in yeast, using the stoichiometric model that was developed in Chapter 2. We calibrated this model using lipidomics data and accounted for promiscuity in the enzymatic properties. We demonstrate the significance of enzymatic promiscuity through MCA, and show that the model's predictions are in agreement with biological observations. This model is to our knowledge the largest and most detailed kinetic model of the lipid metabolism to date.

In **Chapter 4** we develop a formulation that uses MCA-based CCs to generate constraint-based models for the consistent derivation of metabolic engineering strategies. We systematically introduced and parametrized sets of relevant biological constraints to the network. Finally, the advantages of this formulation in comparison to the classical MCA approach and its ability to generate alternative strategies are highlighted.

In **Chapter 5** we enhance the formulation that was developed in Chapter 4 to accommodate reverse metabolic engineering design. The optimization objective was appropriately modified in order to enable the mapping of underlying enzymatic and metabolic restructuring mechanisms of the cell between various physiological states. Furthermore, an additional workflow was defined for the identification of underlying cell mutations using metabolomics data.

Concluding, in **Chapter 6** we summarize the conclusions of this thesis and discuss the applications and future perspectives of the developed tools and workflows in the research field.

The supplementary material for this thesis is available in an electronic format at the following link:

<https://zenodo.org/record/3585882>

Articles included in this thesis

Tsouka S. and Hatzimanikatis V. (FEMS Yeast Research - in press), redLips: a comprehensive mechanistic model of the lipid metabolic network of yeast.

Tsouka S. and Hatzimanikatis V. (in preparation), Exploring the effect of enzymatic coupling on kinetic modeling and metabolic control across the lipid pathways of yeast.

Tsouka S., Hameri T., Ataman M., Miskovic L., Hatzimanikatis V. (in preparation), Response Balance Analysis: investigating metabolic control in a constraint-based formulation for the design of strain engineering strategies.

Tsouka S., Hameri T., Hatzimanikatis V. (in preparation), inverse Response Balance Analysis: a reverse engineering formulation for the mapping of enzymes responsible for diverse phenotypes.

Chapter 2 – redLips: a comprehensive mechanistic model of the lipid metabolic network of yeast

2.1 Introduction

Even the slightest changes in cellular membrane composition, which all serve a specific biological purpose, can affect many cellular functions from signaling cascades to the modulation of membrane fluidity (Guan et al. 2009). Because they are the main structural component of cellular membranes, lipid imbalances have been shown to be involved in various physiopathologies concerning membrane lipid homeostasis (Holthuis and Menon 2014). Yeast is a very prominent model organism for the study of numerous parts of cell metabolism including, but not limited to, lipid-related cellular processes (de Kroon 2017, Klose et al. 2012) because it is easy and inexpensive to cultivate and modify its genome for experiments, and it has a well-documented genome sequence (Santos and Riezman 2012). Consequently, an increasing spectrum of yeast mutants has been made available, providing great opportunities for studies on the effects of lipid metabolism perturbations at molecular and cellular levels. One additional feature of yeast is its high homology to the human genome. Most importantly, the majority of regulatory mechanisms are preserved between the species (Petranovic et al. 2010). This means that yeast could potentially be used as a platform to study lipid dysregulation in humans, making the study of potential causalities and treatments critically easier. The similarities and differences of the two organisms, along with the potentials for comparative analysis

have been reviewed in detail by Nielsen (2009). Unfortunately, the intricacies of how lipids tie to many biological functions, including those leading to disease, remain unknown. This means that comprehensive lipid identification and characterization and detailed studies of lipidomics are needed for a fundamental understanding of cellular metabolism (da Silveira Dos Santos et al. 2014, Han 2016, Ivanova et al. 2009, Kontush and Chapman 2010, Wenk 2005), and as such, recent interdisciplinary approaches are beginning to reveal novel lipid functions and interactions (Harayama and Riezman 2018). Eventually, lipidome profiling could be used as a predictive tool to further enhance our knowledge of the underlying molecular mechanisms typifying lipid dysregulation.

While traditionally established work on cellular lipid metabolism has been limited to the analysis of individual classes of lipids or specific lipid species, progress in mass spectrometry (MS)-based methodologies has allowed the analysis of the entirety of the lipids in a cell. Computational metabolic models of various pathways have emerged in an effort to evaluate the vast omics data available, and many different approaches for their construction and curation with incorporated omics data have been developed (Joyce and Palsson 2006). This mostly involves genome-scale models (GEMs) of metabolism, which are reconstructions of an organism's metabolism from genomic, biochemical, and physiological data, and in principle, contain the majority of known information for the modeled organism. With the increasing availability of omics data, however, comes increasing mathematical complexity, and it can be very complicated to handle the incorporation of experimental data in such large-scale models. The potential of dynamic modeling through the generation of appropriate sets of ordinary differential equations that describe the network topology is also hindered by the model's size. Mathematically, a larger model also leads to an increased solution space, which ultimately contributes to increased uncertainty in the model's predictions. Therefore, it is essential that a network is manageable with respect to size without a loss of information, so the redGEM framework was proposed as a way to systematically reduce GEMs around a biological context of interest with minimal loss of information and connectivity (Ataman et al. 2017). On the other hand, due to the rapid discovery of novel species through innovative technologies, a gap is emerging between the existing pathway representations of lipids and lipid structure

databases. An approach aiming to bridge this gap has been proposed, termed Network Integrated Computational Explorer for Lipidomics (NICELips) (Hadadi et al. 2014), and this framework can postulate novel lipid biosynthesis pathways using generalized enzymatic reaction rules. Specific to yeast metabolism, the first GEM of *S. cerevisiae* was published in 2003 (Forster et al. 2003), and over the years, multiple yeast GEMs have been updated and published by several research groups (Lopes and Rocha 2017). Due to inconsistencies in annotation, a community consensus reconstruction has been developed, with its latest versions being Yeast 7 and Yeast 8 (Aung, Henry and Walker 2013, Lu et al. 2019). Very recently, a novel method for the representation of lipid requirements in GEMs was proposed (Sanchez et al. 2019).

We thus sought to develop a metabolic model that could act as detailed repository of lipid metabolism for *S. cerevisiae*. Starting from the network provided by Savoglidis et al. (2016), we gathered all relevant reaction and pathway information available in the literature and databases. To ensure consistency with the well annotated GEMs, we incorporated these data into a GEM of the yeast *S. cerevisiae*, expanding its preexisting lipid description. We then performed a systematic reduction of the integrated model around the lipid subsystems to preserve the focus of the model on the lipid metabolic pathways and to simultaneously retain the connections to the rest of the cell metabolism. To ensure consistency of the cell biomass composition, we computed lumped reactions to establish the production of all biomass building blocks (BBBs). These steps made sure that our final model, termed “reduced lipids-centric model” (redLips), is inclusive yet concise and as consistent as the other available yeast GEMs. We have created a detailed thermodynamic database for all the metabolites of the network and performed a complete thermodynamic curation of redLips, a procedure that decreases the mathematical uncertainty and imposes physiological constraints. We also demonstrate how it can be used as a scaffold for lipidomic measurement implementation. redLips can be modified to accommodate simulations and predictions for human (or other) metabolism, thus creating a platform to study lipid regulation for applications across organisms.

2.2 Materials and Methods

2.2.1 Starting reaction network

We used the model of Savoglidis et al. (2016) as a base to gather and reassemble the available knowledge on lipid metabolism to date. The LIPID MAPS classification system distinguishes eight major lipid categories: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides (Fahy et al. 2005, Fahy et al. 2009), and the above cited model focused on the sphingolipid biosynthesis pathway and included some of the glycerophospholipid biosynthetic route. The model was curated using thermodynamic and lipidomics data, and an extensive study on the control asserted by the highly multifunctional enzymes of the system was conducted.

The resulting gathered lipid reactions network (GLRN) was constructed by combining information found in the literature, GEMs, and databases. Primary sources of data include the online repositories *Saccharomyces Genome Database* (SGD, <https://www.yeastgenome.org>, (Cherry et al. 1998)), KEGG (<https://www.genome.jp/kegg/>), and Lipid Maps (<https://www.lipidmaps.org>) as well as relevant journal publications and books (Dickinson and Schweizer 2004).

2.2.2 Consistent reduction of models

An issue that arises when modeling only a part of cell metabolism is the connection to the rest of the network. For example, if the lipid network was to be studied without including the TCA cycle, ATP would need to be obtained through an artificial transport reaction from the extra-model domain to the intra-model domain, though there is no such compartmental transport in reality. This can lead to uncertainty on the concentration levels as well as to a major question of the relevant flux constraints. To create a consistent and reliable model, we would need to constrain the flux values of all these transport reactions (which would include mostly cofactors) to realistic values.

To overcome this issue, we decided to effectively couple our model with a GEM that will account for any non-realistic assumptions that would have to be made. We did

this by first incorporating our detailed lipid network into a GEM of choice, thus expanding the lipid metabolism pathways already present, then we utilized the redGEM framework to obtain a reduced model using our original subsystems as the starting network.

redGEM is a framework developed by Ataman *et al.* (Ataman et al. 2017) to systematically and consistently reduce genome-scale models. It focuses on chosen parts (subsystems) of the metabolic networks that are then connected to each other up to a user-defined degree of connection. This measure describes the distance in terms of reaction steps between a subsystem pair and can be either imposed by the user for all subsystem pairs or can be equal to the intrinsic minimum distance between each pair. Subsequently, the resulting core network is connected to the biomass building blocks (BBBs) using lumpGEM (Ataman and Hatzimanikatis 2017).

In redGEM, a graph search algorithm is employed to identify all possible connections between metabolites belonging either to the same subsystem or different ones (excluding cofactors). The first step is the intra-expansion of the starting network, connecting metabolites within each subsystem of interest. Then these subsystems are step-wise connected to each other, first adding the one-step connections, then the two-step connections (which will involve an intermediate), etc., thus creating the core network. The degree of connection is symbolized as $D\#$, where $\#$ is the number corresponding to the desired connection length.

After the network expansion, lumpGEM is used to identify sets of biosynthetic subnetworks that will synthesize each BBB that cannot already be produced by the core network. In other words, lumpGEM's objective is the minimization of the number of reactions that need to be added to the core to allow the production of each BBB. These sets are then collapsed into elementally balanced lumped reactions. lumpGEM first identifies the minimal subnetwork of reactions needed to connect the expanded network to each BBB. Subsequently, all alternative subnetworks of this minimal size can also be computed and are then translated into a single lumped reaction that is tested for feasibility in terms of stoichiometry and thermodynamics. Various consistency checks are performed to ensure the minimal loss of information during the reduction process. These checks include flux variability and essentiality studies in

both stoichiometric and thermodynamic levels of curation between the GEM and the reduced model.

2.2.3 Genome Scale Model (before integration)

We integrated the GLRN into the well-known and well-studied iMM904 GEM (Mo, Palsson and Herrgard 2009), which is annotated, ensuring that it was straightforward to match each reaction and its metabolites between the two networks. iMM904 also includes a large number of cellular compartments compared to most yeast GEMs. The interested reader may refer to (Lopes and Rocha 2017), (Sanchez and Nielsen 2015) and (Osterlund, Nookaew and Nielsen 2012) for a more detailed review of the development and evolution of various *S. cerevisiae* GEMs.

2.2.4 Gathered Lipid Reactions Network (before integration)

The GLRN encompasses more than 500 enzymatic reactions and 300 metabolites. We considered 7 cellular compartments where all the reactions take place, which are the cytosol, mitochondria, endoplasmic reticulum (ER), peroxisomes, Golgi apparatus, vacuole, and nucleus (as well as extracellular space). The model can be organized into 15 subsystems: glycolysis, pyruvate metabolism, fatty acid biosynthesis, fatty acid mitochondrial biosynthesis, fatty acid elongation, fatty acid degradation, phospholipid biosynthesis, sphingolipid biosynthesis, sterol biosynthesis and esterification, sterol metabolism, mevalonate pathway, dolichol biosynthesis, cardiolipin biosynthesis, carnitine shuttle, and triacylglyceride decomposition. We did not consider any membrane compartments or lipid bodies in our study since our thermodynamic calculations do not hold for non-aqueous solutions, as will be explained shortly, and we instead opted for consistency over extensive detail.

The localization assignment for each reaction was made according to the Yeast7 and Yeast8 consensus GEMs (Aung, Henry and Walker 2013, Lu et al. 2019). For the reactions that are not included in this model, the N-terminal amino acid sequence of the associated gene was used to predict localization (Emanuelsson et al. 2007).

2.2.5 Genome Scale Model (after integration)

After we integrated the GLRN into the GEM, the integrated model had 2181 reactions and 1551 metabolites. The lipid-related reaction subsystems of iMM904 that were mostly expanded were the fatty acid biosynthesis and degradation, as well as the sterol biosynthesis and esterification, all of which existed mostly as lumped reactions or were missing parts of the pathways. The phospholipid and sphingolipid biosynthetic pathways originally included mostly mass-imbalanced and pooled reactions and were also greatly enhanced, with parts like phospholipid remodeling being added. Similarly, the lipid species that were added to the model mostly included fatty acids of different carbon chain lengths, complex sphingolipids, monolyso-glycerophospholipids, and fatty acid biosynthesis and degradation as well as sterol intermediates, over all of the cellular compartments.

2.2.6 Lipidomics – biosynthetic fluxes

Lipids are an essential component of the cell's various membranes and are critical for cell survival. Thus, being essential to biomass formation, they should be present in the modeled assumption of the biomass composition. However, few GEMs even consider these lipids as part of the growth requirements, let alone encompass the lipid network in detail. We have identified 37 metabolites that should be considered, which are 4 phospholipids (phosphatidylethanolamine, phosphatidylcholine, phosphatidylserine and phosphatidylinositol), 4 lyso-phospholipids (lyso-phosphatidylethanolamine, lyso-phosphatidylcholine, lyso-phosphatidylserine and lyso-phosphatidylinositol), 20 complex sphingolipids, ergosterol, 4 sterol esters (ergosterol, episterol, lanosterol and zymosterol esters), dolichol, as well as long and very long chain fatty acids.

We did not wish to alter the biomass reaction already defined in iMM904, so we defined 35 additional biosynthetic reactions. These reactions are all single (or double) species exchange reactions, all of which are essential to cell growth. This artificial representation corresponds to elementary fluxes of the aforementioned lipid species towards biomass formation. These fluxes can be constrained based on experimental concentration measurements (when available) as:

$$\mu(\bar{Y}_i - sd) \leq v_i \leq \mu(\bar{Y}_i + sd),$$

where v_i are the biosynthetic reaction fluxes, \bar{Y}_i the mean of the lipidomic content measurements, sd the experimental measurements' standard deviation of \bar{Y}_i , and μ the specific growth rate of the cell as calculated from the flux through the biomass objective function.

It is important to note that when a species is already considered in the biomass composition of the GEM, the experimental constraint is altered accordingly to consider the corresponding amount required for each contribution.

2.2.7 Thermodynamics

Next, we performed a complete thermodynamic curation of the integrated model using TFA (Salvy et al. 2019) to further reduce the solution space of the problem and help identify reaction directionalities. For a reaction to be feasible in the assumed directionality, the net change in Gibbs free energy of a reaction ($\Delta_r G'^o$) must be negative in this direction.

Lipids are very complex molecules, and thermodynamic information about them, such as the Gibbs free energies of formation and dissolution constants, is scarce. Where available, experimental observations indicating a pathway direction were used, which in turn provided insight for the whole reaction network. Otherwise, group contribution methods were used, which predict properties of complex molecules by using group or atom properties (Mavrovouniotis 1990, Mavrovouniotis 1991). Thus, very complicated molecules can be decomposed into a number of simple groups, and their individual contributions to the total properties can be estimated.

Since thermodynamic properties depend on the pH of the environment, we needed to assign a pH value to each considered compartment (Orij et al. 2009, Paroutis, Touret and Grinstein 2004, Preston, Murphy and Jones 1989). For cross-membrane transport reactions, we also needed to take the membrane potential difference, if any (Cohen and Venkatachalam 2014), into account. Finally, all of the calculated changes in Gibbs free energy needed to be adjusted with the associated compartmental ionic strength (Ataman 2016). All of these values can be found in Table 2.1.

Table 2.1. Values for pH and ionic strength (in M) for each model compartment, and cross-membrane potentials (in mV) for each set of these compartments (where applicable) – opposite arrow direction will correspond to the same value with opposite sign.

#	Compartment	pH	Ionic Strength (M)	Cross-membrane potential (mV)
I	Cytosol	7	0.25	n/a
II	Endoplasmic Reticulum	7.2	0	n/a
III	Golgi Apparatus	6.35	0	n/a
IV	Mitochondria	7.5	0.25	IV → I: 180
V	Nucleus	7	0	V → I: 15
VI	Peroxisome	8.2	0	n/a
VII	Vacuole	6.17	0	n/a
VIII	Extracellular	5	0	VIII → I: -60

To estimate the properties of a lipid containing a fatty acyl carbon chain, we needed to assume a chain length for each of the attached R groups. We chose C16:0 (where the first number denotes the carbon chain length and second denotes the number of unsaturations on this chain) for all species, since this chain length represents the vast majority of lipids in eukaryotes. Regardless, this assumption does not carry much weight in our model, since the group contribution method used to estimate the Gibbs free energy of a reaction considers only the groups that undergo a molecular change. Consequently, if the R group is not the reactive part of the molecule participating in the reaction, its length will not affect the calculated $\Delta_r G'^o$ value. One more assumption that needed to be made was that no reactions occurred inside membranes. It is known that this is not the case for numerous lipid biotransformations, but since all thermodynamic properties have been measured with the assumption of an aqueous solution and are computed accordingly, it was a necessary assumption.

With these in mind, we curated a thermodynamic database containing all the thermodynamic properties of the model's metabolites, such as pKa, standard Gibbs free energy of formation, formula, charge, etc. These properties were calculated through Chemaxon (<https://www.chemaxon.com>). This database covers 90.4% of the

integrated network's metabolites, which allowed us to calculate 87.4% of the $\Delta_r G'^o$ of the network reactions.

Assumptions made about the thermodynamic constraints, such as temperature and pH, or even uncertainty in the calculation and the standard deviation of measurements, can render networks computationally infeasible. Additionally, especially concerning lipid metabolism, channeling phenomena can lead to apparently infeasible reactions in a certain directionality. Regardless, since we were confident in most of the reaction directionalities in our network, we could adjust some thermodynamic constraints to attain feasible solutions in what we consider physiological conditions. More specifically, in order to retain consistency with yeast physiology, we relaxed 62 thermodynamic feasibility constraints in terms of the $\Delta_r G'^o$. These constraints correspond in majority to lipid species transport reactions across intracellular compartments. This is actually a case for which our computations may not hold, though, since lipid species do not cross membranes in the same way as most others. The complete list of the $\Delta_r G'^o$ relaxations can be found in supplementary Table S2.3.

2.2.8 Fatty Acid chain lengths

Lipid species consist of R groups of acyl chains exchanged between themselves or provided by free fatty acids. These chains vary in size and usually contain an even number of carbon atoms from 4 to 28 as well as often one unsaturation. As mentioned previously, the most abundant fatty acids in yeast have a 16-carbon chain, and the second most abundant have an 18-carbon chain, which together comprise more than 70% of the total fatty acid population (Daum et al. 1999, Schneiter et al. 1999). In this model, we consider acyl chains only of even chain lengths varying from C8 to C26. Because any chain length or combination thereof could react to form a lipid species, we treated the fatty acids (in both inactive and coenzyme A [CoA]-activated form) as metabolite pools, which comprised all of the fatty acyl providers. We also defined a metabolite pool for polyprenol diphosphates, which include species possessing 14 to 22 prenyl units.

2.2.9 Lipidomics – concentrations

As mentioned above, experimental measurements can be used to constrain fluxes and effectively couple them to biomass formation. The metabolic concentrations of species can also be constrained through lipidomics as:

$$\ln(\bar{X}_i - sd) \leq LC_i \leq \ln(\bar{X}_i + sd),$$

where \bar{X}_i is the mean of the concentration measurements, sd the standard deviation of the experimental measurements of \bar{X}_i , and LC_i are the natural logarithms of the concentrations for each compound.

2.2.10 Media

To ensure that the maximum growth rate predicted by the model reflects a typical growth rate for yeast in aerobic conditions (about 0.32-0.48 h⁻¹), we constrained the maximum uptake of glucose (which we considered to be the sole carbon source) to 4 mmol·gDW⁻¹·h⁻¹ (Orij et al. 2012). The other uptakes allowed were the following inorganics: hydrogen, water, ammonium, oxygen (limited to 20 mmol·gDW⁻¹·h⁻¹), phosphate, and sulfate. We also had the option to allow a basal uptake of exogenous ethanolamine (we chose a value of up to 0.02 mmol·gDW⁻¹·h⁻¹) to activate the reaction catalyzed by ethanolamine kinase (EKI1), the first step of phosphatidylethanolamine (PE) synthesis via the Kennedy pathway.

2.3 Results and Discussion

2.3.1 redGEM output model

To form our lipid-focused reduced metabolic model, we first applied the redGEM and lumpGEM algorithms to the previously defined subsystems of interest, GLRN, and the glycolysis pathway. We also included the electron transport chain (ETC) reactions to the starting subsystems to ensure consistent energy associations and that the growth rates were as equivalent to the GEM as possible. For this reduction, we set the degree

of connection to 3, which means that pairwise subsystem connections of up to 3 steps each will be added during the subsequent network expansions. The resulting model encompassed 1130 reactions, of which 639 were enzymatic, 419 were transport or boundary, 35 were biosynthetic (as described in the Materials and Methods section), and 37 were lumped reactions. Additionally, the reduced model included 800 metabolites, 404 of which were unique across compartments.

After formation of the reduced model, to ensure and evaluate its function and the minimal loss of information from the integrated model, we conducted consistency checks in terms of enzyme essentiality and thermodynamic flux variability. The results from these tests can be found in supplementary Tables S2.1 and S2.2. These tests showed that, as expected, redLips exhibits equal or less variability in terms of flux ranges compared to the integrated GEM, since some information will unavoidably be lost through the reduction process. In any case, all of the flux values in the solution space of redLips are a subset of the integrated GEM's solution space, as they would otherwise be inconsistent. Similarly, redLips has more essential enzymes than the integrated GEM, though the essential enzymes of the latter are all a subset of the former. This discrepancy can occur mainly because of two reasons: First, it is possible that some of the enzymes that are essential for redLips and non-essential for the GEM participate in lumped reactions, thus are indispensable for growth. Additionally, this means that these enzymes catalyze reactions that are present in all the computed alternatives for one (or more) BBB for the minimal subnetwork size. Second, some alternative pathways compensating for the loss of this enzymatic activity might be lost due to the reduction process, making it essential in the reduced model.

2.3.2 Overview of the reactions and metabolites in each expansion step

In order to get a clearer picture of redLips' structure and the overall network connectivity, we took a closer look at the reactions added to the model in each expansion step with respect to reactions that can carry flux. In the following discussion, the number of reactions comprising the graph search output will be given in parentheses next to the number of feasible (flux carrying) reaction additions. The starting subsystems include 540 reactions and 609 metabolites (307 unique ones

across compartments) in total. The complete list of subsystems included in redLips and the respective numbers of reactions in each of them, along with the number of reactions added in each expansion of the starting network are given in detail in Table 2.2.

In the D1 expansion of the model (one-step connections between core subsystems), most of the added reactions were transport reaction across compartments. The starting subsystems did not include any transport reactions, and all of the existing ones that connect metabolites belonging to the starting network were added at this stage since they are one step connections. Concerning the central carbon pathways, one reaction from the TCA cycle was included, namely the oxidation of succinate to fumarate and the reaction catalyzed by transaldolase from the pentose phosphate pathway. At this stage, a total of 199 (243) reactions were added to the core model along with 12 (22) new metabolites, of which 9 (17) were unique across compartments.

In the D2 expansion, a total of 44 (57) reactions were added to the D1 model. This seems like a significantly smaller number than in the previous step, though since the vast majority of the computed one-step reactions were transport reactions, this number is much larger than the enzymatic reactions that were added to the core during the D1 expansion. These reactions involve 35 (43) new metabolites, 30 (38) of which are unique, and include: the condensation of acetyl-CoA and oxaloacetate to form citrate in the cytosol and the peroxisomes (TCA cycle) and the reactions catalyzed by transketolase activity (pentose phosphate pathway).

The D3 expansion of the model encompassed 51 (71) additional reactions, including two more reactions from the pentose phosphate pathway. In terms of compounds, 57 (75) new metabolites were added, of which 56 (72) were unique across compartments. Lastly, a final graph search added the reactions in which only core metabolites participate and that had not already been added to the model in any expansion step, which were most commonly transport reactions for cofactors and boundary reactions. In our case, there were 250 (594) reactions that matched those criteria.

Table 2.2. List of subsystems included in redLips, and the corresponding number of reactions that were added in each step of the reduction process. Total number of reactions per subsystem and the percentage coverage of the corresponding integrated GEM subsystem is also reported. Boldface denotes the lipid pathway subsystems. FGS: Final Graph Search, PP: Post-Processing. (*)The biomass reaction representing cell growth is not part of either the starting network or any expansion step. (**)The lumped reactions are not part of the expansion steps, and they are computed and added to the model after D3 and before FGS.

Subsystem	Starting Network	D1	D2	D3	FGS	PP	Total # of reactions (% coverage of the integrated GEM)
Alanine and Aspartate Metabolism	-	-	4	-	-	-	4 (44.4%)
Alternate Carbon Metabolism	1	-	-	10	-	-	11 (40.7%)
Anapleurotic Reactions	-	5	-	-	-	-	7 (63.6%)
Arginine and Proline Metabolism	-	-	-	2	-	-	2 (6.1%)
Cardiolipin Biosynthesis	7	-	-	-	-	-	7 (100%)
Carnitine Shuttle	4	-	-	-	-	-	4 (100%)
Citric Acid Cycle	-	1	2	-	5	3	11 (84.6%)
Complex Alcohol Metabolism	-	-	-	2	-	-	2 (7.4%)
Cysteine Metabolism	-	-	-	1	2	-	3 (30%)
Dolichol Biosynthesis	30	-	-	-	-	-	30 (100%)
Fatty Acid Biosynthesis	67	-	-	-	-	-	67 (100%)
Fatty Acid Biosynthesis Mitochondrial	39	-	-	-	-	-	39 (100%)
Fatty Acid Degradation	99	-	-	-	-	-	99 (100%)
Fatty Acid Elongation	28	-	-	-	-	-	28 (100%)
Glutamate Metabolism	-	-	-	1	4	-	5 (29.4%)
Glutamine Metabolism	-	-	-	3	1	-	4 (100%)
Glycerolipid Metabolism	-	1	3	-	1	-	5 (55.6%)
Glycine and Serine Metabolism	-	-	1	5	1	1	8 (42.1%)
Glycolysis/Gluconeogenesis	12	-	4	3	1	-	20 (100%)
Glycoprotein Metabolism	-	-	2	1	-	-	3 (42.9%)
Histidine Metabolism	-	-	-	1	-	-	1 (7.1%)
Methane Metabolism	-	-	-	-	1	-	1 (50%)
Methionine Metabolism	-	-	-	1	3	-	4 (20%)
Mevalonate pathway	10	-	-	-	-	-	10 (100%)
NAD Biosynthesis	-	-	-	-	4	-	4 (16.7%)

Nucleotide Salvage Pathway	-	-	-	-	14	-	14 (16.9%)
Oxidative Phosphorylation	17	-	-	-	-	-	17 (89.5%)
Pentose Phosphate Pathway	-	1	2	2	-	-	5 (38.5%)
Phospholipid Biosynthesis	60	-	-	-	-	-	60 (100%)
Purine and Pyrimidine Biosynthesis	-	-	1	4	5	-	10 (19.2%)
Pyruvate Metabolism	8	1	2	1	-	-	12 (92.3%)
Riboflavin Metabolism	-	-	-	1	-	-	1 (7.1%)
Sphingolipid Biosynthesis	58	-	-	-	-	-	58 (100%)
Sterol Biosynthesis	31	-	-	-	-	-	31 (100%)
Sterol Metabolism	10	3	1	-	-	-	14 (100%)
TAG Decomposition	3	-	-	-	-	-	3 (100%)
Threonine and Lysine Metabolism	-	-	-	1	-	-	1 (5.3%)
Tyrosine, Tryptophan, and Phenylalanine Metabolism	-	-	1	4	1	-	6 (13.6%)
Valine, Leucine, and Isoleucine Metabolism	-	-	-	3	-	-	3 (15.8%)
Other	-	-	-	-	-	-	2 (20%)
Biomass Synthesis	35	-	-	-	-	-	36(*) (100%)
Lumped Reactions	-	-	-	-	-	-	37(**) (n/a)
Pooling Reactions	21	-	-	-	-	-	21 (100%)
Exchange Reactions	-	-	-	-	60	-	60 (36.4%)
Transport Reactions	-	187	21	4	148	-	360 (59.6%)

It is interesting to note that as we increased the user-defined degree of connection between the core subsystems, more amino acid biosynthetic routes were added. Also, some parts of the metabolism were located many steps away from our core network as we defined it, so ultimately were not added. One example of this was the TCA cycle that started to form for the D1 and D2 model expansions, though no new reactions were added in the D3 model. It therefore remained incomplete, missing three reactions to convert α -ketoglutarate to succinate through succinyl-CoA and one reaction to balance the intermediate byproducts. To ensure a more comprehensive

and consistent network, we included these four reactions in our model *a posteriori*. For the complete list of reactions of the model, the interested reader may refer to supplementary Table S2.3.

2.3.3 Generated lumped reactions

After the network expansion, we generated lumped reactions connecting the required BBBs to ensure their adequate production for the desired amount of growth. The biomass composition, as defined in iMM904, is comprised of 42 BBBs, and 14 of those could be sufficiently produced by our generated core network. Therefore, we needed to generate associated lumped reactions for 28 BBBs. As mentioned previously, lumpGEM computes the minimal set of reactions (called a subnetwork) that need to be added to the core network to produce a target BBB, which are then lumped into one reaction. For each of these subnetworks, all alternative subnetworks of the same size were also computed, to allow for flexibility of the network in terms of biosynthetic routes. In total, 38 lumped reactions were computed that corresponded to lumped subnetworks of various numbers of reactions. A detailed report on the number of generated lumped reactions per BBB and the size of the computed minimal subnetworks can be found in Table 2.3.

At this point of the workflow, we made several interesting observations about the ability of the core network to produce several BBBs. Even though their biosynthetic routes were explicitly present in the model, two BBBs, namely PC and ergosterol, could not be produced by the core network. This production was hindered by the lack of adenosyl-methionine, which the core could synthesize from adenosine triphosphate (ATP) and methionine by methionine adenosyltransferase, though methionine was another BBB that could not be produced by the core network. Using lumpGEM, we estimated the minimal set of reactions that we would need to add to the model to enable the production of methionine. Two alternative subnetworks were computed, each consisting of 11 reactions. It is noteworthy that in both subnetworks, the algorithm computes the most efficient methionine pathway to be the textbook biosynthetic route from aspartate. This additionally serves as an excellent validation point: the algorithm will always compute the most efficient biosynthetic pathways, which should be -and are- the physiologically observed ones. As seen in Figure 2.1,

this pathway converts aspartate to homoserine, followed by homocysteine, which will finally be converted to methionine (Mountain et al. 1991). The subnetworks also include methionine biosynthesis through sulfate assimilation (Thomas et al. 1992), with a few extra reactions included for mass balancing. The only difference between the two alternative subnetworks lies in the dehydrogenation of L-aspartate semialdehyde to homoserine; this reaction can use NADH or NADPH as a cofactor (model reactions HSDxi and HSDyi, respectively).

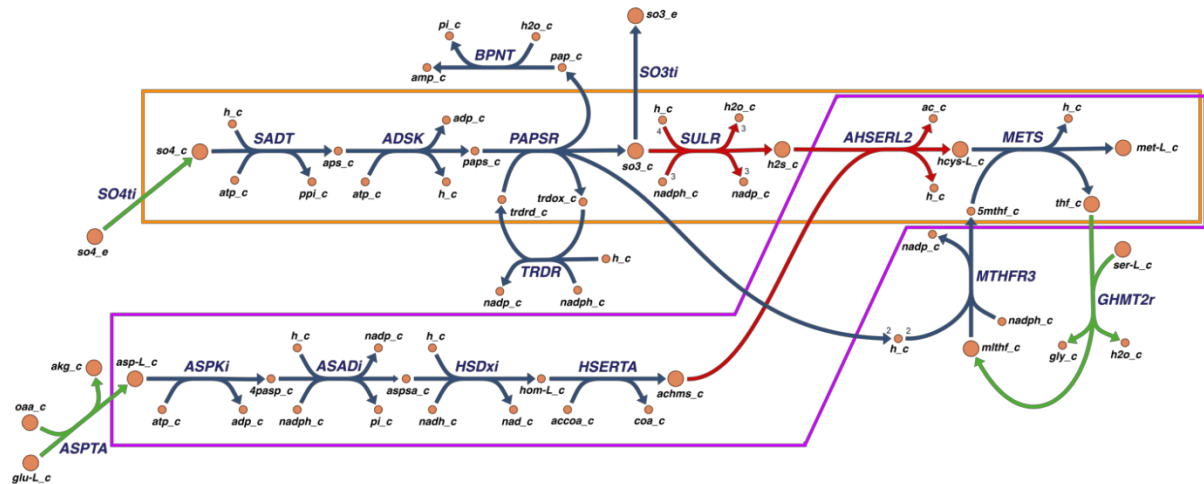


Figure 2.1. The L-methionine minimal subnetwork (in blue). The purple box highlights the textbook methionine biosynthetic route starting from aspartate. The orange box highlights the sulfate assimilation pathway for methionine biosynthesis. Reactions in red are part of the core network and part of the biosynthetic routes. Reactions in green are part of the core network but not part of the biosynthetic routes and serve the mass balancing of the subnetwork.

Interestingly, the lumped reaction computed for the synthesis of methionine was still insufficient for the production of ergosterol and PC because, in both of these pathways, adenosyl homocysteine was produced though not consumed by any other reaction of the core. Therefore, additional lumped reactions needed to be generated to remove this product and mass balance the two pathways. It just so happens that the minimal subnetworks required for both of these cases produced methionine and

were identical and unique. This subnetwork consisted of 4 reactions, which is considerably smaller size than the 11-reaction methionine subnetworks. Furthermore, since ergosterol and PC share the same subnetwork, the computed lumped reaction only needed to be added to the model once, resulting in the addition of 37 lumped reactions. Finally, these pathways can be observed graphically in Figure 2.2 and Figure 2.3. Figures 2.1, 2.2, and 2.3 were created using the ESCHER web application (King et al. 2015).

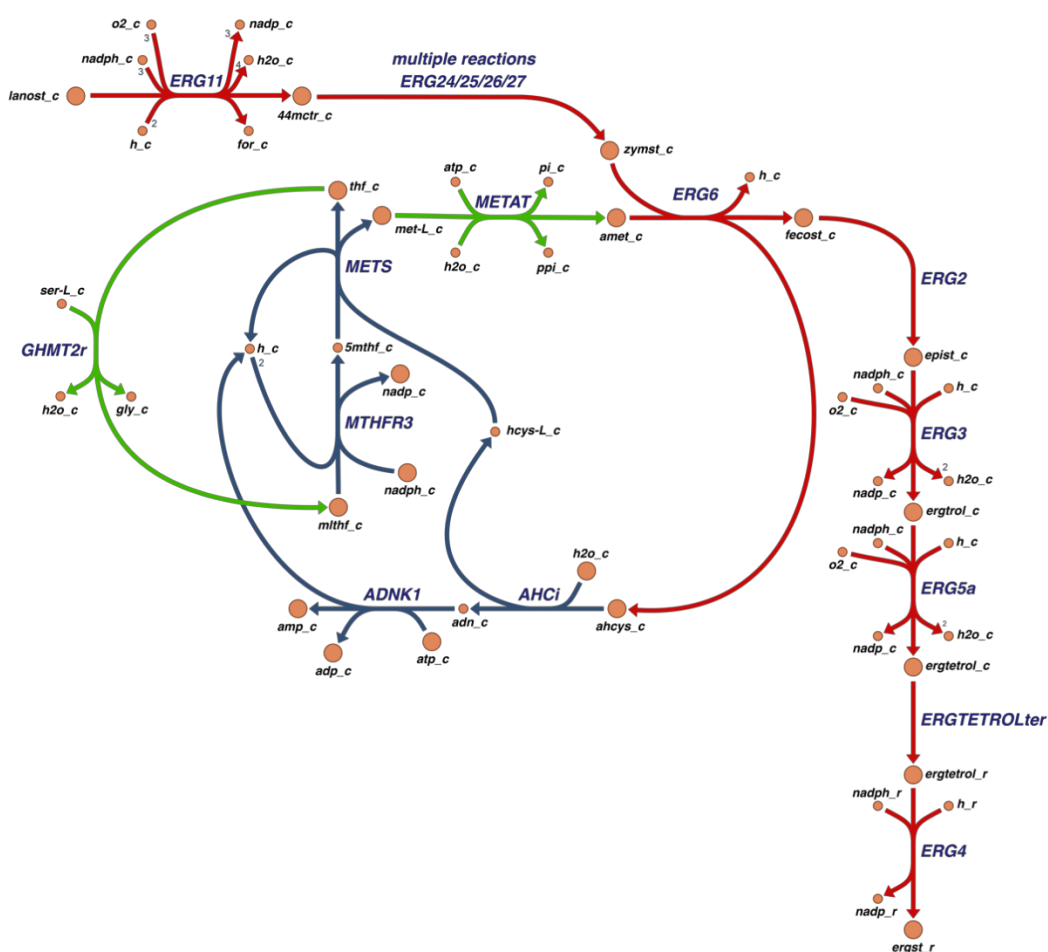


Figure 2.2. The ergosterol minimal subnetwork (in blue). Reactions in red are part of the core network and part of the biosynthetic route. Reactions in green are part of the core network but not part of the biosynthetic route, serving instead as the mass balance for the subnetwork.

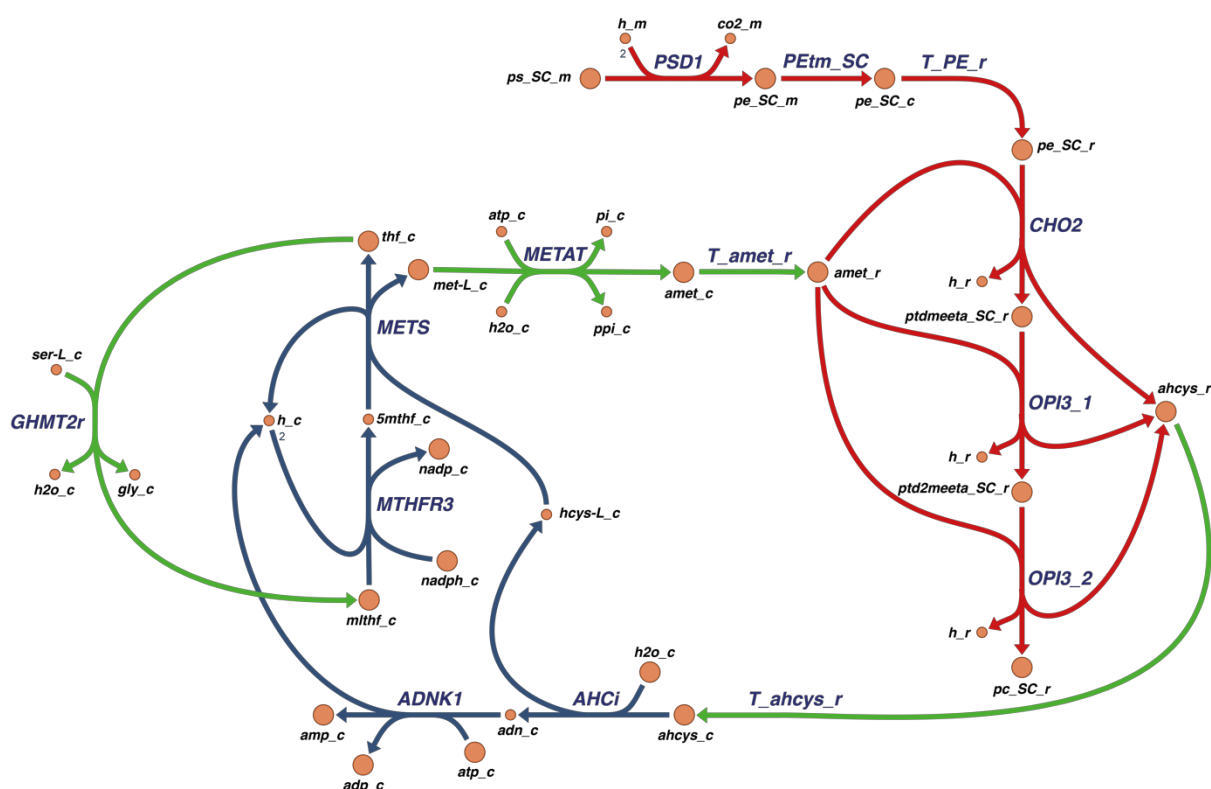


Figure 2.3. The phosphatidylcholine (PC) minimal subnetwork (in blue). Reactions in red are part of the core network and part of the biosynthetic route. Reactions in green are part of the core network but not part of the biosynthetic route, serving instead as the mass balance for the subnetwork.

2.3.4 Thermodynamics

The thermodynamic curation of redLips stemmed from the curation of the integrated GEM, as described in the Materials and Methods section. We used the same data to ensure that all reactions in the network are thermodynamically feasible, by imposing the relevant physiological constraints. The compounds whose properties could not be computed contain an acyl-carrier protein (ACP) molecule, which is a large and complicated molecule with a stereochemical structure that cannot be computed by GCM, as well as other related or bound species. The coverage of our database amounts to 89% of the metabolites of redLips, meaning 85.5% of the $\Delta_r G'^0$ values for the

network reactions could be computed. These computations included the relaxation of 62 thermodynamic constraints, as described in the Materials and Methods section. The complete thermodynamic curation data can be found in supplementary Table S2.3.

Table 2.3. Biomass building blocks for iMM904, the size of the subnetworks generated by lumpGEM, and the corresponding number of lumped reactions. (*)Produced by the core network.

Biomass Building Block	Size of Subnetwork	# of generated lumped reactions
1,3-beta-D-Glucan	3	1
AMP	10	1
L-Arginine	7	1
L-Asparagine	1	1
CMP	7	1
L-Cysteine	5	1
dAMP	13	1
dCMP	10	1
dGMP	15	2
dTMP	12	2
Ergosterol	4	1
Glycogen	3	1
GMP	12	1
L-Histidine	12	1
L-Isoleucine	9	4
L-Leucine	8	1
L-Lysine	8	2
L-Methionine	11	2
Phosphatidylcholine	4	1
L-Phenylalanine	7	1
L-Proline	4	2
Riboflavin	18	1
L-Threonine	5	2
Trehalose	2	1
L-Tryptophan	9	1
L-Tyrosine	7	2
UMP	6	1
L-Valine	3	1
Glycine	(*)	(*)
L-Alanine	(*)	(*)
L-Aspartate	(*)	(*)
L-Glutamate	(*)	(*)

L-Glutamine	(*)	(*)
L-Serine	(*)	(*)
Mannan	(*)	(*)
Phosphatidate	(*)	(*)
Sulfate	(*)	(*)
Phosphatidyl-1D-myo- inositol	(*)	(*)
Phosphatidylethanolamine	(*)	(*)
Phosphatidylserine	(*)	(*)
Triglyceride	(*)	(*)
Zymosterol	(*)	(*)

2.3.5 Gene Essentiality Analysis and Comparison

redLips was curated to include gene-reaction relationships in the form of logical rules. These rules were assigned through an exhaustive search in other yeast GEMs, majorly iIN800 (Nookaew et al. 2008) and iMM904 (Mo, Palsson and Herrgard 2009, Zomorodi and Maranas 2010), and in literature through the *Saccharomyces Genome Database* (SGD, <https://www.yeastgenome.org>, (Cherry et al. 1998)). Available experimental evidence for gene essentiality were gathered from literature through the Phenotype repository of SGD. All the genes whose deletions would result in inviability or auxotrophy beyond our defined media were classified as essential.

In order to benchmark and evaluate the performance of our model, we performed gene essentiality analysis for redLips and iMM904, for single-gene knockouts, and compared the results. The detailed predictions for each of the models are available in supplementary Table S2.4. redLips encompasses 459 genes opposed to 905 for iMM904. Out of these, 439 are common between the two models (Figure 2.4a). The 20 genes that are part of redLips and not iMM904 are all encoding enzymes catalyzing lipid related reactions.

redLips predicted correctly 50 genes as essential (true positive) and 372 genes as non-essential (true negative). Nine genes were predicted falsely as essential (false positive) and 28 as non-essential (false negative) (Figure 2.4b). Out of the 28 false negative predictions, one gene, namely YJL097W (PHS1) is not part of the iMM904

gene annotation. PHS1 encodes the enzyme that catalyzes the elongation of very long chain fatty acids, which are then used as building blocks for complex sphingolipids. The false negative prediction occurred because of the definition of biomass composition requirements in the model. Sphingolipids were not considered as BBBs, thus their formation, or lack thereof, does not affect the predicted growth. Moreover, among the rest of the false negative predictions we identified four genes, namely YBR265W (TSC10), YDL015C (TSC13), YKL004W (AUR1), and YMR296C (LCB1), which all are essential for sphingolipid production, either directly or through the metabolism of very long chain fatty acids. Similarly, the YMR013C (SEC59) gene was a false negative prediction because dolichol species were not considered in the biomass composition.

Interestingly, by imposing a minimum flux value on at least three of the defined biosynthetic reactions for lipid species (namely Bipc_a, Bmipc_a, and Bdolp), it was possible to attain the correct prediction of true positive for four out of five of these genes, with the exception of LCB1 (Figure 2.4c). The imposed flux value was equal to 10^{-6} mmol·gDW⁻¹·h⁻¹, which corresponds to the smallest BBB flux contributing to biomass in the network. This shows the significance and value of the addition of these reactions and highlights the importance of a consistently defined cell lipid composition. The results of this analysis when these fluxes were enforced are available in supplementary Figure S2.1.

When comparing the results of the two models for their common genes (Figure 2.4d), redLips performed better in predicting experimentally essential genes; iMM904 predicted 45 true positives which were a subset of the 50 predicted by redLips. The five additional genes encode enzymes which catalyze reactions belonging to lipid pathways, serine metabolism and glycolysis (Figure 2.4f). Correspondingly, the 27 false negative genes for redLips were a subset of iMM904's 32. Four of them could be turned into true positives if the sphingolipid requirements of the biomass were modified as mentioned previously. Other false negative genes included YDR208W (MSS4), YLR240W (VSP34), and YNL267W (PIK1), all related to PI synthesis. This part of the network is fairly complex and in possession of multiple alternative biosynthetic reaction routes. Furthermore, PI derivative species have been known to be especially active in signalling and membrane trafficking (Downes, Gray and Lucocq 2005, Krauss

and Haucke 2007), the mechanisms of which are either unknown or not included in the model.

redLips and iMM904 each predicted nine false positive genes, four of which were common. In redLips, these predictions occurred mostly due to alternative pathways missing from the network; while this was expected due to redLips being a reduced model, about half of these genes were false positives for iMM904 as well (Figure 2.4e). The rest of the false negatives stemmed from the inclusion of ergosterol in the biomass composition of the model. While ergosterol is essential to yeast cells, mutants incapable of synthesizing it are viable by accumulating ergosterol precursors in their membranes (Kato and Wickner 2001, Liu et al. 2017), an effect that was not included in either of the models.

In conclusion, the gene essentiality analysis and comparison of redLips opposite iMM904 showcases the ability of redLips to make accurate predictions, and in most cases performing better than the GEM. Genes that are not part of iMM904's annotation were included in the gene-reaction relationships of redLips, and the vast majority of genes that encode enzymes which catalyze lipid related reactions were predicted correctly as essential or non-essential.

2.3 Results and Discussion

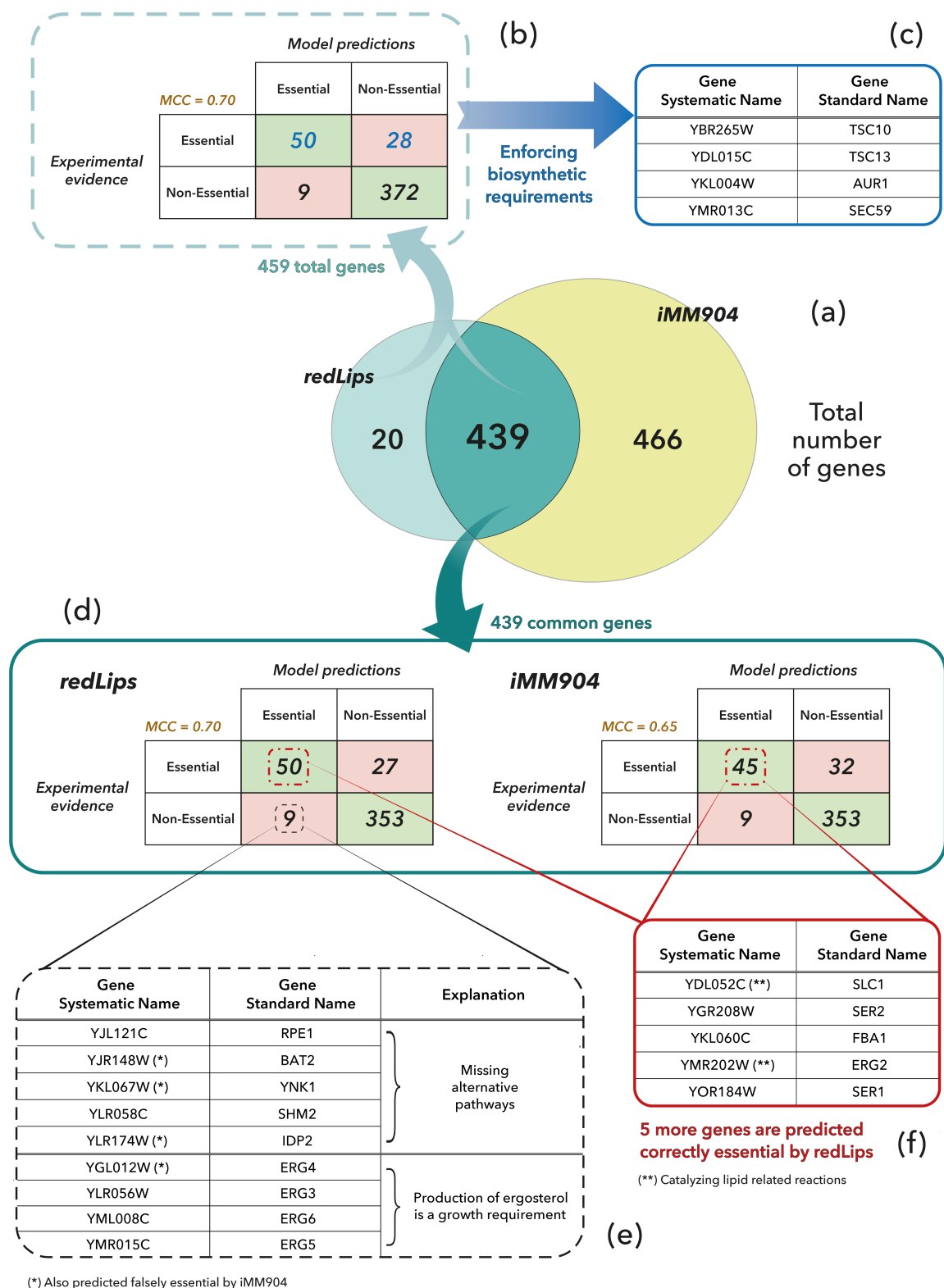


Figure 2.4. (a) Venn diagram of the genes included in redLips and iMM904. (b) Gene essentiality analysis in redLips and comparison with experimental evidence. (c) Improvements that can be made to the predictions by enforcing lipid biosynthetic requirements. (d) Gene essentiality analysis in both redLips and iMM904 for the enzymes they have in common; and comparison with experimental evidence. (e) The nine genes that correspond to false positive predictions of

redLips and explanations of the occurrence. (e) The five true positive predicted genes of redLips that iMM904 predicts falsely negative. The Matthews Correlation Coefficient (MCC)(Matthews 1975) is also reported for each case.

2.3.6 Comparison with yeast GEMs

Comparing redLips with the available yeast GEMs provided insight into the network in terms of comprehensiveness and pathway connectivity. As mentioned previously, our model includes all the major lipid species and respective biosynthetic pathways.

One common difference between models that can lead to very dissimilar numbers in reactions and species is the considered fatty acyl chains. As stated in the Materials and Methods section, glycerophospholipids possess two fatty acyl chains (four in the case of CL). In our model, we defined a fatty acid pool that participates in the formation of these species, thus considering only one metabolite of each class with an attached, generic fatty acyl (the assumed length is still C16:0 for thermodynamic calculations). Some other GEMs, namely Yeast7 and Yeast8, consider four individual fatty acid species as reactants in these biotransformations: C16:0, C18:0, C16:1, and C18:1. Naturally, this leads to a very large combinatorial number of reactions and species. For example, if we consider a species with two fatty acyl tails, there are ten possible combinations leading to ten model metabolites. This number can grow exponentially if one considers the large number of fatty acyl derivatives and remodeling reactions. The same issue arises with fatty acyl-CoAs. Since the aim of redLips is to be used as a scaffold for omics integration and its nature is not context limited to specific species or pathways in order to enable versatility, this is a pitfall we aimed to avoid in order to preserve a concise representation. As mentioned earlier, a model including all the combinatoric occurrences will be difficult to curate and handle. It is important to note that, in the case of available experimental data or focused studies, separately considering these species can be beneficial for the accuracy of predictions (Sanchez et al. 2019), and should be taken into account by expanding the associated parts of the network species and reactions accordingly.

Another difference between redLips and other models is that some other models, for example iIN800 (Nookaew et al. 2008), include multiple identical reactions if their respective associated enzymes are encoded by multiple genes or in cases of multiple enzyme paralogs. This practice is acceptable according to genome annotation, but it leads to misleading computations; the mechanistic representation of the network calculates the net flux through each reaction irrespective of which enzyme is catalyzing it. This means that the resulting net flux value for this particular reaction will be the sum of all the discrete flux values for each reaction copy. To resolve this point, we considered only unique reaction occurrences in redLips that represent the net reaction rate for each biotransformation. The exception to this rule is the case where the same biotransformation occurs in different cellular compartments. Since the metabolites in each compartment are modeled separately, this consideration does not result in duplicate reactions—in mathematical terms, the stoichiometric matrix will not have duplicate columns. In circumstances where enzymatic or kinetic properties are relevant for a study and require a separate consideration for these instances, the model can simply be modified to incorporate them.

We also present a detailed comparison of our network to the other yeast networks in Table 2.4. Included in this table are the number of lipid-related reactions, species, and cellular compartments considered in each model. To ensure accurate comparability, we curated the number of reactions and species of interest for all considered models. The criteria we used were as follows: (i) We considered only one generic instance of metabolites possessing one or more fatty acyl chains. This applies both to species and reactions. As discussed previously, each model considers a different number of fatty acyl chain lengths, and in combination, this can lead to misleadingly different statistics. (ii) We didn't consider metabolite pools or pooling reactions. Similarly, each model considers various diverse metabolite pools that can be heavily connected to the network by a large number of pooling reactions. (iii) We didn't consider duplicate reactions unless they occurred in different compartments. (iv) We didn't consider transport and boundary reactions. Since each model considers a different number of cellular compartments, the number of transport reactions varies accordingly. (v) We didn't consider disconnected reactions; there were rare occurrences of reactions in which both reactants and products did not participate in any other reaction in the

network. These reactions serve for annotation purposes and most probably will be gap-filled in the future, but they do not contribute to the functionality of the model.

Using this comparison, we can see in Table 2.4 that redLips covers at least as many species as the other GEMs and more reactions than most of them. The major differences in the non-curated numbers of species and reactions can be attributed to the reasons listed above as well as the number of compartments of each model. If we go through the reactions per pathway, we can see that the majority of differences stem from the biosynthetic routes for PI derivatives, such as glycosyl-phosphatidylinositol (GPI) anchors for proteins and inositol and PI polyphosphates. These molecules play a major role in cell signaling, which was beyond the scope of redLips at this time. Signaling cascades in lipid metabolism is a vast area of study on its own, and we feel that it would be best served with a dedicated model. Another difference, especially concerning the Yeast7 and Yeast8 models, was in phospholipid biosynthesis. The larger numbers in these models are due to the consideration of five additional cellular compartments, including membranes, and the assignment of reactions occurring in more than one of them.

Table 2.4. Detailed comparison between redLips and other yeast GEMs in terms of reactions, species, and cellular compartments. The curated numbers of lipid reactions and species are given in parentheses next to the non-curated numbers if these numbers differ.

Model:	redLips	iMM904 (Mo, Palsson and Herrgard 2009)	iIN800 (Heavner and Price 2015, Nookaew et al. 2008)	iTO977 (Heavner and Price 2015, Osterlund et al. 2013)	Yeast7 (Aung, Henry and Walker 2013)	Yeast8 (Lu et al. 2019)
# of lipid reactions (Curated # of lipid reactions)	481 (451)	291 (264)	348 (281)	331 (324)	1758 (442)	1787 (460)
Cardiolipin Biosynthesis	7	4	2	3	270 (4)	270 (4)
Carnitine Shuttle	4	3	1	1	3	3
Fatty Acid Biosynthesis	67	33	71 (63)	64	25	25
Fatty Acid Biosynthesis Mitochondrial	39	13	34	34	15	15
Fatty Acid Elongation	28	6	33 (28)	28	30	30
Fatty Acid Degradation	99	51	46 (43)	47	100	100
Glycerolipid Metabolism	7	9	15 (14)	7	29 (14)	31 (16)
Glycoprotein Metabolism	5	7	15 (4)	4	3	3
GPI Biosynthesis	0	0	0	8	9	19
Isoprenoid Biosynthesis	19	1	0	0	22	22
Mevalonate Pathway	14	14	13 (12)	11	14	14
Phospholipid Biosynthesis	60	50	51 (35)	44	686 (78)	690 (82)
PI Signaling System	0	0	0	8	15	15
Sphingolipid Biosynthesis	58	63 (36)	36 (25)	25	257 (63)	257 (63)
Sterol Metabolism	41	36	22 (19)	25 (23)	43 (32)	43 (32)
TAG Decomposition	3	1	4 (1)	1	48 (4)	48 (4)
Other	0	0	0	15	11	13
Pooling Reactions	30	0	5	5	178	189
# of lipid species (Curated # of lipid species)	241 (237)	156 (143)	184 (178)	231 (220)	500 (233)	523 (237)
# of involved compartments	7 + ex	7 + ex	2 + ex	3 + ex	12 + ex	12 + ex

2.4 Conclusions

In conclusion, redLips is a metabolic model that captures the complexity of lipid metabolism by preserving and uniting the vast majority of known lipid reactions and pathways while avoiding the pitfall of excessive—and often times redundant—detail. It was created by gathering, merging, and upgrading existing lipid metabolic pathways, integrating them into the iMM904 GEM of *S. cerevisiae* and subsequently reducing this model around the major lipid-related subsystems using the redGEM and lumpGEM frameworks. Additionally, it is consistent with the organism biochemistry as well as thermodynamic principles and can be further constrained through lipidomics measurements, applied both as flux and concentration bounds. redLips can be used as a concise platform for studying lipid metabolism across different species, and is a valuable tool for health or industry related research. We believe that this model will continue to accommodate future discoveries through the incorporation of new reactions and species as well as providing a coherent base to link cell signaling routes and building kinetic models.

Chapter 3 – Exploring the effect of enzymatic coupling on kinetic modeling and metabolic control across the lipid pathways of yeast

3.1 Introduction

3.1.1 Stoichiometric Modeling of Metabolism

Over the past decades, computational models of metabolism have emerged as a valuable tool for hypothesis assessment and strategy design. These models are based on a mathematical representation of connectivity in the represented network, associating reactions with metabolites. This representation, called the stoichiometric matrix, stores the information of all the metabolites and their stoichiometric participation in each reaction of the network. For the study of these models, tools like Flux Balance Analysis (FBA) (Orth, Thiele and Palsson 2010) are well-established. FBA is a constraint-based method that allows the calculation of the steady state flux distribution throughout the metabolic network through an optimization algorithm. However, the resulting solution is not unique, but rather infinite solutions exist within a solution space which is delimited by the model's bounds and constraints. A consistent way to reduce the allowable solution space of the problem and to enforce physiological limitations is with the introduction of thermodynamic feasibility constraints. The three main approaches that have been used to this end are the Energy Balance Analysis (EBA) (Beard, Liang and Qian 2002), the Network-Embedded Thermodynamic analysis (NET analysis) (Zamboni, Kümmel and Heinemann 2008),

and the Thermodynamics-based flux analysis (TFA) (Salvy et al. 2019), as reviewed thoroughly by Ataman and Hatzimanikatis (2015).

Nevertheless, stoichiometric models are not able to capture dynamic metabolic responses to perturbations in cellular and process parameters, nor they can consider regulation at the enzyme and post-translational levels (Miskovic et al. 2015). These studies are vital for the elucidation of the mechanisms of cellular regulation and membrane homeostasis. Kinetic models are necessary in order to investigate these subjects.

3.1.2 Kinetic Modeling of Metabolism

Kinetic models are an essential tool for the study of cellular metabolism and for understanding the dynamics of its regulation mechanisms. However, the assignment of appropriate rate expressions for the reactions in the network is challenging, since enzymatic mechanisms are seldom fully characterized. Moreover, information about the parameter values in these rate expressions is scarce, rendering the construction of large-scale kinetic models a difficult endeavor (Miskovic et al. 2015). To overcome this uncertainty, the majority of studies have used simple mass-action kinetic rate expressions to describe the reactions in the network. However, it has been shown that properties such as the reaction displacement from thermodynamic equilibrium and the enzyme saturation need to be taken into account for accurate predictions (Chakrabarti et al. 2013, Soh, Miskovic and Hatzimanikatis 2012). The Optimization and Risk Analysis of Complex Living Entities (ORACLE) framework (Miskovic and Hatzimanikatis 2010) accounts for this uncertainty by incorporating stoichiometric, thermodynamic and physiological constraints *a priori* to applying kinetic data to the model and produces populations of thermodynamically feasible kinetic models by sampling the kinetic parameter space (Wang, Birol and Hatzimanikatis 2004, Wang and Hatzimanikatis 2006a, Wang and Hatzimanikatis 2006b).

In most metabolic networks, groups of compounds are conserved throughout the network and are thus subject to conservation constraints. The total concentration of each of these groups, which are called conserved moieties, remains constant over time even though the individual concentrations might change. In a biological sense, the conserved moieties are transferred from one biomolecule to another inside the

metabolic network, but are not synthesized or degraded in this network, or exchanged with the environment. In a mathematical sense, a conserved moiety in a metabolic model introduces linearly dependent rows in the stoichiometric matrix, since the summation of all the components' mass balances is invariant over time and thus constant and equal to zero. Additionally, the concentration levels of moieties have been reported to affect the metabolic control over fluxes and concentrations within the network, by affecting the thermodynamic equilibrium displacements of reactions (Hameri et al. 2019, Kholodenko, Sauro and Westerhoff 1994, Wang, Birol and Hatzimanikatis 2004). Consequently, it is important to identify the existence and properties of moieties in the metabolic network before proceeding to the kinetic parametrization of the problem.

3.1.3 Lipid Regulation

Lipids are a diverse group of biomolecules that serve multiple purposes in an organism. They comprise the major components of cellular membranes, act as energy storage sources and participate in signaling cascades. Lipid metabolic pathways need to be tightly regulated in order to maintain a physiologically healthy operation of cellular processes. Lipid metabolism is responsible for providing the required lipids in the correct proportions for the correct function of cell membranes and be able to adapt their production levels in response to external stimuli. Imbalance of lipid species can lead to severe metabolic disorders including, but not limited to, Parkinson's and Alzheimer's disease (Kosicek and Hecimovic 2013, Santiago and Potashkin 2013), as well as diabetes (Markgraf, Al-Hasani and Lehr 2016, Santiago and Potashkin 2013) and liver steatosis (Bradbury 2006). Cancer cells also have an atypically active lipid metabolism that allows their rapid proliferation (Alves et al. 2016, Ogretmen and Hannun 2004, Vriens et al. 2019).

Kinetic models of the lipid metabolism have been around for years (Farquhar et al. 1965). Studies on this subject have employed both deterministic and stochastic approaches to the problem in an effort to understand the mechanisms of lipid regulation in both mammalian (Bhattacharya et al. 2014, Hubner et al. 2008, Knoblauch et al. 2000, Watterson et al. 2013) and yeast cells (Alvarez-Vasquez et al.

2011, Alvarez-Vasquez et al. 2005, Alvarez-Vasquez et al. 2004, Alvarez-Vasquez et al. 2007, Savoglidis et al. 2016, van Eunen et al. 2013). Although multiple efforts have been made to accurately capture the dynamic properties of the lipid metabolic network, the vast majority of studies are context-specific and context-limited. This means that their networks are based on *ad hoc* reductions of genome-scale metabolic models (GEMs) and their parameters were fitted to specific datasets and thus cannot serve multiple studies.

3.1.4 Metabolic Control Analysis

Sensitivity analysis has successfully been applied in multiple engineering fields and can be expanded to accommodate systems biology (Hatzimanikatis and Bailey 1996). Metabolic control analysis (MCA) (Kacser, Burns and Fell 1995) is a widely used tool that has facilitated the study of the local sensitivity of metabolic networks to perturbations near a steady state (Fell and Sauro 1985, Reder 1988b), and of cellular regulation (Hatzimanikatis and Bailey 1997, Heinrich and Rapoport 1974). MCA quantifies the magnitude to which a change in a system parameter will affect a system variable, which in our case translates into the effect that altering an enzyme activity will have on reaction rates or metabolic concentrations in reference to a given steady state. As a local sensitivity analysis tool, the limitation of this technique is that it can only be applied to estimate responses to small perturbations in biological systems.

3.1.5 Aim and Contribution

In conclusion, the main obstacle in the construction of consistent kinetic models is the scarcity of well-characterized enzymatic mechanisms and their parameter values, which leads to uncertainty in the model's predictions. This uncertainty is amplified for large-scale metabolic networks, which is why most studies consider only small parts of them. However, the inconsistent reduction of metabolic networks introduces more uncertainty in the models, and can lead to unrealistic predictions. To address these issues, we sought to develop a large-scale kinetic model that focuses on the lipid metabolic pathways. We used the redLips model (Chapter 2) as a basis for our study, which was constructed through a systematic reduction around the lipid network of *S.*

cerevisiae. redLips already includes thermodynamic feasibility constraints, which are essential for the calculation of metabolite concentration levels and reaction displacement from thermodynamic equilibrium, and is an excellent scaffold for the integration of experimental data. Since metabolic models can represent multiple physiologies within their flux and concentration solution space, a representative profile needs to be selected prior to building a kinetic model. Consequently, the resultant kinetic model will be able to describe dynamic responses to perturbations around the selected profile. Thus, to ensure consistency with the organism's physiology, we first analyzed the diverse physiological states that our model can describe, in terms of metabolite concentrations and metabolic fluxes. Subsequently, we selected a representative profile in terms of reaction fluxes and metabolite concentrations around which the kinetic model could be built. In order to accurately capture the enzymatic coupling effect that arises from the promiscuous and multifunctional enzymes which are typical in lipid metabolism, we used the notion of apparent inhibitors when devising kinetic rate laws for the reactions in the network. We generated populations of kinetic parameter sets through the sampling-based ORACLE workflow and demonstrated how the consideration of enzymatic coupling is essential for factual predictions. The kinetic version of the redLips model along with the statistically analyzed sets of kinetic parameters can be used to perform dynamic simulations and to facilitate the discovery and design of metabolic engineering strategies. Furthermore, to our knowledge, it is the largest and most detailed kinetic model of the lipid metabolism to date.

3.2 Materials and Methods

3.2.1 Kinetic modeling techniques and metabolic control

Within the MCA method two important variables are calculated: flux control coefficients (FCCs) and concentration control coefficients (CCCs). These coefficients are defined as the fold change of metabolic fluxes and metabolite concentrations, respectively, in response to changes of system parameters, usually enzyme activities.

The elasticities with respect to a parameter or variable are also computed. While CCs are global quantities, elasticities are local quantities and express the sensitivity of a reaction to a change in the respective parameter or variable. The complete derivation of this method and the corresponding mathematical expression can be found in (Wang, Birol and Hatzimanikatis 2004).

The control coefficients are defined as follows:

- Flux control coefficient (FCC): $C_p^v = \frac{d \ln v}{d \ln p} = \frac{p}{v} \frac{dv}{dp} = \frac{p \Delta v}{v \Delta p}$
- Concentration control coefficient (CCC): $C_p^x = \frac{d \ln x}{d \ln p} = \frac{p}{x} \frac{dx}{dp} = \frac{p \Delta x}{x \Delta p}$

In all cases, concentrations are represented by x , parameters by p and fluxes by v .

As mentioned previously, there exists a certain degree of uncertainty associated with kinetic models, which is propagated through the workflow and ultimately translated into uncertainty in the control coefficients. This is why the generation of a very large number of kinetic models is essential, and any uncertainty can then be statistically quantified and analyzed. ORACLE employs smart Monte Carlo sampling techniques in order to reduce this uncertainty and allow for the definition of more complex kinetic mechanisms, while introducing the minimum possible amount of bias in the calculations (Miskovic et al. 2019b). The key steps of the workflow are the following (Miskovic et al. 2015): Firstly, the stoichiometry of the metabolic network is defined, and constraints and data stemming from fluxomics and metabolomics are added if available. The model is then curated by introducing thermodynamic feasibility constraints. FBA and TFA analyses are performed to compute thermodynamically feasible flux distribution profiles for the system, and representative physiological steady states are identified. Next, metabolite concentration levels are sampled according to the chosen physiology. The displacement of enzymatic reactions from thermodynamic equilibrium is also computed based on these concentrations. For each reaction, a kinetic mechanism is assigned based on available literature, and experimental values of kinetic parameters can be added. Then, the enzymatic degrees

of saturation are sampled, and populations of thermodynamically and stoichiometrically feasible kinetic models are parameterized consistently to the sampled concentration and flux profiles. Finally, each resulting kinetic model is evaluated on stability. Experimental or literature data can also be used to validate these kinetic models. The ones that are unstable or inconsistent with data are rejected in this pruning phase, while the ones that pass this test can be used for dynamic simulations and MCA analysis. Control coefficients are calculated and perturbation analysis of the system is performed. These results can be analyzed to formulate hypotheses for the design of metabolic engineering strategies.

3.2.2 The effect of enzymatic coupling on Control Coefficients

Promiscuous enzymes catalyze multiple reactions in the network and give rise to enzymatic coupling effects. Enzymatic coupling requires an appropriate reformulation of the parameter elasticity matrix, as defined above. In the absence of the enzymatic coupling effect, the parameter elasticity matrix Π is an identity matrix, since for every reaction there exists only one corresponding enzymatic activity, and the elasticity matrix element with respect to this activity will be equal to unity. However, in the case of enzymes that catalyze multiple reactions in the metabolic network, changes in the activity of one of those enzymes will have an equal impact on the maximal reaction rate values of all the reactions that it catalyzes. Therefore, for each of those enzymes, a single estimated change needs to be calculated, accounting for all the individual predicted changes for each catalyzed reaction. Consequently, a structural change in the Π matrix needs to be introduced (Savoglidis et al. 2016), reducing its dimension of enzymatic activities appropriately, which will in turn result in a reduction of equal magnitude in the dimensions of the calculated CC matrices.

3.2.3 The updated redLips network

For this study we used the redLips model that was described in Chapter 2, which focuses on the lipid metabolic network of the yeast *S. cerevisiae*. The model was slightly modified to facilitate the assignment of kinetic mechanisms to each reaction in the network. Firstly, we updated the elementary biosynthetic drains of lipid

species, which were modeled as exchange reactions, to occur in two steps: initially the biomolecule is transported from the cytosol to the extracellular domain and subsequently from the extracellular domain across the system boundary. Secondly, we allowed a number of metabolites to be able to diffuse outside of the cell by adding the appropriate transport reactions. These reactions are permitted to carry a maximum flux value of 10^{-6} mmol·gDW⁻¹·h⁻¹, representing a small diffusion rate that occurs naturally between the cell and its surroundings. Not all metabolites can diffuse through membranes without dedicated transporters, and according to this fact, the criteria we used for the choice of these compounds are the following: (i) the metabolite cannot be phosphorylated, (ii) its structure cannot be complex, (iii) it doesn't already have an assigned transport to the extracellular domain. Additionally, the added reactions for each species must be able to carry flux. This led to the addition of transport reactions for 52 species. Finally, the resulting model encompasses 1275 reactions and 890 metabolites and was curated to include thermodynamic feasibility constraints for all the added reactions. This curation led to 90.1% coverage in terms of metabolite thermodynamic properties and resulted in the estimation of the associated change in Gibbs free energy for 86.8% of the network's reactions.

3.2.4 Experimental data integration

In order to integrate experimental measurements to our network, and reduce the solution space of the problem, we introduced sets of constraints that allow us to impose bounds on both fluxes and concentrations. To this end, we used lipidomics measurements provided by the Riezman lab (University of Geneva) that corresponded to 28 lipid species. These data can be found in supplementary Table S3.1.

For the fluxes, we used the equations described in the Materials and Methods section of Chapter 2, to effectively couple the biomass production rate to the flux through each elementary lipid biosynthetic reaction as follows:

$$\mu(\bar{Y}_i - sd) \leq v_i \leq \mu(\bar{Y}_i + sd),$$

where v_i are the biosynthetic reaction fluxes, \bar{Y}_i the mean of the lipidomic content measurements, sd the experimental measurements' standard deviation of \bar{Y}_i , and μ

the specific growth rate of the cell as calculated from the flux through the biomass objective function.

For the concentrations, we were faced with a common – yet underdiscussed – issue, which is the distribution of species within cellular compartments. Since concentrations of lipids are measured after cell lysis there is no available data on the intracellular allocation of each molecule. To address this issue, we used the sets of constraints proposed by Tokic, Miskovic and Hatzimanikatis (2019) to impose bounds on each species per cellular compartment while maintaining consistency with the experimental measurements. These constraints are formulated as follows.

For the concentration C_M of a metabolite M measured in the range $C_M \in [\underline{C}_M, \overline{C}_M]$ we have:

$$\ln(C_M) \geq \sum a_i \ln(C_{Mi})$$

where C_{Mi} is the concentration of the metabolite M in compartment i, and a_i is the volume fraction of compartment i with respect to the entire cell.

The concentration C_M will clearly also be subject to the bounds:

$$\ln(\underline{C}_M) \leq \ln(C_M) \leq \ln(\overline{C}_M)$$

Accordingly, each concentration C_{Mi} will be subject to the bounds:

$$\ln(\underline{C}_{Mi}) \leq \ln(C_{Mi}) \leq \ln(\overline{C}_{Mi})$$

\underline{C}_{Mi} and \overline{C}_{Mi} are calculated as:

$$\underline{C}_{Mi} = \max \left[\frac{\underline{C}_M + (a_i - 1) * UB}{a_i}, LB \right]$$

$$\overline{C}_{Mi} = \min \left[\frac{\overline{C}_M + (a_i - 1) * LB}{a_i}, UB \right]$$

where LB and UB are the physiological lower and upper bound on intracellular metabolite concentrations, with values $10^{-7}M$ and $10^{-2}M$, respectively.

Since the provided lipidomics measurements were given in nM units, we used a specific cell volume value of $2 \text{ mL}\cdot\text{gDW}^{-1}$ as reported by Van Urk et al. (1988) to convert them to $\text{mmol}\cdot\text{gDW}^{-1}$ units, which multiplied by the specific growth rate yield the units considered for reaction rates in our model, namely $\text{mL}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$. The metabolite concentration in our network are modeled in M, thus no conversion was required.

3.2.5 Sampling the solution space

As the mathematical problem still remains underdetermined after these curation steps, there is a large solution space of thermodynamically feasible states that needs to be explored in order to confidently choose a representative flux and concentration reference steady state around which we can build populations of kinetic models. To this end, we sampled the solution space using an artificially centered hit and run (ACHR) sampling algorithm. We generated 50'000 samples for the fluxes and concentrations each, and then we selected the sample closest to the center of the sampled space as our representative state for both cases.

3.3 Results and Discussion

3.3.1 Bidirectional Reactions and Flux Directionality Profiles

An important aspect to be considered in every metabolic model is that of the physiology of the studied organism. Especially for building kinetic models, a steady state that represents well the observed physiology needs to be selected prior to the generation of the kinetic parameters. To this end, we investigated the model's feasible steady state flux distributions, by studying the flexibility of the network in terms of flux variability and bidirectional reactions (BDRs). Most of the enzymes catalyzing each reaction can operate in both directions, rendering the vast majority of the

reactions kinetically reversible. However, even if a reaction is kinetically reversible, it is not a necessity that it is also bidirectional. In reality, there will always be both a forward and reverse flux, but in terms of net fluxes many reactions will only operate towards a certain direction in physiological conditions. The FBA formulation usually predicts that most reactions can carry flux in both directions, but the TFA formulation greatly decreases their number since thermodynamic feasibility constraints are imposed.

Specific to redLips, 121 reactions were bidirectional after the thermodynamic curation of the model. The vast majority of these reactions consisted of transports between cellular compartments, specifically 76 transport and 2 boundary reactions. The remainder of the BDRs belonged to the TCA, oxidative phosphorylation, glycolysis and gluconeogenesis, the carnitine shuttle, and to a smaller degree to the pentose phosphate pathway, nucleotide salvage pathway, pyruvate and amino acid biosynthesis. A big number of BDRs leads to an astronomical amount of combinations for the steady-state flux distribution of the network, or flux directionality profiles (FDPs). Theoretically, n BDRs correspond to 2^n FDPs. Hence, it was crucial to determine which of these reactions can be further constrained according to biological information, and ultimately to choose one prominent and physiologically representative FDP.

As a first step we constrained reactions for which evidence of a certain operational directionality exist. Specifically, we made the following assumptions: (i) glycolysis operates towards the production of pyruvate, (ii) the TCA cycle operates as a full cycle, and (iii) the pentose phosphate pathway (PPP) operates according to the direction of glycolysis, from fructose 6-phosphate towards glyceraldehyde 3-phosphate. We further constrained a number of transport reactions for which evidence of directionality exists, such as nucleotide transports. In total, we imposed directionalities on 38 reactions, which lead to the constraint of 41 reactions to operate in one direction. Next, we constrained the biomass production to be at least 99% of the maximal computed production, and performed thermodynamic flux variability analysis (TVA) to examine how the flexibility of the network was altered. 8 reactions became unidirectional for this case. Subsequently, we made the assumption that the cell will always strive for minimal enzyme usage, which translates to a minimal sum

of fluxes through the network. The objective of minimizing the sum of fluxes is also known as parsimonious approach (Lewis et al. 2010), and applying it to the network lead to 17 additional reactions assuming a directionality. Ultimately, the number of BDRs was reduced to 55, mostly consisting of species transports, for which no relevant assumptions could be made. Theoretically, 55 BDRs would lead to up to 2^{55} FDPs. This number of possible FDPs is enormous and further expert knowledge and experimental data are required in order to further reduce the number of BDRs. Moreover, even after reducing more this number, further analysis of alternative FDPs should be performed. Here, we have chosen one of these FDPs because it satisfies the available data, physio-chemical constraints, and our assumptions. The complete list of bidirectional reactions and the step by step determination of their respective directionalities can be found in supplementary Table S3.2.

It is interesting to note that if we do not predefine any directionalities before applying the maximal biomass demand and parsimonious TFA, most reactions considered in our assumptions will present the same directionality as when enforcing those assumptions. For example, glycolysis will be operating towards the production of pyruvate; the only reaction retaining its bidirectionality in the glycolysis pathway will be the one catalyzed by glucose-6-phosphate isomerase (PGI), a reaction that has been known to be capable of operating in both forward and reverse directions.

3.3.2 Conserved moieties

In the redLips network we identified 17 conserved moieties. These include well reported and expected groups, such as the pyrimidine nucleotides, other nucleotide-decomposed molecules such as quinones and quinols, cytochromes, and carnitine, ACP and Coenzyme A (CoA) bound species among others. Most of the moieties consist of two or three metabolites, while the two largest groups consist of 85 and 129 species, for the ACP and CoA bound moieties respectively. The full list of conserved moieties can be found in supplementary Table S3.3.

3.3.3 Kinetic mechanisms – Apparent inhibition through competing substrates

One of the most important steps in building a kinetic model is the assignment of mechanisms that accurately describe the enzyme kinetics for each reaction. Lipid associated enzymes are multifunctional and promiscuous, which means that they are capable of catalyzing multiple reactions in the network and are able to bind several different species (Carbonell, Lecointre and Faulon 2011, Tidhar and Futerman 2013). In the traditional inhibitory action, the inhibitor species competitively bind to the active sites of the enzyme, but no biochemical transformation takes place. In our instance, this is not the case; the multiple alternative substrates that the identified enzymes can accept in its catalytic sites bind to the enzyme and form complexes that lead to product formation. This mechanism is still a form of competitive inhibition but requires a special form of a kinetic rate expression to describe each of the individual reactions. To this end, we used the notion of apparent inhibitors and the derived rate expressions as defined by Savoglidis et al. (2016).

For the case of our model we identified 53 enzymes that catalyze a total of 346 reactions in the redLips network. Within these enzymes we identified four distinct physiological possibilities that may – or may not – give rise to apparent inhibition and are worth discussing:

- The first case (case I) is the simplest case of substrate competition; for example, the enzyme encoded by SCS7, exhibiting sphingolipid alpha-hydroxylase activity. In this instance we have one enzyme catalyzing the alpha hydroxylation of sphingolipid-associated very long chain fatty acids, which corresponds to three reactions all occurring in the endoplasmic reticulum. Thus, the substrates of these enzymes are competing for the same binding site of the enzyme and will act as apparent inhibitors to one another. Similarly, in the case of isoenzymes we consider one reaction group that encompasses all of them and their catalyzed reactions, since the reactions' substrates will be competing for the active sites of both these enzymes simultaneously, analogously to case I. There is however one exception to this rule which relates to selectivity, or binding affinity; the enzyme might be able to bind multiple species but be strongly selective towards one of them, as is the case with the

example of the three elongases, ELO1, ELO2 and ELO3. For simplicity, in such circumstance we will consider each of the enzymes separately as multiples of case I.

- The second case (case II) can be observed when an enzyme is multifunctional; for example, the enzyme encoded by FOX2, exhibiting both 3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase activities. This enzyme catalyzes 22 reactions, all occurring in the peroxisomes, but not all substrates bind to the same active site of the enzyme. Therefore, in this case we need to consider two reaction groups, each consisting of 11 reactions, the substrates of which act as inhibitors only within their assigned group.
- The third case (case III) arises when the same enzyme is found in multiple cellular compartments; for example the enzyme encoded by AUR1, exhibiting phosphatidylinositol:ceramide phosphoinositol transferase activity. This enzyme operates in both the endoplasmic reticulum and the Golgi apparatus, catalyzing ten reactions in total. The substrates of these reactions cannot compete for binding if they are not located in the same compartment. Thus, for this case we need to define two reaction groups, one per cellular component, each consisting of five reactions.
- It is important to note that we do not consider cofactors as apparent inhibitors. For instance, there are multiple reactions that involve either a NAD⁺/NADH or a NADP⁺/NADPH cofactor pair interchangeably. In reality the phosphorylated and non-phosphorylated species compete for the binding site, but our focus was to investigate these inhibitory mechanisms and their effect solely in the lipid species pathways for this study. Thus, the final case (case IV) encompasses promiscuous enzymes that apart from catalyzing multiple lipid-related reactions, they can catalyze these reactions with multiple cofactor pairs. For this case we defined separate reaction groups per cofactor pair, since the apparent inhibitors will not be affected by the reacting cofactor, and it makes it easier to visualize in terms of numbers.

The aforementioned cases of apparent inhibition led us to define a total of 62 reaction groups to describe the enzymatic coupling effects in the network. The procedure we used is the following: (i) We identified the reactions which are catalyzed by the same enzyme or enzymatic activity. (ii) We separated these reactions if they occurred in

multiple cellular compartments. (iii) We further separated these reactions if the catalytic activity was not the same for all of them. For this step we used the Enzyme Commission (E.C.) number to define the activity type, under the assumption that one enzymatic active site accommodates only reactions with a specific E.C. number. Thus, if two (or more) reactions catalyzed by a common enzyme have been assigned different E.C. numbers (up to the second level) we assume that there is more than one active site involved, and consequently no apparent inhibition is taking place. Cases of catalytic promiscuity in a single active site have recently been documented (Velez Rueda et al. 2019), although none of the reported proteins are part of the redLips network. All of the identified enzymes along with the corresponding case, defined reaction groups and number of catalyzed reactions can be found in Table 3.1. The complete list of reactions per group can be found in supplementary Table S3.4.

Table 3.1. Promiscuous enzymes in the redLips network, identified by their standard gene name, along with the corresponding case, defined reaction groups and number of catalyzed reactions.

Gene Std. Name	Case	Groups Defined	# Reactions Catalyzed
<i>ALE1</i>	I	<i>ALE1</i>	4
<i>ARE2</i>	I	<i>ARE2</i>	5
<i>AUR1</i>	III	<i>AUR1_ER</i> <i>AUR1_GOLGI</i>	5 5
<i>BTS1</i>	I	<i>BTS1</i>	2
<i>CEM1</i>	I	<i>CEM1</i>	8
<i>DPL1</i>	I	<i>DPL1</i>	2
<i>ECI1</i>	I	<i>ECI1</i>	3
<i>ELO1</i>	I	<i>ELO1</i>	2
<i>ELO2</i>	I	<i>ELO2</i>	2
<i>ELO3</i>	I	<i>ELO3</i>	3
<i>ERG20</i>	I	<i>ERG20</i>	2
<i>ERG25</i>	IV	<i>ERG25_A</i> <i>ERG25_B</i>	6 6
<i>ERG26</i>	IV	<i>ERG26_A</i> <i>ERG26_B</i>	2 2
<i>ERG27</i>	I	<i>ERG27</i>	2
<i>ERG9</i>	I	<i>ERG9</i>	2
<i>ETR1</i>	I	<i>ETR1</i>	8
<i>FAA1</i>	II	<i>FAA1_A</i>	4

3.3 Results and Discussion

		<i>FAA1_B</i>	5
<i>FAA2</i>	I	<i>FAA2</i>	14
<i>FAA3</i>	I	<i>FAA3</i>	3
<i>FAS1</i>	II	<i>FAS1_A</i>	8
		<i>FAS1_B</i>	8
		<i>FAS1_C</i>	2
<i>FAS2</i>	II	<i>FAS2_A</i>	8
		<i>FAS2_B</i>	8
<i>FAT1</i>	I	<i>FAT1</i>	4
<i>FOX2</i>	II	<i>FOX2_A</i>	12
		<i>FOX2_B</i>	12
<i>HTD2</i>	I	<i>HTD2</i>	8
<i>IFA38</i>	I	<i>IFA38</i>	7
<i>IPT1</i>	I	<i>IPT1</i>	5
<i>ISC1</i>	I	<i>ISC1</i>	15
<i>LAC1</i>	III	<i>LAC1_CYTOSOL</i>	2
		<i>LAC1_ER</i>	2
<i>LCB3</i>	I	<i>LCB3</i>	2
<i>LCB4</i>	I	<i>LCB4</i>	2
<i>LRO1</i>	I	<i>LRO1</i>	2
<i>OAR1</i>	I	<i>OAR1</i>	8
<i>OLE1</i>	I	<i>OLE1</i>	8
<i>OPI3</i>	I	<i>OPI3</i>	2
<i>PHS1</i>	I	<i>PHS1</i>	7
<i>PLB3</i>	I	<i>PLB3</i>	4
<i>POT1</i>	I	<i>POT1</i>	12
<i>POX1</i>	I	<i>POX1</i>	16
<i>PXA1</i>	I	<i>PXA1</i>	11
<i>RER2</i>	I	<i>RER2</i>	11
<i>SCS7</i>	I	<i>SCS7</i>	3
<i>SCT1</i>	I	<i>SCT1</i>	2
<i>SPS19</i>	I	<i>SPS19</i>	4
<i>SRT1</i>	I	<i>SRT1</i>	5
<i>SUR1</i>	I	<i>SUR1</i>	5
<i>SUR2</i>	I	<i>SUR2</i>	2
<i>TAZ1</i>	I	<i>TAZ1</i>	2
<i>TES1</i>	III	<i>TES1_CYTOSOL</i>	10
		<i>TES1_PEROXISOME</i>	14
<i>TGL1</i>	I	<i>TGL1</i>	5
<i>TGL3</i>	I	<i>TGL3</i>	2
<i>TSC13</i>	I	<i>TSC13</i>	7
<i>YDC1/YPC1</i>	I	<i>YDC1 / YPC1</i>	2

3.3.4 Kinetic mechanisms – Rest of the network

For the rest of the reactions of the redLips network, we also had to assign appropriate kinetic mechanisms to describe their enzymatic properties. It is important to note that in the case of boundary reactions no mechanism is assigned since they are artificial representations serving to mass-balance the network, thus these reactions are not a part of the resulting kinetic model. Within the ORACLE framework, the biomass reaction will also be split into individual biosynthetic fluxes for each building block that contributes to cell growth. In total 1161 reactions were assigned kinetic mechanisms, with the majority being uni-uni reactions which were described by Michaelis-Menten kinetics. These reactions were 404 in number, followed by the 346 reactions that were assigned a form of apparent inhibition kinetics. The rest of the reactions follow a form of reversible hill kinetics, depending on their stoichiometry, convenience kinetics if one or more of their stoichiometric coefficients is larger than unity or the number participating metabolites is more than four, simple mass-action kinetics if the participating reactants are all defined as small molecules or inorganics, and irreversible mass-action kinetics for the cases of enzymatic irreversibility, pooling or lumped reactions and the elementary biosynthetic and diffusion reactions. A remark must be made for the enzyme Phosphofructokinase-1 (PFK1), which converts fructose 6-phosphate and ATP to fructose 1,6-bisphosphate and ADP. Phosphofructokinase activity is also subject to allosteric control, with ATP inhibiting the enzyme, and AMP along with fructose 2,6-bisphosphate reversing the inhibition by allosteric activation. Thus, a special kinetic mechanism needs to be assigned to this particular reaction in order to accurately capture these phenomena. Other enzymes of the network have also been reported to being subject to allosteric or another type of control but were not considered in this study. The statistics of the assigned mechanisms over the whole network and their description can be seen in Table 3.2. The complete mathematical expressions can be found in Appendix A.

Table 3.2. Overall assigned kinetic mechanisms in the redLips network according to each enzymatic function and stoichiometry.

Stoichiometry	Kinetic Mechanism	Number of reactions
<i>Uni - Uni</i>	Reversible Michaelis Menten	404
	Irreversible Michaelis Menten	125
	Mass-action kinetics	28
	Convenience kinetics with Apparent Inhibition	56
<i>Bi - Bi</i>	Reversible Hill	138
	PFK specific mechanism	1
	Irreversible Michaelis Menten	3
	Mass-action kinetics	1
	Convenience kinetics with Apparent Inhibition	185
<i>Bi - Uni / Uni - Bi</i>	Reversible Hill	32
	Mass-action kinetics	2
	Convenience kinetics with Apparent Inhibition	74
<i>Ter - Bi / Bi - Ter</i>	Reversible Hill	7
	Irreversible Michaelis Menten	1
	Mass-action kinetics	1
	Convenience kinetics with Apparent Inhibition	25
<i>More complex stoichiometry</i>	Convenience kinetics	37
	Convenience kinetics with Apparent Inhibition	6
<i>Other stoichiometry</i>	Irreversible Michaelis Menten	33
	Mass-action kinetics	2

Another fine point that needs to be addressed at this stage concerns the parametrization of the kinetic rate expressions when enzymatic coupling effects are present. Consider the example case of two reactions that are catalyzed by the same enzyme and involve the same co-substrates irrespective to apparent inhibition effects. Since in essence the same biomolecule creates a complex in the same binding site of the same enzyme for both of these reactions, the binding affinity between this molecule and this enzyme, expressed through the respective Michaelis constant, needs to be equal in the rate expressions for these two reactions. This property was enforced in the sampling algorithm of the ORACLE workflow.

3.3.5 The importance of accounting for the enzymatic coupling effect

Apart from the need to account for common binding affinities, enzymatic coupling also requires an appropriate reformulation of the parameter elasticity matrix, as presented in the Materials and Methods section. Depending on the consideration – or lack thereof – of enzymatic coupling effects, we can distinguish four distinct cases of mathematical complexity: The simplest case, and most common in literature, is the disregard of enzymatic coupling effects, meaning that the apparent inhibition terms are absent in the rate expressions of all reactions. The second case presents a simplistic approach to modeling enzymatic coupling, which includes the restructuring of the parameter elasticity matrix Π , but without accounting for apparent inhibition in the kinetic rate expressions. In both of these cases, the kinetic parameter sets are identical, but the calculation of corresponding CCs will be different. On the other hand, we have the occasion where enzymatic coupling effects are considered directly in the kinetic mechanisms and reaction rates, which gives rise to two more cases with respect to the inclusion – or lack thereof - of the restructuring of the parameter elasticity matrix Π . Bearing in mind the network complexity and consequent uncertainty, we generated 9'400 sets of kinetic parameters for this model using the ORACLE framework, accounting for enzymatic coupling effects in the kinetic mechanisms assigned. We additionally generated 10'000 sets of kinetic parameters without considering any enzymatic coupling effects in the reaction kinetics. The corresponding FCCs and CCCs were calculated for each parameter set of both of these cases according to the expressions defined in the Materials and Methods section for both the cases of identity and restructured Π matrix.

According to the definition of CCs, a positive or negative CC value means that by increasing the corresponding enzyme activity, the concerned flux or concentration value will also increase or decrease, respectively. In Figure 3.1, we present a comparison of the results generated for each of the aforementioned four cases with reference to the phosphatidylinositol synthase (PIS1) reaction. The biosynthesis of PI is a major part of the phospholipid biosynthetic pathways and serves well in demonstrating the differences between the four examined cases. In complete absence of the enzymatic coupling effect (Figure 3.1a), the enzymatic activities that appeared to affect most this reaction flux were the inositol-3-phosphate synthase (INO1),

hexokinase (HXK1), and inositol monophosphatase (INM1). All of these enzymes presented positive control over PIS1, meaning that an increase in their activity would result in an increased flux through the reaction. This trend is logically explained since all three enzymes catalyze the reactions upstream of PIS1, and lead to the production of inositol, one of the two building blocks of phosphatidylcholine (PI).

Moving to the case where a restructured Π matrix was used but the kinetic mechanisms remained devoid of apparent inhibition terms, we already noted some key differences between the CC distributions and rankings. Of course, since the only difference between this and the previous case is the form of the parameter elasticity matrix, only the CCs of enzymes catalyzing more than one reaction were affected. The three top enzymatic activities affecting PIS1 one remained the same as in the previous case, although inositol phosphosphingolipid phospholipase C (ISC1) appeared in the fourth position. ISC1 catalyzes 15 reactions. Sphingolipid degradation through hydrolysis, catalyzed by ISC1, produces myo-inositol phosphate, a precursor to inositol and therefore a vital component of PI biosynthesis, while PI itself participates in the biosynthesis of sphingolipids. Another enzyme that had more control over PIS1 when the restructured parameter elasticity matrix was taken into account, is inositol phosphorylceramide synthase (AUR1). AUR1 catalyzes five reactions that encompass the synthesis of the first complex sphingolipids group, namely inositol phosphoceramides (IPCs), by attaching the head group of PI to a ceramide and producing an IPC and diacylglycerol.

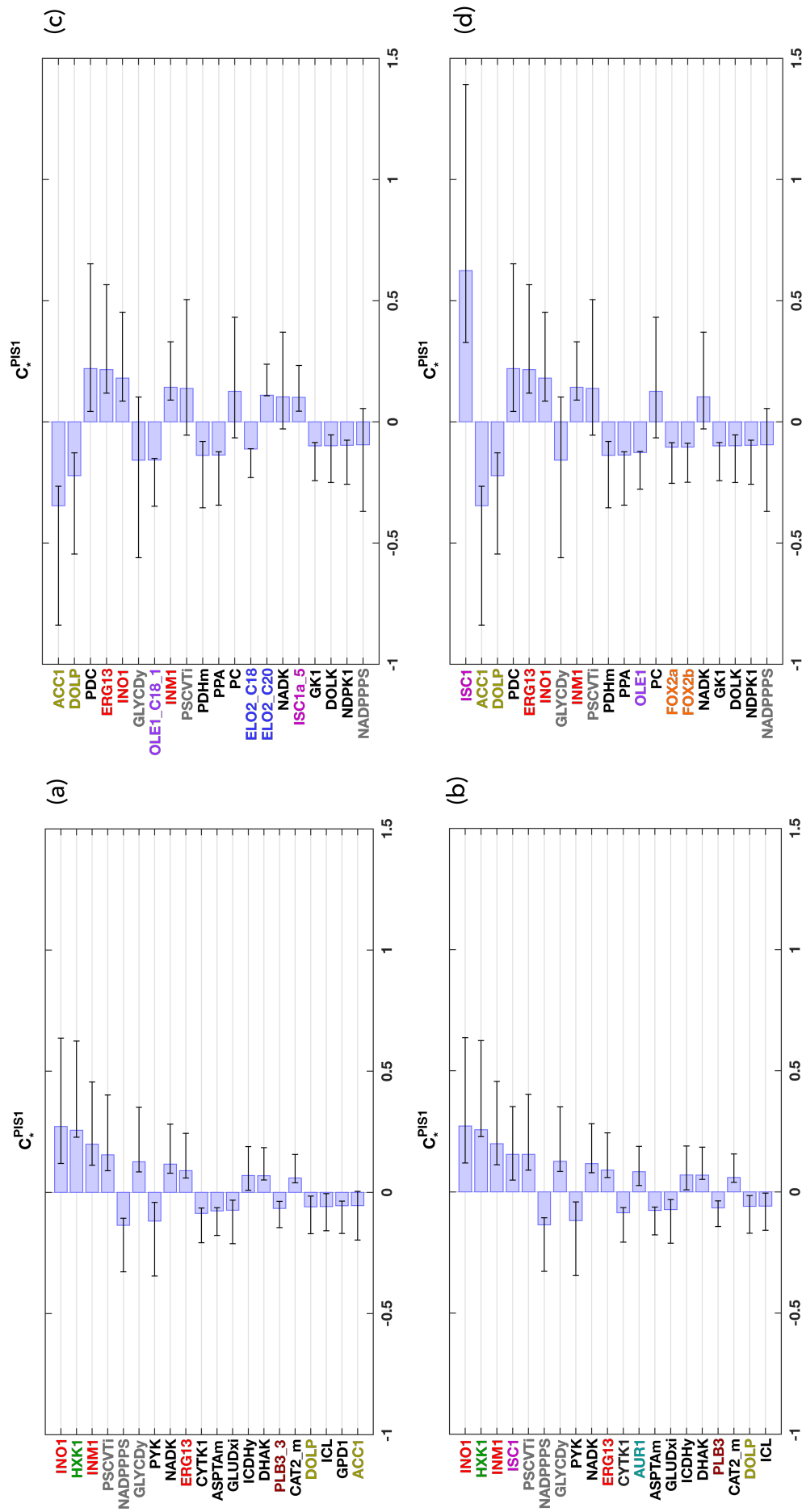


Figure 3.1. Top 20 ranked Flux Control Coefficients corresponding to the reaction PIS1, for the cases of (a) no enzymatic coupling and without restructuring Π matrix, (b) no enzymatic coupling with restructuring Π matrix, (c) enzymatic coupling without restructuring Π matrix, (d) enzymatic coupling with restructuring Π matrix. The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively. Colors denote parameters of interest and are discussed in the text.

When accounting for the enzymatic coupling effect, but not using a restructured Π matrix (Figure 3.1c), the enzymes that exhibited the most control over PIS1 were significantly different. The top ranked ones were acetyl-CoA carboxylase (ACC1), endoplasmic reticular dolichol kinase (DOLP or SEC59), and pyruvate decarboxylase (PDC or PDC1), with the two first having negative and the latter positive control over the reaction. This result suggested a connection between the synthesis of PI and acetyl-CoA, and even the dephosphorylation of dolichol-phosphates. These connections have not been previously reported in literature and could lead to significant discoveries in the regulatory mechanisms of PI synthesis. The previously top ranked enzymes appeared in a lower ranking, with INO1 at the fourth and INM1 at the eighth place, while HXK1 did not appear at all in the top 20 controlling enzymes. It is interesting to note that the values of the CCs for those particular enzymes in both of the cases were more or less equal, but that was not the case with ACC1, which appeared twentieth in rank for the first case, was absent in the second case, and first in the third case. This occurrence in fact strongly suggests that the effect that enzymatic coupling will have in the calculation of CCs for enzymes that catalyze a single reaction is not straightforward, and it strengthens the importance of its consideration. We could also identify some of the promiscuous enzymes appearing in lower rankings of the top 20 for this case, of course in separated forms since the Π matrix was considered as an identity matrix and the contributions of each enzymatic control were not combined into a single CC. Namely, these enzymes were the OLE1_C18_1 (fatty acid desaturase), ELO2_C18 and ELO2_C20 (fatty acid elongase), as well as ISC1a_5 (inositol phosphosphingolipid phospholipase C).

Finally, accounting for the enzymatic coupling effect and using a restructured Π matrix (Figure 3.1d) further altered the resulting highest control asserting enzymes. As mentioned earlier, only the CCs of enzymes catalyzing more than one reaction were affected between this and the previous case, resulting from the form of parameter elasticity matrix. By adding up the control contributions for promiscuous enzymes, ISC1 climbed up to exerting the most control over PIS1. The succeeding ranked enzymes were unaffected, with ACC1, DOLP and PDC following. Fatty acid desaturase (OLE1) appeared also in this case, but in a lower rank, 3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase (FOX2a and FOX2b) appeared in ranks 14

and 15 respectively, while fatty acid elongase (ELO2) did not appear in the top 20 ranked enzymes.

These observations led us to the classification of the enzymatic activities appearing in each ranking into a few comprehensive categories, which are accordingly denoted in Figure 3.1 with different colors. A number of enzymes, namely INO1, INM1, and ERG13 (in red), appeared in all four cases in the top 20 ranks, and even though the mean value of each of their distribution might have changed, the sign of the control remained unaffected. On the other hand, while some enzymes such as 3-phosphoshikimate 1-carboxyvinyltransferase (ARO1 or PSCVTi in model), NADP⁺ phosphatase (NADPPPS) and glycerol dehydrogenase (GCY1 or GLYCDy in model) (in grey) appeared in all four cases, their control sign changed between the case pairs of consideration or not of apparent inhibition effects in the kinetic mechanisms (Figure 3.1 a, b versus c, d). Especially for the case of GLYCDy, the mean value of the distribution varied dramatically between the two instances, while for the other two enzymes only the top (or bottom) quartiles of their distributions crossed to the opposite sign. Another difference between those two case pairs was the control of the enzymes ACC1 and DOLP (in yellow), which appeared very low in rank when the inhibitory effect in the kinetic mechanisms was neglected, but rose to very high control values when it was not. HXK1 (in green) appeared in a high rank only for the cases where no apparent inhibition effects were considered (Figure 3.1 a, b). Some of the most striking results of this study involve promiscuous enzymes, and the effect that the consideration of the enzymatic coupling effect has on their predicted control. It is clear from these results that that this inclusion, with regard to both apparent inhibition terms in the kinetic mechanisms and a restructured Π matrix, will not necessarily lead to greater (or lesser) CC values for promiscuous enzymes. As we observed, it may very well enhance the combined enzymatic control, like in the case of ISC1 (in magenta), AUR1 (in turquoise) and FOX2a/b (in orange), decrease it, like in the case of ELO2 (in blue) and OLE1 (in purple), or even not affect it at all, like in the case of PLB3 (in dark red). More specifically, ISC1 was not found to exert significant control over PIS1 if the fact that it catalyzes several reactions was not considered (Figure 3.1 a, c), but this changed dramatically for the cases it was (Figure 3.1 b, d). Apart from ISC1, AUR1 and PLB3 were the only promiscuous enzymes to

appear in the top ranking only for the cases of no apparent inhibition kinetics (Figure 3.1 a, b). On the contrary, OLE1, ELO2, and FOX2a, b, appeared in top ranking only for the cases of apparent inhibition kinetics.

This analysis investigated the effect that accounting for enzymatic coupling has on metabolic control predictions. It was demonstrated that the effect of enzymatic coupling is not negligible. On the contrary, predictions varied significantly between the studied cases. Moreover, it was evident that by considering this effect only through the restructuring of the parameter elasticity matrix and not in the kinetic rate expressions was not adequate. The assignment of kinetic rate expressions that account for apparent inhibition induces a coupling of the competing metabolites. This coupling affects the dynamic behavior throughout the metabolic network and needs to be taken into consideration. Thus, in order to model properly the properties and consequences of enzymatic promiscuity, the enzymatic coupling needs to be taken into account in both the rate laws and the Π structure.

3.3.6 Validation of the model's predictive capabilities

Having demonstrated the way to consistently account for the enzymatic coupling effect, we investigated the predicted control exerted on key reactions of the network, using the kinetic parameters and CCs calculated for this case.

A very well-known example of internal cell regulation is the use of the Kennedy pathway for the production of the phospholipids phosphatidylethanolamine (PE) and phosphatidylcholine (PC) in the case of impaired production by the CDP-DAG pathway (Carman and Henry 1989, Carman and Henry 1999, Carman and Zeimetz 1996). Mutants defective on these enzymes depend on exogenous supply of ethanolamine and choline to synthesize CDP-ethanolamine and CDP-choline and subsequently the aforementioned phospholipids, respectively. In Figure 3.2 we present the FCCs for the control of the enzymes of the CDP-DAG pathway on the reactions of the two Kennedy pathways. We can see that for both branches of the pathway, the enzyme phosphatidylserine synthase (CHO1) has the strongest effect on all reactions, and the values are strictly negative. CHO1 catalyzes the first step of the CPD-DAG pathway which is the synthesis of phosphatidylserine (PS) from CDP-DAG

and serine. This result is in agreement with the physiological observations (Carman and Henry 1989, Carman and Henry 1999, Carman and Zeimet 1996) since according to MCA definitions a negative CC value corresponds to increased flux through the reaction in the case of decreased enzyme activity. We also observe negative control from the enzyme that catalyzes the synthesis of the main precursor CDP-DAG, namely phosphatidate cytidyltransferase (CDS1), leading us to the hypothesis that the suppression of the CDP-DAG pathway in favor of the Kennedy pathway could be starting one step before, at the very synthesis of the compound. The enzymes catalyzing the next step of the pathway which is the transformation of PS to PE in three cellular compartments (phosphatidylserine decarboxylases PSD1 and PSD2) exhibit likewise slightly negative control but of very small magnitude. It is very interesting to note that the last two enzymes of the pathway catalyzing the three methylation reactions that convert PE to PC (phosphatidylethanolamine methyltransferase CHO2 and methylene-fatty-acyl-phospholipid synthase OPI3) affect differently each of the branches of the Kennedy pathway. Concerning the PE production branch, the values of the corresponding CCs are strictly positive, while for the PC production branch values are virtually zero, so they do not affect it. For the first case, however, a hypothesis can be made that in the event of increased synthesizing activity of PC from PE, the Kennedy pathway will need to be more active in order to compensate for the increased demand of PE, since these enzymes are both downstream of the PE synthesis. All of the above remarks serve to validate our model's predictions against experimental observations and lead to the formulation of novel hypotheses for the underlying mechanisms of regulation in these pathways.

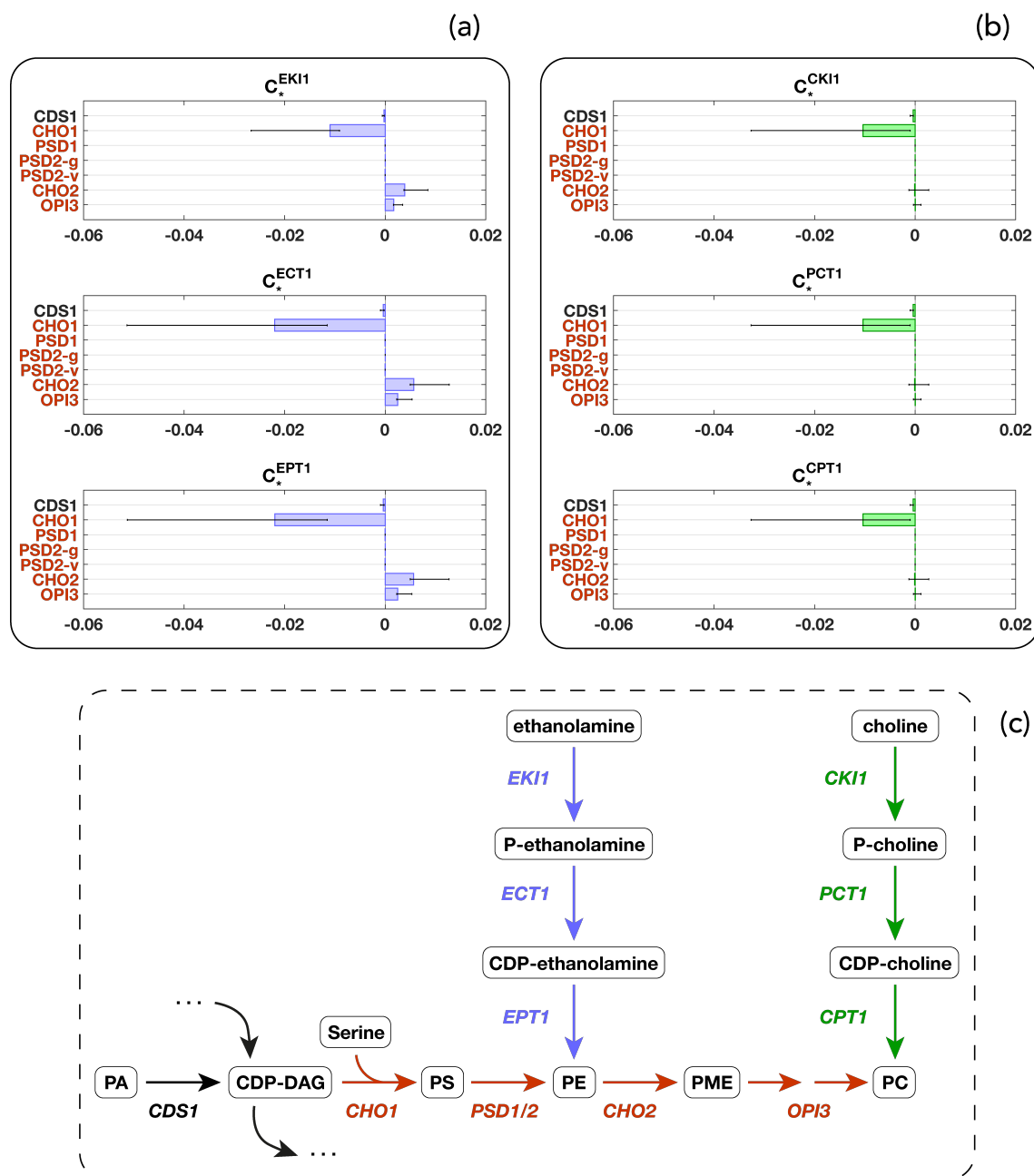


Figure 3.2. Flux Control Coefficients corresponding to the reactions of the Kennedy pathway for the biosynthesis of (a) PE (in purple) and (b) PC (in green), with respect to the enzymes catalyzing the CDP-DAG biosynthetic pathway (in red). The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively. (c) The CDP-DAG and Kennedy biosynthetic pathways, with corresponding colors.

Another point of interest is the concentration levels of cardiolipin (CL) in the system. CL is a major component of mitochondrial membranes and mitochondrial functions, and abnormalities in its structure or production have been associated with numerous diseases such as Barth syndrome, ischemia and cardiomyopathy (Baile, Lu and Claypool 2014, Schlame and Greenberg 2017). Barth syndrome has been associated with the human tafazzin gene G4.5, which is highly homologous to the yeast gene TAZ1 (lyso-phosphatidylcholine acyltransferase). Tafazzin is reportedly responsible for CL remodeling, which consists of the diacylation of newly synthesized CL to monolyso-CL (MLCL), which then undergoes condensation with an acyl-CoA to form mature CL. This process is essential in maintaining the unique CL acyl composition for optimal performance of the mitochondrial associated functions. Mutants defective in tafazzin have been reported to accumulate large quantities of MLCL and present with lower levels of remodeled CL, along with other phospholipid imbalances (Gu et al. 2004). In Figure 3.3 we can see the control of TAZ1 on the network's metabolite concentration levels, in descending order.

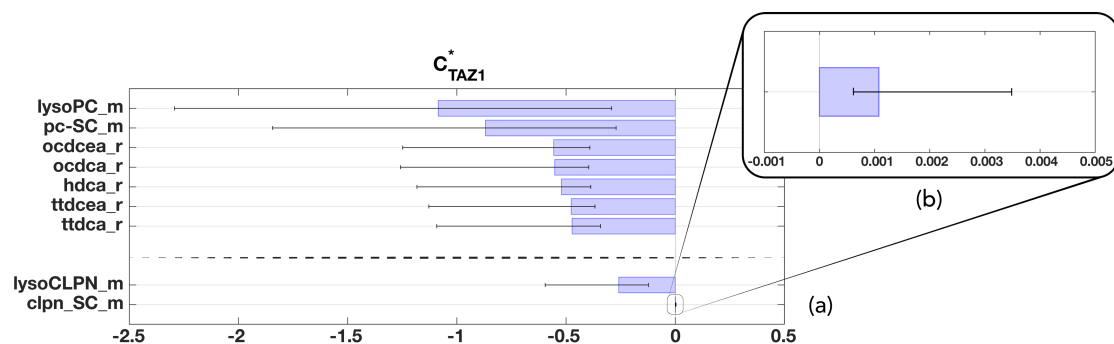


Figure 3.3. (a) Top 7 ranked Concentration Control Coefficients (top) corresponding to the enzyme activity TAZ1 along with the lower ranked MLCL and CL species (bottom), and (b) zoom in on the bar corresponding to CL. The bars and error bars denote the mean value and the lower and upper quartiles of the distribution respectively.

We observed that the most affected concentrations are the ones of PC and monolyso-PC (MLPC) in mitochondria, along with a number of free fatty acids in the ER. These results are in agreement with experimental observations concerning altered phospholipid composition of TAZ1 mutants (Gu et al. 2004). Interestingly, the CL species in tafazzin mutants include mostly species with increased C18:1, C18:0 and 16:0 acyl chains, which are exactly the ones predicted to increase in concentration in our calculations. Additionally, we present the effect that altered TAZ1 expression would have in CL and MLCL levels, which is negative and minorly positive, respectively. The signs and magnitudes of the CCs express that in the case of reduced TAZ1 activity, accumulation of MLCL and lesser depletion of CL will occur in the cell, which is in agreement with the observations of Gu et al. (2004).

This and the above analysis show that our model's predictions are in very good agreement with physiological data. This validation procedure is essential in the construction of kinetic models in order to prune the generated sets of kinetic parameters if necessary. Having thus demonstrated that the kinetic version of redLips can accurately capture experimental observations, novel interactions within the metabolic network can be identified and evaluated.

3.4 Conclusions

In conclusion, we demonstrated the consistent construction of a kinetic version of the redLips model, by taking into account the network complexities and limitations thereof. The kinetic redLips model is the most detailed kinetic representation of lipids to date, and one of the largest kinetic models available. Lipid related enzymes are multifunctional and promiscuous, and special care needs to be taken when developing and assigning kinetic rate expressions. The notion of apparent inhibitors was used to allow the definition of appropriate enzymatic mechanisms and the discrepancies this consideration induces in the results was investigated. The thermodynamically feasible flux directionality profiles of the network were assessed and a representative steady state was chosen. Experimental lipidomics were used to curate the model in

terms of biosynthetic fluxes and concentrations, by taking into account the compartmentalization of species and the relative volume of each cellular compartment, and conserved moieties were identified. Populations of kinetic parameter sets were generated using the ORACLE framework through extensive sampling, and the corresponding MCA-based control coefficients were calculated and statistically analyzed. The importance of accounting correctly for the enzymatic coupling effect was demonstrated, and enzymatic control over key reaction rates and metabolite concentrations was examined. We validated the model against experimental observations and showed that its predictions were in very good agreement with reported biological responses. Within these studies, novel regulatory interactions were identified. We believe that the systematic workflow presented can be used as a guideline to building consistent kinetic models of metabolism, while allowing for the inclusion of complex phenomena such as enzymatic promiscuity. Furthermore, the constructed model can be used to study dynamic responses in response to perturbations in the lipid metabolic pathways, and could be a useful tool in the elucidation of the mechanisms that control membrane homeostasis.

Chapter 4 – Response Balance Analysis: investigating metabolic control in a constraint-based formulation for the design of strain engineering strategies

4.1 Introduction

Bioprocess optimization has been a very prominent area of study over the past decades. Several factors contribute to the improvement of how an organism handles several metabolic tasks, from growth to yield of products. Evolution itself is a prime example of how an organism can efficiently adapt and perform in various conditions, ranging from environmental and media related conditions, to gene expression and intracellular flux and concentration redistributions.

Genome editing techniques have significantly improved over the years and are enabling the metabolic engineering of organisms that can achieve a targeted physiological state. Two general approaches exist for reaching this desired state; cutting off existing control loops that will result in the modification the endogenous regulatory architecture or, heterologous pathways can be integrated to disruptively overcome native control patterns (Bailey 1991). The latter method requires testing if the integration of DNA fragments into the original genome sequence successfully perturbs cellular regulation in a desired fashion, and can be relatively laborious. The former technique demands knowledge about cellular control so that the DNA sequence can be modified effectively and without unfavored side effects. Widely

available gene-protein-reaction associations of various organisms provide invaluable information about their metabolism and enable the construction and elaboration of mathematical models that can be studied computationally in order to interrogate and analyze their cellular behavior (Gombert and Nielsen 2000). The various emerging mathematical formulations can yield candidate metabolic engineering strategies for consistently achieving the target physiological state via genome editing. Strain design can be accommodated by the identification of pathways that enable production of desired compounds using known native and/or known non-native reactions, as well as *de novo* reactions that have not been observed to biologically occur. Additionally, strategies can include gene deletions and/or up- and down-regulation of enzymes. For an extensive review of the most prominent computational tools and workflows, the interested reader may refer to the following reviews (Costa, Hartmann and Vinga 2016, Long, Ong and Reed 2015, Wang et al. 2017).

Kinetic models of metabolism are mathematical descriptions of cellular dynamics and provide useful information about the regulatory architecture of an organism. The construction of kinetic models involves numerous sources of uncertainty as highlighted in previous studies (Andreozzi, Miskovic and Hatzimanikatis 2016, Chakrabarti et al. 2013, Hameri et al. 2019) and reviews (Miskovic et al. 2015, Saa and Nielsen 2017). This uncertainty can stem from a large allowable solution space native to stoichiometric mechanistic models, deviation errors in experimental measurements, or lack of information about the network in both a structural and a parametrization level. Consequently, sampling-based workflows that generate populations of models in an effort to statistically quantify and overcome the introduced uncertainty have been developed (Chakrabarti et al. 2013, Khodayari and Maranas 2016, Tran, Rizk and Liao 2008, Wang, Birol and Hatzimanikatis 2004). One of these workflows is the Optimization and Risk Analysis of Complex Living Entities (ORACLE) (Miskovic and Hatzimanikatis 2010), which, given a representative cell physiology, systematically generates sets of kinetic parameters through sampling.

Metabolic control analysis (MCA) (Kacser, Burns and Fell 1995) has been established as an efficient method for studying control in biological systems, as it does not require mathematical integration of complex systems of ordinary differential equations. Furthermore, Hatzimanikatis, Floudas and Bailey (1996a), (1996b) have shown that

MCA-based flux and concentration control coefficients (FCCs and CCCs, respectively) can be used to devise a Mixed Integer Linear Programming (MILP) formulation for querying cellular responses upon enzymatic perturbations. Their study highlights that to achieve a given perturbed state for a simple pathway, multiple alternative metabolic engineering strategies can be suggested. On the scale of simple linear and branched pathways, and serving as proof of concept, control coefficients emerging from kinetic modeling were used successfully to propose metabolic engineering strategies. The advantage of this method is that physiologically relevant bounds and constraints can be imposed to the system, as opposed to the classical MCA. However, to our knowledge, their proposed MILP formulation has not been applied to larger scale nonlinear kinetic models.

We hereby present Response Balance Analysis (RBA), a workflow that utilizes populations of kinetic parameter sets and their corresponding control coefficients (CCs) in order to consistently derive metabolic engineering strategies. We used an *E. coli* metabolic model (Hameri, Fengos and Hatzimanikatis 2019) which describes the central carbon pathways in aerobic growth conditions and was curated with thermodynamic feasibility constraints and experimental data where available. We used the ORACLE framework in order to generate multiple sets of kinetic parameters along with their corresponding flux and control coefficients. We then demonstrated how through RBA these data can be formulated in a MILP problem and how it can be used to efficiently analyze, enumerate, and propose alternative metabolic engineering strategies. RBA can be a powerful tool for the design of sophisticated approaches, and it can accommodate the incorporation of physiological data in reaction rate, concentration and enzyme expression levels.

4.2 Materials and Methods

4.2.1 Metabolic Control Analysis notions

In MCA, the concentration control coefficients, C_p^x , and the flux control coefficients, C_p^v , are defined as the fractional change of metabolite concentrations x and metabolic

fluxes v , respectively, in response to a fractional change of system parameters p (Kacser and Burns 1973a). These CCs serve as measurable outputs that provide information about the levels of control that system parameters have on the studied biological system and physiology. From the log(linear) formalism (Hatzimanikatis, Floudas and Bailey 1996a, Reder 1988a), C_p^x and C_p^v can be derived through the following expressions:

$$C_p^x = -(NVE)^{-1}NV\Pi$$

$$C_p^v = EC_p^x + \Pi$$

where N is the stoichiometric matrix, V is the diagonal matrix whose elements are the steady-state fluxes, E is the elasticity matrix with respect to metabolites and Π is the matrix of elasticities with respect to parameters.

Hence, flux and concentration control coefficients are computed for each reaction flux i and metabolite concentration j with respect to the system parameters k as:

$$C_{p_k}^{v_i} = \frac{d \ln v_i}{d \ln p_k} = \frac{p_k}{v_i} \frac{dv_i}{dp_k}$$

$$C_{p_k}^{x_j} = \frac{d \ln x_j}{d \ln p_k} = \frac{p_k}{x_j} \frac{dx_j}{dp_k}$$

4.2.2 Kinetic model description

In this study, we used a kinetic model that describes the aerobically grown physiology of *E. coli* (Hameri, Fengos and Hatzimanikatis 2019) for demonstrating the RBA workflow. This model was systematically reduced from the *E. coli* iJ01366 genome-scale model (Orth et al. 2011a) around the originally defined reaction subsystems of glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, glyoxylate cycle, pyruvate metabolism and the electron transport chain (ETC). The

reduction was performed through the redGEM and the lumpGEM algorithms (Ataman et al. 2017, Ataman and Hatzimanikatis 2017), thus ensuring the preservation of as much information as possible as well as that thermodynamic feasibility constraints are respected. This model constitutes of 337 metabolites participating in 647 reactions, which are in turn associated with 271 enzymes that serve as parameters in the RBA formulation. The model was curated with thermodynamic feasibility constraints using TFA (Salvy et al. 2019) and relevant fluxomics data (McCloskey et al. 2014). The representative steady state profiles of the metabolite concentrations and metabolic fluxes were chosen with Principal Component Analysis (PCA) as detailed in the relevant publication, and the kinetic model with associated populations of kinetic parameters was built using the ORACLE workflow (Miskovic and Hatzimanikatis 2010).

4.3 Results and Discussion

4.3.1 Response Balance Analysis formulation

The MCA control coefficients can be used to compute how a given metabolic flux or metabolite concentration would respond to parameter perturbations of the system. These deviations can be calculated as follows:

$$F_i = \sum_{k=1}^{\# \text{ of } p} C_{p_k}^{v_i} \times E_k$$
$$M_j = \sum_{k=1}^{\# \text{ of } p} C_{p_k}^{x_j} \times E_k$$

where

$$F_i = \ln (v_i/v_{i,0})$$

$$M_j = \ln (x_j/x_{j,0})$$

$$E_k = \ln (p_k/p_{k,0})$$

and F , M and E are the logarithmic deviations in flux, metabolite concentration and parameter with respect to their respective reference steady states $v_{i,0}$, $x_{i,0}$ and $p_{i,0}$ (Figure 4.1).

Similar to mass balances in the Flux Balance Analysis (FBA) mathematical formulation (Orth, Thiele and Palsson 2010), these equations can be written in matrix-vector form forming a constraint-based linear problem, on which we can impose bounds on the problem's variables. Since the enzyme expression levels of a cellular system could either be up- or down-regulated, but not both at the same time, we made use of integer variables to model these properties in our system. The integer variables along with their accompanying constraints, in combination with the above linear equations, give rise to a MILP mathematical formulation for studying system control. For the full mathematical derivations of the formulation, one may refer to the original publications (Hatzimanikatis, Floudas and Bailey 1996a, Hatzimanikatis, Floudas and Bailey 1996b).

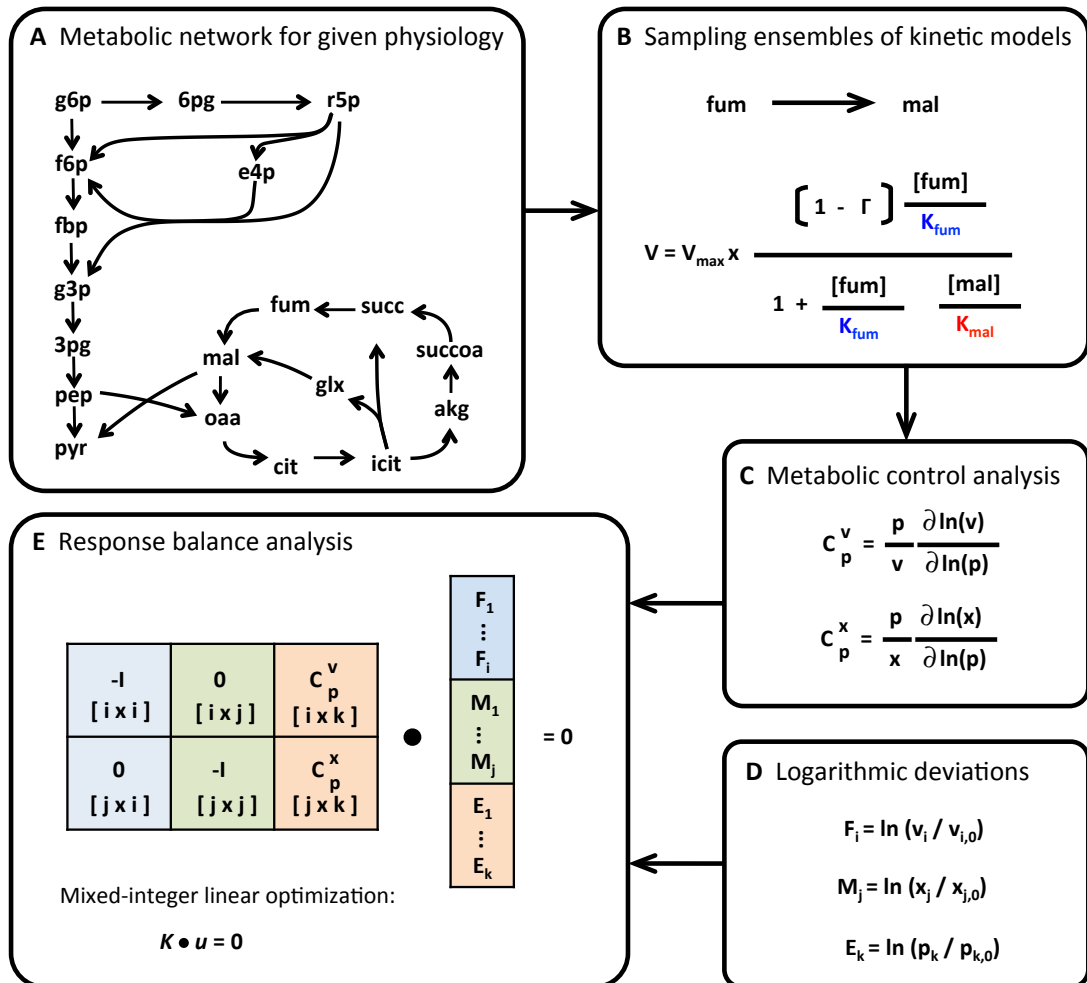


Figure 4.1. Schematic representation of the RBA formulation. Diagram providing information about the various steps necessary for constructing an RBA model from kinetic models for studying system response to perturbations.

4.3.2 Navigating the degrees of freedom

One of the biggest advantages of the MILP problem formulation, is that it allows us to set bounds for all the aforementioned logarithmic deviations (fluxes, metabolite concentrations and parameters), as well as introduce additional relevant constraints to the system. The variable bounds can be inferred from physiological assumptions or experimental measurements, while the associated constraints can impose design

limitations, such as a maximum number of total genetic manipulations. Still, there is a large amount of uncertainty associated with the many degrees of freedom of this formulation in terms of appropriate bound, constraint, and objective selection. Combined with the selection of a representative kinetic parameter set, the feasible solution space can be vast and induce complexity in the computation and interpretation of results. In the following subsections we discuss each of these aspects of the formulation and how we endeavored to tackle them.

4.3.2.1 Selecting a set of kinetic parameters

One of the core points of the RBA formulation is the appropriate selection of a kinetic parameter set from which the CCs can be calculated. In the original model publication (Hameri, Fengos and Hatzimanikatis 2019), 50'000 sets of kinetic parameters were generated for the given kinetic model using the ORACLE framework. As a representative set for our analysis, we decided to use the one that corresponded to the vector of FCCs that was closest to the mean of the FCC distribution across all kinetic parameter sets, with respect to a flux of interest. We chose the uptake of glucose as a case study, as represented by the reaction `GLCptspp` in the model. In order to demonstrate the result variability that this selection of parameters can induce, we additionally selected a number of “extreme” parameter sets through PCA. We used 9 components to describe the kinetic parameter space with respect to the flux of interest, which lead to a coverage of 96.63% of the space variance. We selected the minimum and maximum corresponding parameter sets for each component (2 x 9), leading to a total of 19 sets that were included in our case study. Henceforth, the utilized sets of parameters will be referred to as the reference set, and extreme sets #1-18, each pair of which corresponds to each principal component in the order of sequential maximal coverage of the space. Accordingly, the models constructed through the RBA formulation using each of the kinetic parameter sets will be referred to as the reference model, and extreme models #1-18.

4.3.2.2 Selecting the objective function(s)

In constrained-based problem formulations, optimization of an appropriately selected objective function can lead to a finite number of alternative solutions, or even a unique solution in some cases. The objective function for this problem can be intuitively set as the maximization (or minimization) of the flux/concentration of interest, subject to the set variable bounds, depending on the desired goal of the strategy design. Another rational objective function can be formulated by taking into consideration the notion that the cell will always strive to complete each task with minimal energy consumption, thus with minimal enzymatic resource usage. In that case, the objective can be set as the minimization of the number of parameter alterations in order to achieve a given lower bound on the logarithmic change of the variable of interest. This function can be even further modified to minimize the magnitude of parameter logarithmic changes instead or on top of their number.

4.3.2.3 Defining the bounds of variables

Metabolic engineering modifications of pathways inescapably result in altered reaction rates as well as metabolite concentration levels. RBA, being a constraint-based methodology, allows for the setting of appropriate bounds on these quantities. Reaction fluxes through the network cannot be allowed to increase or decrease infinitely; a substantial increase could lead to undesirable effects on the network, while a value of zero is considered not biologically realistic. Both fluxes and concentrations need to thus be constrained within physiological bounds, conditional to each case study. It should also be noted that the reactions are not allowed to switch their net directionality within this formulation, since the kinetic parameters and in turn the CCs have been computed based on the specific flux directionality profile of the reference steady state.

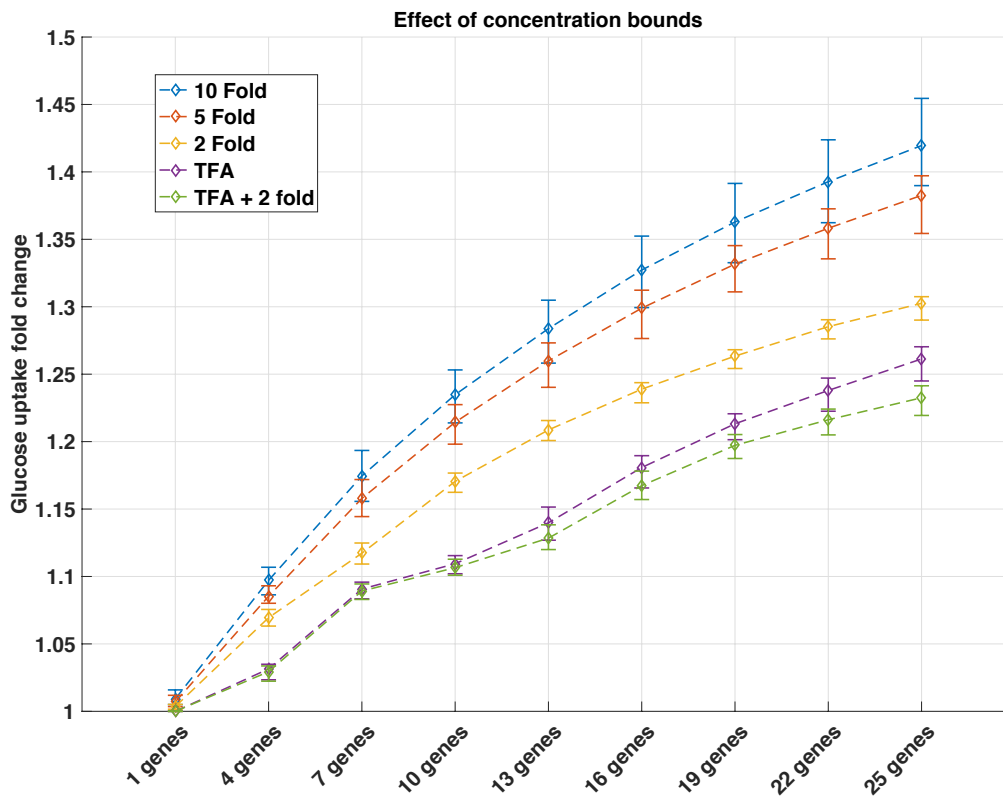


Figure 4.2. Effect of the metabolite concentration bounds on the achievable fold change in glucose uptake for various numbers of genetic manipulations. The flux bounds were maintained flexible at 100-fold in order to study the impact of the concentration bounds on flux predictions. The maximal allowed enzymatic perturbation was 2-fold in magnitude. The whiskers and the diamonds indicate the interquartile ranges and the means of 19 models considered, respectively.

Regarding metabolite concentrations in particular, they should first and foremost obey thermodynamics, as calculated during the thermodynamic feasibility curation of the model. Therefore, we set the logarithmic deviation bounds according to the minimal and maximal allowable concentration values that were calculated using thermodynamic variability analysis (TVA), in order to allow reasonable but consistent flexibility in the metabolite concentration levels. Moreover, these values need to neither be very high nor very low, in order to remain within physiological bounds (i.e. not exceed toxicity levels) and not trigger responses that are unable to be captured by the model (Hatzimanikatis, Floudas and Bailey 1996a). For that reason, we

additionally constrained the concentration deviations to not be more than 2-fold, subject to the thermodynamically permissible values. We assessed the impact of the tightness of concentration bounds on the formulation in Figure 4.2, and note that thermodynamics significantly constrained the predicted achievable fold change of the objective of interest. Looser bounds such a two, five and ten-fold deviation around the reference metabolite concentration state led to significantly larger and most likely biologically unrealistic predictions. Hence, it is crucial to constrain these bounds to physiologically realistic limits through assumptions or metabolomics.

4.3.2.4 Defining the bounds of parameters

In a similar fashion to concentration bounds, severe changes in the manipulated parameter variables could comparably affect the model's predictions. Large overexpression or underexpression of enzymes can significantly influence the organism's growth, or even lead to an excess of toxic external substrates. After applying the physiologically relevant concentration bounds to the RBA model, we compared how the results were affected for allowable enzymatic perturbations of various magnitudes. As the amount of allowable change in enzymatic expression levels increased, we noticed that the predicted change in the glucose uptake reaction rate also increased significantly (Figure 4.3). Moreover, we observed again the positive effect that a larger number of gene modifications had on the flux of interest. It is worth noting that, by increasing the number of gene modifications or the allowable magnitude of the enzymatic perturbation, the predicted fold change of the target reaction flux varied considerably more across the extreme kinetic parameter sets (Figure 4.3 whiskers). Thus, in the interest of accurate calculations it is essential that the enzyme perturbation boundaries are fixed to a physiologically reasonable range.

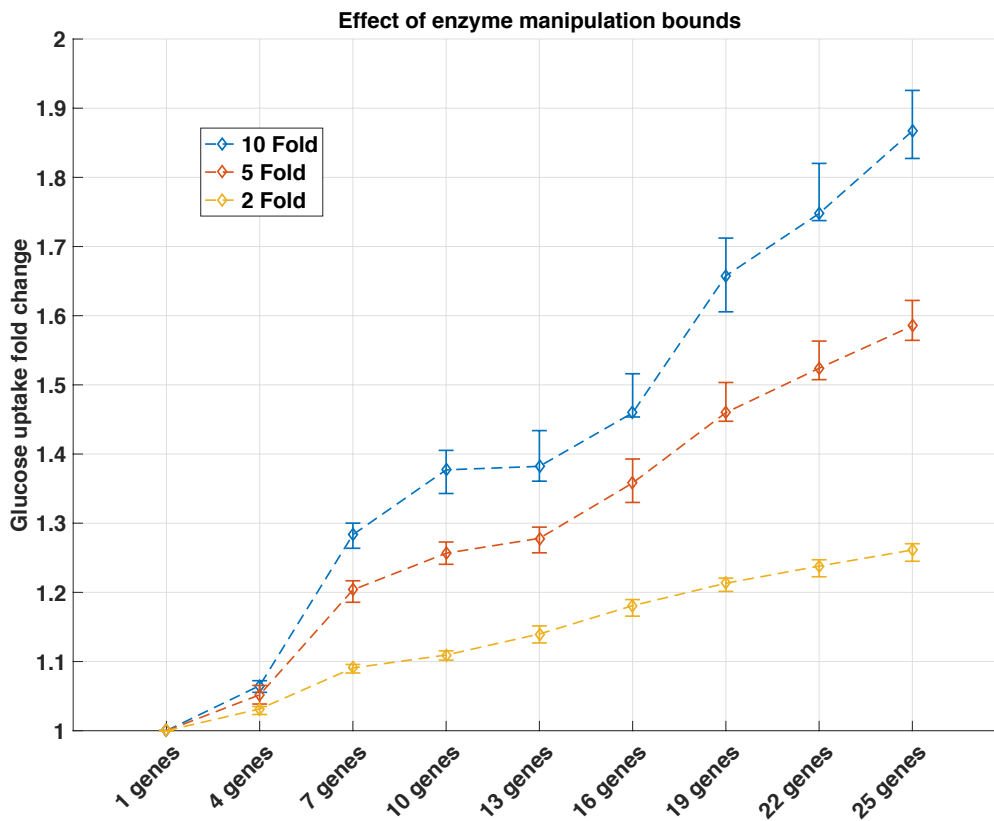


Figure 4.3. Effect of the enzymatic perturbation magnitude bounds on the achievable fold change in glucose uptake for various numbers of genetic manipulations. The flux bounds were maintained flexible at 100-fold in order to study the impact of the enzymatic perturbation magnitude bounds on flux predictions. The concentration bounds were fixed to the thermodynamically feasible calculated values. The whiskers and the diamonds indicate the interquartile ranges and the means of 19 models considered, respectively.

4.3.2.5 Fixing the number of parameter manipulations

The kinetic model encompasses 271 parameters that are possible modification targets. In reality, some cases may require up to 40 or so genes to be engineered in order to achieve a target state. Allowing a larger number of manipulations in the model gives more flexibility to the network, so better objective values can be attained. Figures 4.2 and 4.3 both demonstrated that increasing the number of manipulations has a larger impact on the relevant variable of study. For our case study, increasing the number of manipulations increased the flux through the uptake of glucose in a

logarithmic trend. However, the number of alternative solution sets of parameters given a fixed number of total manipulations can be subject to massive combinatorics. To address this complexity and to focus our analysis in the most prominent parameter candidates, we focused our analysis in the following sections to no more than 10 simultaneous manipulations at a time.

4.3.3 Manipulation-limiting concentrations

Having demonstrated that the bounds of concentration deviations can significantly affect the maximal fold increase of the glucose uptake, we were interested to answer the following question: Which metabolite concentrations are the ones that limit this maximal value? In other words, we wanted to investigate how many, and which, of the concentrations need to “violate” their respectively imposed bounds in order to achieve a higher objective fold increase.

Figure 4.4 shows the effect of allowing a fixed number of metabolite deviations to exceed their bounds has on the maximal fold change of glucose uptake. This analysis was performed for various numbers of parameter manipulations. We observed that in the cases of one and two manipulations, if the concentration bounds remained enforced (zero violations), the flux through glucose uptake could not be modified. However, when we allowed some concentration deviations to exceed their bounds (two and one respectively for the cases of one and two manipulations), this was not the case. For a larger number of parameter manipulations, some fold increase could be achieved even without violating the thermodynamic and physiological bounds imposed. In all cases though, the potential violations pushed the theoretical maximum fold increase of the glucose uptake to higher values.

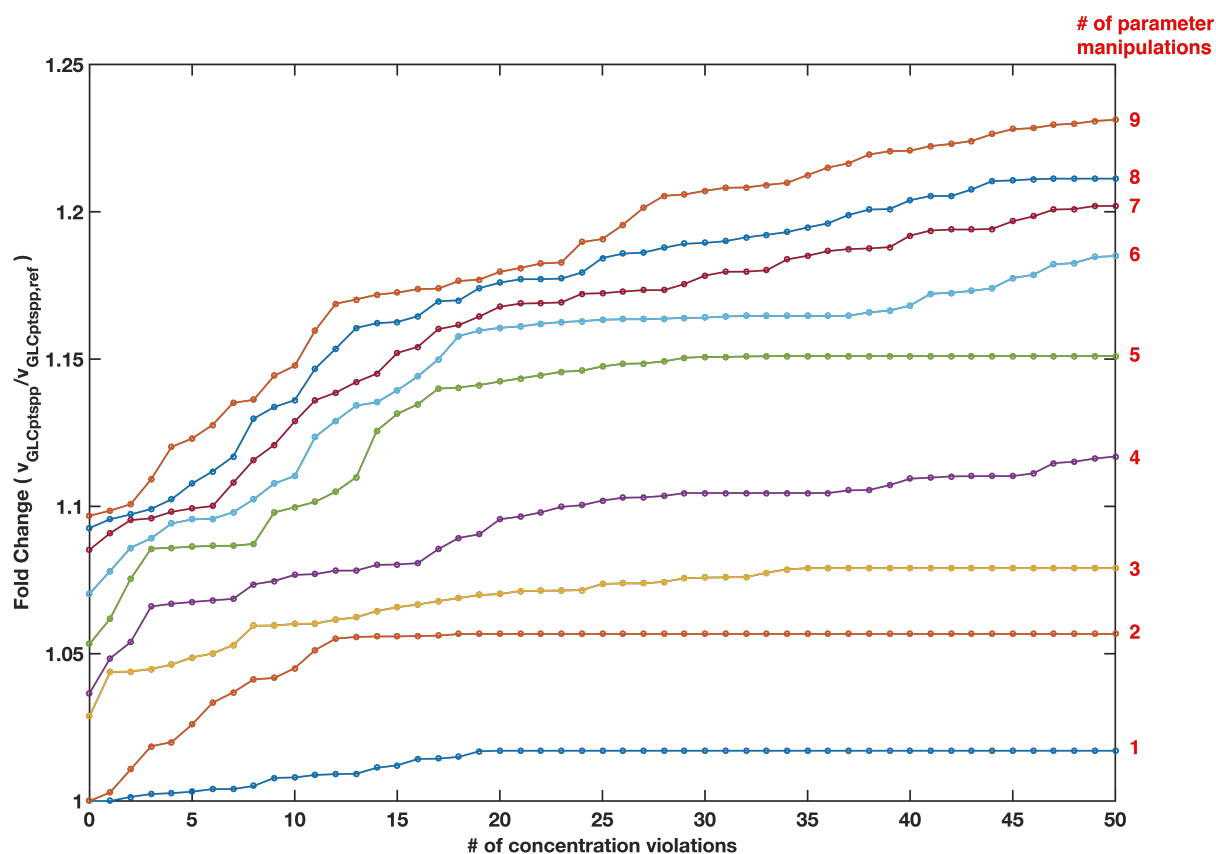


Figure 4.4. Maximal attainable fold increase for glucose uptake with respect to the number of metabolite concentration violations, for a various numbers of parameter manipulations. The flux bounds were maintained flexible at 100-fold, the non-violated concentration bounds were fixed to the thermodynamically permissible plus physiologically feasible bounds (TFA+2-fold), and the parameter manipulation magnitude bounds were set to 5-fold. For all cases only the reference model was used.

Subsequently, we took a closer look at these results, focusing on the case of three maximum violations, for one to nine parameter manipulations (Table 4.1). We noticed that for each fixed number of manipulations, the sets of three metabolite concentrations are unique, and involve 9 total species concentrations that need to be more deviated than allowed by the thermodynamically implied bounds. Irrespectively of the number of parameter manipulations, protons (both in the cytosol and peroxisome cell compartments), AMP, and phenylalanine need to be more positively deviated, while glucose, CTP, dTTP, glutamine, and acetyl phosphate need

to be more negatively deviated in all cases. This analysis revealed that the pH can be a limiting factor to metabolic design, since the concentration of protons depends on the pH value of each compartment. We can additionally use this analysis to focus on each of the other identified molecules and draw hypothesis about their role in the system limitations, and/or potentially investigate these interplays and ways to overcome them *in vitro*.

Table 4.1. Violating metabolite species for the case of three metabolite concentration violations, for a range of parameter manipulations. Each set of three species was unique for each number of parameter manipulations.

		# of parameter manipulations								
		1	2	3	4	5	6	7	8	9
Metabolite concentrations	H ⁺ (cytosol)	X	X	X	X	X	X	X	X	-
	H ⁺ (peroxisome)	X	X	X	-	-	X	-	-	-
	Glucose (cytosol)	-	-	X	X	X	-	-	-	-
	CTP (cytosol)	-	-	-	X	X	X	X	X	-
	AMP (cytosol)	X	-	-	-	-	-	-	-	X
	dTTP (cytosol)	-	X	-	-	-	-	-	-	-
	Phenylalanine (cytosol)	-	-	-	-	-	-	X	X	-
	Glutamine (cytosol)	-	-	-	-	-	-	-	-	X
	Acetyl phosphate (cytosol)	-	-	-	-	-	-	-	-	X

4.3.4 Selecting top candidate strategies

As it has been similarly shown in previous MCA studies on this model (Hameri, Fengos and Hatzimanikatis 2019), we have ranked the FCCs based on their control over the uptake of glucose. For this ranking we used the mean value of the FCC distribution over all the generated kinetic parameter sets. The top 10 enzyme activities controlling glucose uptake in our network and their respective type of control can be found in Table 4.2. This manner of ranking is the most commonly used method to get a rough sense of which parameters in the network affect most the variable of interest. We were thus interested to investigate if the same parameters would be prominently selected as modification targets in the RBA framework.

Table 4.2. Top ranked parameters based on their control over glucose uptake flux GLCptspp. Ranking was computed based on the mean values of 50'000 sets of Control Coefficients.

Parameter Name	Control on GLCptspp
O2tpp	positive
O2tex	positive
CO2tpp	positive
GLCtex	positive
PPC	negative
PGI	positive
GLCptspp	positive
RPI	negative
PFK	positive
G6PDH2r	positive

A big advantage of the RBA formulation over classical MCA, is the ability to generate multiple alternative strategies for a given objective. As an example case, we generated all the possible alternative sets of parameter manipulations, given that the number of manipulations was set to five and the concentration deviation bounds were set to the thermodynamically and physiologically permissible ranges. We performed this analysis across all of the selected 19 sets of kinetic parameters. Firstly, we computed the maximal logarithmic increase of the glucose uptake for each model, subject to the aforementioned constraints. Secondly, we imposed a lower bound to the glucose uptake equal to 95% of this maximal value. Using the reference kinetic parameter set, this maximal fold change was 5.33%. This number varied from 4.70% for the worst-case model (extreme #6) to 7.07% for the best-case model (extreme #14). The amount of alternative sets of five enzymes was found to vary from one unique solution set (extremes #4, 14, 17), to a maximum of 16 (extreme #12). For the case of the reference model we identified four alternative solution sets. The alternative solutions for each extreme model can be found in supplementary Table S4.1.

We could pinpoint two distinct parameters that appeared in all of the generated sets across all models: hexokinase (HEX1) and periplasmic D-glucose transport via proton symport (GLCt2pp). Both enzymes needed to be down-regulated to about 50%, which also happens to the bound we imposed for this calculation. The other parameters that seemed to be crucial for the established objective were pyruvate kinase (PYK), which appeared in all solution sets except two (both of which corresponded to extreme #14), periplasmic D-glucose transport via PEP:Pyr phosphotransferase system (GLCptspp), which is the enzyme catalyzing the objective reaction, and glucose transport via diffusion from the extracellular domain to the periplasm (GLCtex). GLCptspp and GLCtex were interchangeable in the solution sets; they never appeared in the same computed set but one of them was part of every computed set with the exception of three (all of which corresponded to extreme #14). The computed alternative sets with the respective parameter fold changes for the reference model can be found in Table 4.3.

The question that arises in the face of multiple alternative solutions is how to properly rank them and how to select optimal strategies across this vast field. To this end, we propose a ranking based on the biologically relevant assumption that the cell will

strive to utilize the least amount of resources to perform metabolic tasks. Thus, we can classify our alternatives based on the sum of changes in expression levels they would introduce, in ascending order. We have defined this sum of changes as the sum of the absolute values of the logarithmic deviations for each parameter of each solution set. This manner of ranking can be seen in practice in Table 4.3.

Table 4.3. Computed alternative strategies for the increase of glucose uptake. The number of manipulations is set to five, the manipulation magnitude bounds are set to 2-fold, the flux deviation bounds are set to 100-fold and the concentration deviation bounds are set to TFA+2-fold. These calculations were made using the reference kinetic parameter set.

		Parameter deviation $p_k/p_{k,0}$				
		Alternative 1	Alternative 2	Alternative 3	Alternative 4	
Parameters	HEX1	0.50	0.50	0.50	0.50	
	PYK	0.70	0.71	0.71	0.71	
	GLCt2pp	0.50	0.50	0.50	0.50	
	GLCptspp	1.06	1.06	1.06	-	
	G6PP	1.64	-	-	2	
	GLCtex	-	-	-	1.03	
	ATPM	-	1.01	-	-	
	PPK2r	-	-	1.16	-	
		3	1	2	4	Ranking

Most of the parameters that appeared across all of the alternative sets are enzymes that catalyze reactions directly linked to glucose uptake or glycolysis. It is interesting to note that the top ranked parameters from classical MCA (Table 4.2) do not appear in any of the alternative parameter sets. Some of these highly ranked parameters exhibit very large control over multiple fluxes and concentrations across the metabolic network. However, these responses are hindered by the relevant bounds we impose on the system. We can thus conclude that the RBA formulation will favor parameters that have less control over the network, ensuring that cellular balance will not be excessively perturbed. Furthermore, the ability to account for these non-physiological – or non-permissible thermodynamically – responses also presents as one of the main reasons we propose this formulation as superior to the classical MCA.

4.4 Conclusions

In conclusion, the RBA framework enables the consistent and sophisticated design of metabolic engineering strategies using MCA-based control coefficients. RBA is computationally faster and simpler than other approaches since the derivation of control coefficients does not require the numerical integration of non-linear kinetic models. To our knowledge, this type of approach has never been applied to large or genome scale kinetic models of metabolism. Using a previously published large-scale kinetic model of *E. coli*, we demonstrated that the RBA formulation can be applied to large-scale metabolic networks. We used the PCA method to select a number of representative sets of kinetic parameters among their population, in order to effectively represent the uncertainty and flexibility of the kinetic model in respect to parametrization. One of the main advantages of RBA is that, being a constraint-based modeling method, it can accommodate the integration of biologically relevant bounds and constraints, which ensure that the proposed strategies are consistent with the entire system capabilities and limitations thereof. Since the RBA model predictions can be sensitive to the user-defined bounds on the allowable reaction flux, metabolite concentration and enzymatic expression deviations, the importance of including relevant physiological constraints was discussed extensively. Even though our study

4.4 Conclusions

focused on a simpler test case, viable metabolic engineering strategies were shown to be readily derived using this formulation. Alternative solutions can also be generated and evaluated on their efficiency and potential *in vitro* implementation. We believe that this formulation will provide a refined alternative to computational genetic design, due to its simplicity and modularity, and that it will continue to be enhanced through the introduction of ever-growing omics data, and additional specialized constraints and objectives.

Chapter 5 – inverse Response Balance Analysis: a reverse engineering formulation for the mapping of enzymes responsible for diverse phenotypes

5.1 Introduction

5.1.1 Evolutionary biology and regulatory responses

All living beings are a result of a long and complicated evolutionary process, spanning through thousands or millions of years. Even within a short time frame, an organism will adapt to environmental triggers in order to achieve self-preservation and internal regulation of its functions. The elucidation of the evolutionary steps that led to cells as we know them is a vast field of study that has occupied scientists for decades (Archibald 2015, Jensen 1976). Metabolic pathway evolution can be induced by mutations in single-copy gene sequences, gene duplication and ensuing divergence, as well as biochemical noise (Weng, Philippe and Noel 2012). These events, with the addition of natural selection, give rise to enzymatic diversity through specialized enzymes and consequently to increased metabolic efficiency. Promiscuous enzymes have been also implicated in the evolutionary process. It has been postulated that the notion of a “flexible genome” gives the cell a physiological advantage (Greenspan 2001). Additionally, multifunctional and promiscuous enzymes might require very few mutations to improve their potentially superfluous functions, should they become physiologically relevant, and without eliminating their primary, native

activity (Khersonsky and Tawfik 2010, O'Brien and Herschlag 1999). Directed evolution addresses these points and aims to efficiently manipulate and tailor biocatalysts (Lutz 2010). Along with genome editing techniques, the operation of organisms in a modified physiological state has been enabled through metabolic engineering. By studying the internal mechanisms that allow the cell to adjust to each new state, one can hypothesize about the regulatory mechanisms that lead to homeostasis and evolutionary traits.

5.1.2 Mutations in health and disease

On the other hand, even if some metabolic mutations can be beneficial, quite often an organism will develop undesirable mutations which lead to disease. A disease will occur if the function of an essential enzyme is compromised, or if a control mechanism for a metabolic pathway is disturbed. Such cases can be mutations in metabolic enzymes which alter or inhibit their catalytic action, or mutations involving the structure of regulatory proteins or transport mechanisms of metabolites. Nevertheless, a mutation in a single enzyme does not necessitate that an individual will suffer from a disease. Several different enzymes may be able to catalyze the same biotransformation, or alternative pathways may be available to retrieve the same end product through a variety of metabolic intermediates. For the study of human health, it is crucial to identify and diagnose the pathogenic mutations early and accurately in order to provide the necessary treatment. Some of the greatest threats of the universal human health in our days are caused by the disruption of the metabolic homeostatic mechanisms, including obesity, insulin resistance, non-alcoholic fatty liver disease, type 2 diabetes and cardiovascular complications (Hotamisligil 2006). Although these disorders have been associated with genetics and have been closely linked to oxidative stress, nutrition and chronic inflammation, details about their triggers, evolution, prevention and treatment still remain unknown (Heindel et al. 2017, Rani et al. 2016).

5.1.3 Formulations for the study of metabolic control

Mathematical models of metabolism can be of great utility in the elucidation of the cellular function of organisms. Metabolic Control Analysis (MCA) is the mathematical quantification of how changes in enzymatic activities affect the metabolic network in terms of flux and concentration distributions. For the mathematical formulation and further details on MCA and Control Coefficients (CCs), the reader is referred to Chapters 3 and 4. Recently, a method based on MCA and inspired by reverse engineering strategies was proposed (Savoglidis et al. 2016). This method, called Inverse Metabolic Control Analysis (IMCA), makes use of the calculated CCs and aims to translate phenotypic changes in flux and/or concentration levels to changes in enzymatic activities. However, similarly to the classical MCA, IMCA fails to take into account physiological bounds for any changes in the network, or present the ability to generate alternative hypotheses.

Response Balance Analysis (RBA) is a constraint-based method that allows the development of metabolic engineering strategies, as presented in Chapter 4. Using the RBA formulation as a basis, it is possible to introduce additional objectives and constraints which allow for the definition of the reverse engineering problem. Instead of designing a genome editing strategy in order to reach a target state or profile, the aim could be to pinpoint the inferred changes in the genome of a mutant, whether it be a beneficial evolutionary step or a diseased state, given each individual's phenotype.

5.1.4 Aim and scope

We hereby present inverse Response Balance Analysis (iRBA), a reformulation of the RBA workflow, for the mapping of enzymes that are responsible for different phenotypes. We employed two different metabolic models to demonstrate the different capabilities of iRBA. For the first case study we used an *E. coli* model that was consistently reduced from the iJ01366 genome-scale model (Orth et al. 2011b), as presented in Chapter 4. We selected a representative physiological state for the model that exhibited a suboptimal growth phenotype and generated a population of kinetic parameter sets along with their corresponding CCs. We then applied iRBA to

investigate the evolutionary changes in the organism's enzymatic activities that would lead the cell to a phenotype of increased growth rate. For the second case study we used redLips, a *S. cerevisiae* model that focuses on the lipid metabolic network, as presented in Chapter 2. We used experimental concentration measurements of wild-type strains to constrain the model and we generated a population of kinetic parameter sets and their corresponding CCs. We then applied iRBA using experimental concentration measurements of mutant strains to validate our formulation and investigate the interplay of intracellular enzymatic changes the cell employs in order to salvage its phenotype.

5.2 Materials and Methods

5.2.1 Response Balance Analysis

RBA is a constraint-based formulation, in which the mathematical problem is formed through the calculation of CCs from a selected set of kinetic parameters, and the definition of appropriate physiologically relevant bounds, constraints and objectives. The RBA formulation and its full derivation can be found in detail in Chapter 4.

5.2.2 Models used in this study

For the first part of this study, we used a metabolic model of *E. coli* that operates in aerobic conditions (Hameri, Fengos and Hatzimanikatis 2019). As described in Chapter 4, this model was systematically reduced from the iJO1366 GEM (Orth et al. 2011a) with focus on the originally defined subsystems of glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, glyoxylate cycle, pyruvate metabolism and the electron transport chain (ETC), through the redGEM and the lumpGEM algorithms (Ataman et al. 2017b, Ataman and Hatzimanikatis 2017b). The disparity is that in this case the representative steady states for the fluxes and metabolite concentrations were chosen differently; we first defined a suboptimal growth of the cell as a condition to the model, and then we generated 50'000 samples of flux and concentration profiles describing this state. We subsequently selected as

representative states the profiles that were closest to the mean of the sample distributions.

For the second part of the study, we used redLips, a metabolic model of *S. cerevisiae* that focuses on the lipid metabolic pathways, as presented in Chapter 2. The kinetic model version of redLips consists of 1275 reactions and 890 metabolites, and was constrained using wild-type experimental data provided by the Riezman lab (University of Geneva). The construction and curation of this model as well as the generation of kinetic parameter sets and their corresponding CCs was described in Chapter 3.

Both of these kinetic models with their associated populations of kinetic parameters were built using the ORACLE workflow (Miskovic and Hatzimanikatis 2010). Concerning the *E. coli* model, 50'000 sets of kinetic parameters were generated as in the original publication to describe the newly defined phenotype, while for redLips we generated 9'400 sets, as described in Chapter 3. For both cases the corresponding FCCs and CCCs were additionally calculated.

5.2.3 Integrating experimental data of mutant states

To account for the distribution of species in the multiple cell compartments that cannot be captured in experiments, we derived and used sets of constraints that stem from the cell volume percentage of each considered compartment, as discussed in Chapter 3. However, these constraints were derived for species concentrations, while the RBA and iRBA variables are defined as logarithmic deviations of concentrations. Therefore, in order to apply them to our model, we had to reformulate the mathematical expressions accordingly. The reference concentration vector corresponds to the representative concentration profile that was used in ORACLE to build the kinetic model, and satisfies lipidomics measurements of a wild-type state. The target concentration vector corresponds to lipidomics measurements of mutant (Δ LCB3) strains. The lipidomics measurements for both cases were provided by the Riezman lab (University of Geneva). These data can be found in supplementary Tables S3.1 and S4.2, respectively.

The transformed sets of constraints for the iRBA problem are the following:

For the logarithmic concentration deviation $M_j = \ln(x_j/x_{j,0})$ of a metabolite j measured in the range $x_j \in [\underline{x}_j, \overline{x}_j]$ we have:

$$M_j \geq \sum a_i M_{ji}$$

where M_{ji} is the logarithmic concentration deviation of the metabolite j in compartment i , and a_i is the volume fraction of compartment i with respect to the entire cell.

The logarithmic concentration deviation M_j will be subject to the bounds:

$$\underline{M}_j \leq M_j \leq \overline{M}_j$$

where the range values $[\underline{M}_j, \overline{M}_j]$ can be calculated from the mutant strain lipidomics with respect to the concentration reference state as:

$$\ln(\underline{M}_j) = \ln\left(\frac{\underline{x}_j}{x_{j,0}}\right) = \ln(\underline{x}_j) - \ln(x_{j,0})$$

$$\ln(\overline{M}_j) = \ln\left(\frac{\overline{x}_j}{x_{j,0}}\right) = \ln(\overline{x}_j) - \ln(x_{j,0})$$

and the logarithmic reference concentration can be computed from the relation:

$$\ln(x_{j,0}) = \ln\left(\frac{\sum a_i x_{ji}}{\sum a_i}\right)$$

Additionally, each logarithmic concentration deviation M_{ji} will be subject to the bounds:

$$\underline{M}_{ji} \leq M_{ji} \leq \overline{M}_{ji}$$

where \underline{M}_{ji} and \overline{M}_{ji} are calculated as:

$$\underline{M}_{ji} = \ln \left(\frac{x_{ji}}{x_{ji,0}} \right)$$

$$\overline{M}_{ji} = \ln \left(\frac{\overline{x}_{ji}}{x_{ji,0}} \right)$$

and \underline{x}_{ji} and \overline{x}_{ji} are calculated according to the original formulation as:

$$\underline{x}_{ji} = \max \left[\frac{x_j + (a_i - 1) * UB}{a_i}, LB \right]$$

$$\overline{x}_{ji} = \min \left[\frac{\overline{x}_j + (a_i - 1) * LB}{a_i}, UB \right]$$

where LB and UB are the physiological lower and upper bound on intracellular metabolite concentrations, with values $10^{-7}M$ and $10^{-2}M$, respectively.

5.3 Results and Discussion

5.3.1 Part 1 – Reaching optimal growth states

5.3.1.1 Definition of the problem: evolution for optimal growth

Perhaps the most basic aspect in cell life and evolution is growth. Each organism grows and performs its cellular tasks in a specific rate, one that has been the product of many evolutionary iterations. We were thus interested to study the enzymatic changes that could lead into an increased cell growth rate. To this end, we used the above described *E. coli* model to define a suboptimal growth physiology, and generated 50'000 sets of kinetic parameters using the ORACLE framework. Subsequently, we selected as a target physiological state for the iRBA formulation the

reference steady state that was used in the original publication of this model. This state described optimal growth and was also used as a reference state in Chapter 4.

5.3.1.2 Quadratic objective problem formulation

The potential of the cell to reach a certain physiological profile has been a prominent area of study as a metabolic engineering objective. In stoichiometric modeling, the Minimization of Metabolic Adjustment (MOMA) formulation calculates the network flux distribution that is closest to the original wild-type solution following a gene deletion (Segre, Vitkup and Church 2002). Mathematically, the optimization involves the minimization of the distance between the solution and the target state vectors. In order to formulate this concept for the iRBA problem definition, we first calculated the logarithmic deviation of the desired target state $F_{i,obj}$ from the reference steady state around which the sets of kinetic parameters and associated CCs were computed. Then, we defined a quadratic objective function which minimizes the distance between the solution and the target vectors. Hence, a Mixed Integer Quadratic Programming (MIQP) problem formulation was constructed around the original MILP formulation of RBA (Hatzimanikatis, Floudas and Bailey 1996a, Hatzimanikatis, Floudas and Bailey 1996b). The quadratic objective function can be written as follows:

Minimize:

$$\sum_i^{\#Fluxes} \left(\frac{F_i - F_{i,obj}}{F_{i,obj}} \right)^2$$

Subject to:

$$K \cdot u = 0$$

where, K is the RBA matrix, and u is the vector of logarithmic deviations composed of F , M and E as defined in Chapter 4 (Figure 4.1).

5.3.1.4 Reaching optimal growth – Number of enzymatic adjustments

After constructing the RBA matrix for the selected kinetic parameter sets and formulating the objective function, we were interested to examine how the model's performance was affected by the allowable number of enzymatic manipulations, or in this case enzymatic adjustments. We used the Euclidean distance between the solution and target vectors as a metric to test the performance of our system. Knowing that the target flux profile was thermodynamically feasible in our original network, thus theoretically reachable in an unconstrained system, we were also able to evaluate the effect that the imposed variable bounds would have on the solution. In Figure 5.1 the computed Euclidean distance between the two vectors is presented, as calculated for all of the 19 considered sets of kinetic parameters for various numbers of enzymatic adjustments. We observed that the distance between the two vectors is decreasing when a larger number of enzymatic adjustments is considered. This decrease is monotonic for the median of these distributions (red lines). We also noticed that for an even larger number of enzymatic changes (not depicted), the minimal Euclidean distance reaches a plateau, most probably because of other constraints that were imposed to the system. The bounds for the flux, concentration and enzymatic deviations were all set to 100-fold, a very wide range, in order to evaluate clearly the impact of the objective function to the formulation. These results lead us to the conclusion that in order for the cell to restructure its network enough to be able to reach the selected target physiology, a large number of genetic adjustments have to take place. Additionally, if we impose a narrower range for variable bounds in the model, which would accurately represent physiological constraints, instead of the wider range we used previously, the target state will not be reachable within "one step" of metabolic restructuring, but it would require multiple iterations.

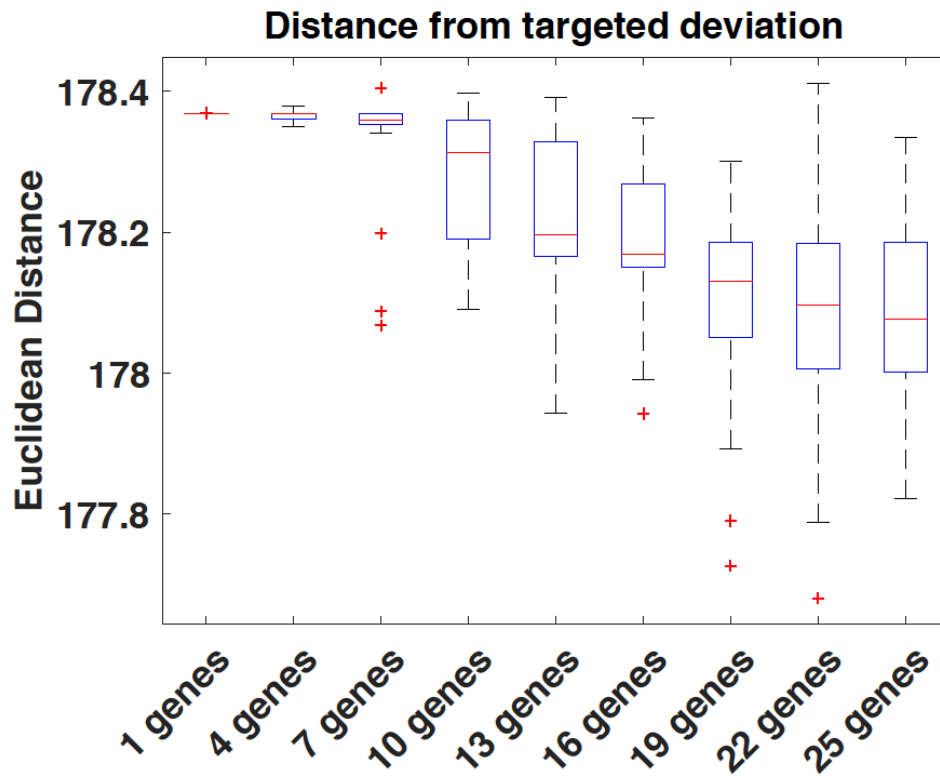


Figure 5.1. Euclidean distance from target deviation vector, for various numbers of enzymatic manipulations across the 19 selected sets of kinetic parameters. For each number of manipulations, the red lines denote the median solution value, the edges of the blue boxes the upper and lower quantiles, the black whiskers extend to the full range of solution values excluding outliers and the red crosses denote the outliers. The bounds for flux, concentration and enzymatic activity deviations are all maintained flexible at 100-fold.

5.3.1.5 Reaching optimal growth – Metabolic and genetic restructuring

Following the evaluation of the objective function, we imposed physiologically relevant bounds and constraints to the model. As discussed in Chapter 4, we set the bounds for concentration deviations to the maximal allowed thermodynamically permissible ones as calculated from Thermodynamics-based Flux Analysis (TFA), subject to a maximum of a two-fold change for each species (referred to as TFA&2-fold bounds). Flux deviations were lightly constrained at 100-fold and the enzymatic

activity deviation was set to 5-fold with a set number of 25 maximal adjustments. As a reminder, the target profile of the quadratic optimization was a flux profile of increased cell growth, which corresponds to the optimal cell growth predicted by the original model. Using the representative set of kinetic parameters to construct the RBA matrix, the resulting optimal solution can be seen graphically in Figure 5.2. The achieved state was overall close to the targeted optimal growth state in terms of Euclidean distance. The targeted growth rate was achieved with less than 1% difference from the targeted deviation (not depicted). The mean Euclidean distance across all reactions from the objective was 0.77, a number that shows that the majority of the network's reactions were successfully pushed to the objective state. The targeted main carbon flow through glycolysis, PPP, and the TCA cycle was very well reached, while the oxidative phosphorylation and alternate carbon metabolism pathways were not as well reached. In particular, the flux deviation furthest from the associated objective was NADH9, one of the network's NADH dehydrogenase reactions. A hypothesis for this inability is that NADH9 is a highly coupled reaction within the model, performing the same function as multiple other reactions with a specific set of electron carriers (demethylmenaquinone-8 versus menaquinone-8 or ubiquinone-8). This reaction has also been reported to be part of a splitting flux reaction pair in a previously published *E. coli* GEM (Feist et al. 2007). Thus, it is possible that the network cannot fine-tune all of these codependent reactions to the desired value simultaneously and in order to achieve the minimum distance for most of them, one of them needs to be further away from the target. The other reaction that was similarly far from the target state is ALDD2y (aldehyde dehydrogenase). This function is also performed by ALDD2x with a different cofactor pair (NAD⁺/NADH versus NADP⁺/NADPH). Additionally, it is highly likely that the physiological constraints we impose on the network inhibit a perfect achievement of the objective. We observed that limiting the number of enzymatic activity adjustments leads to an increased Euclidean distance from the target state (Figure 5.1). Moreover, the thermodynamic feasibility bounds on the concentration deviations and physiologically relevant bounds on enzymatic activity deviations inflict limitations to the metabolic adjustment capabilities of the cell as extensively discussed in Chapter 4.

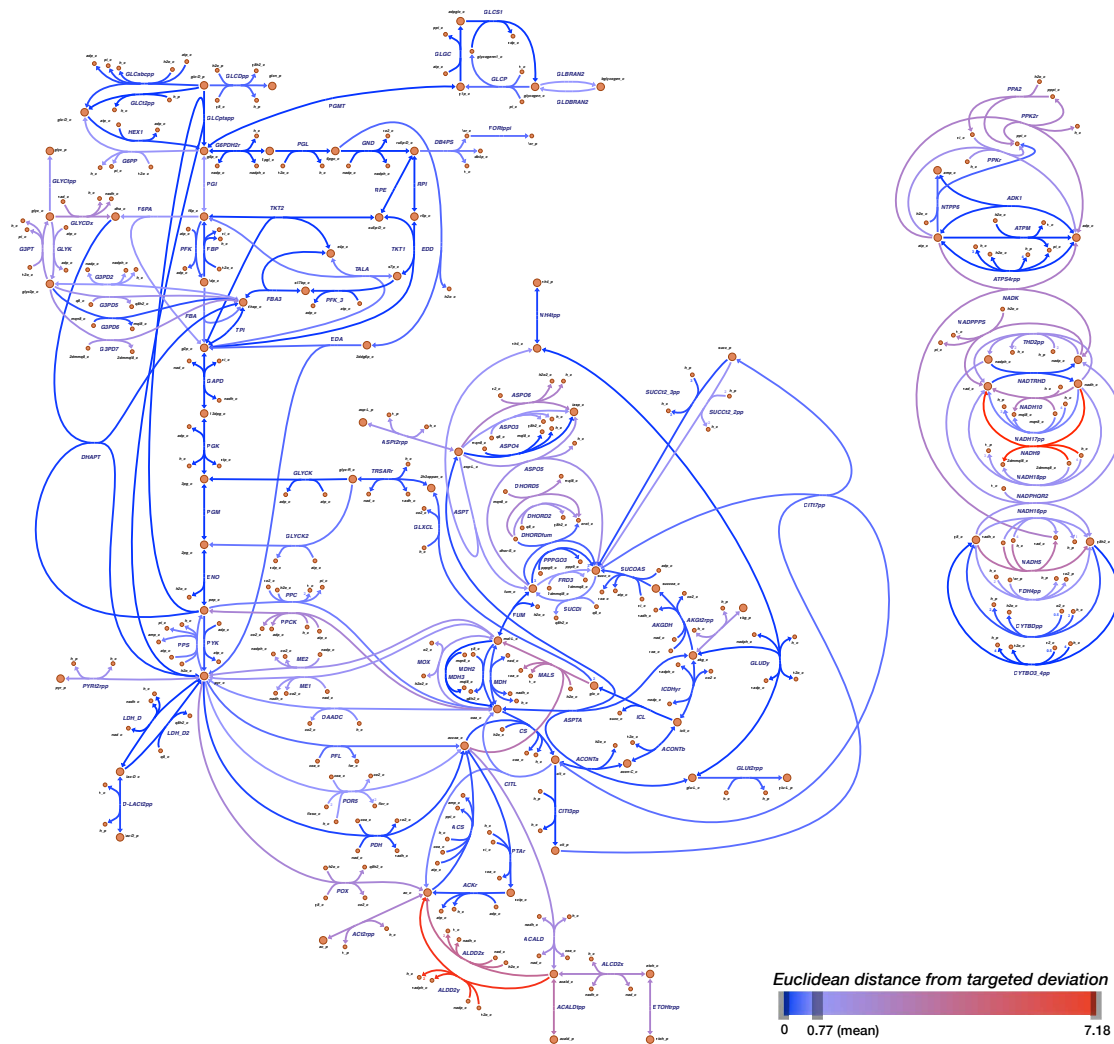


Figure 5.2. Euclidean distance from the target deviation (optimal growth) for each reaction. The bounds for flux, concentration and enzymatic activity deviations are set to 100-fold, TFA&2-fold and 5-fold, respectively.

Concerning the precise enzymatic adjustments computed for the attainment of the target state, most of the predicted enzymes catalyze reactions belonging to the central carbon metabolism, such as glycolysis, TCA cycle, and Pentose Phosphate Pathway. Additionally, some inner membrane transporters were predicted, and the overall fold

changes of the enzymatic activities ranged from 0.15-fold to 1.5-fold, a much lower value than the 5-fold bound that we imposed. This led us to believe that the concentration deviation bounds are the ones that limit the optimization, which is consistent with the statements of the original formulation publications (Hatzimanikatis, Floudas and Bailey 1996a, Hatzimanikatis, Floudas and Bailey 1996b). It should also be noted that the majority of adjustments were downregulations, but no correlation with assigned reaction subsystems could be discerned.

We also constructed the RBA matrix using each of the 18 selected extreme sets of kinetic parameters and performed the same analysis. Remarkably, the network reactions' Euclidean distances from the targeted state for all of these models compared to the ones from the representative model were almost identical. Some minor variations were observed among some of the reactions belonging to the oxidative phosphorylation pathway and to an even lesser extent on the rest of the network.

By examining the predicted enzymatic adjustments across these models, we saw a similar trend as for the representative one, meaning that the majority of the enzymatic adjustments involved enzymes that catalyze reactions from the central carbon metabolism, as well as a large number of enzymes catalyzing reactions from oxidative phosphorylation, and inner and outer membrane transporters. The total number of distinct enzymes appearing across all the 19 models studied was 116 out of the 271 that are included in the network, and only four out of them were predicted to both increase and decrease their activities between models. These enzymes along with the associated statistics are shown in Table 5.1.

We can thus discern that the selection of a kinetic parameter set does not contribute to the predicted genetic and metabolic restructuring of the network in a large degree. It is however possible that more differences could become evident between the 19 examined models' predictions if we constrained further the number of enzymatic adjustments. We should also note that alternative solutions that yield the same optimal minimum distance from the objective are possible for all of the RBA models, though not explored in this study.

Table 5.1. Computed enzymatic adjustment statistics for the 19 selected models: Number of models for which each enzyme is predicted to have altered activity, and the type of enzymatic activity regulation. The names of the enzymes are reported in terms of the reactions they catalyze. Boldface denotes enzymes that are part of the predictions of the representative model.

Enzyme	# of models	Regulation Type	Enzyme	# of models	Regulation Type
CRNt8pp	18	up	NADPHQR2	4	up
TALA	16	down	RPI	4	down
PGI	15	down	SUCt2_3pp	4	down
GLXCL	14	up	THD2pp	1 / 3	up / down
PPC	13	up	THRt4pp	4	up
NADH18pp	11 / 1	up / down	ASNtex	3	down
GLYCLTt2rpp	11	up	ASPt2rpp	3	up
LDH_D	11	down	CA2tex	3	up
EDD	10	down	CLtex	3	up
MG2t3_2pp	10	down	CO2tpp	3	down
PYK	10	up	CYTBDpp	3	down
THRt2pp	9	up	FDH5pp	3	up
ACtex	8	down	FUM	3	down
GLBRAN2	8	down	GLCptspp	3	down
GLCDpp	8	up	GLYCLTtex	3	up
MOX	8	down	MALtex	3	up
POX	8	down	MDH2	3	up
PPS	8	up	ME2	3	down
TRSARr	8	up	TYRtex	3	down
ASPO3	7	down	VALtex	3	down
CITt3pp	7	down	AKGDH	2	down
GLYctpp	7	up	COBALT2tex	2	up
PFK_3	7	up	DHORD5	2	up
ACKr	6	down	DHORDfum	2	up
CITtex	6	up	EDA	2	down
FE3tex	6	up	F6PA	2	up
SUCDi	6	up	FDH4pp	2	down
Zn2tex	6	up	FORtppi	2	down
ASPTA	5	down	G3PD7	2	down
DHAPT	5	up	GLCtex	2	down
DHORD2	1 / 4	up / down	GLUDy	2	down
FBA	5	down	GLUt2rpp	2	down
FLDR2	5	up	LEUt2rpp	2	up
G6PP	5	down	MALS	2	down
MDH3	5	down	MDH	2	down
O2tpp	5	down	NADPHQR4	2	up
PFL	5	down	NADTRHD	2	down
SO4tex	5	down	NI2tex	2	up
CU2tex	4	up	O2tex	2	down
CYTBO3_4pp	4	up	PGK	1 / 1	up / down
GLYCLTDx	4	down	TKT1	2	down
HIStex	4	up	TKT2	2	up
NADH16pp	4	down	TPI	2	up
			AKGt2rpp	1	up
			AKGtex	1	up

5.3 Results and Discussion

<i>ALDD2y</i>	1	up
<i>ASPO5</i>	1	down
<i>DHAtp</i>	1	down
<i>FBP</i>	1	up
<i>FE2tp</i>	1	up
<i>FORtex</i>	1	down
<i>G3PD2</i>	1	up
<i>G3PD5</i>	1	down
<i>G3PT</i>	1	down
<i>GAPD</i>	1	down
<i>GLCNtex</i>	1	up
<i>GLCt2pp</i>	1	down
<i>GLYCK</i>	1	up
<i>GLYt2rpp</i>	1	up
<i>GND</i>	1	up
<i>Htex</i>	1	up
<i>ILEtex</i>	1	up
<i>ME1</i>	1	up
<i>MNT2pp</i>	1	down
<i>NADH10</i>	1	down
<i>NADK</i>	1	down
<i>PDH</i>	1	down
<i>PGM</i>	1	down
<i>POR5</i>	1	down
<i>PPPGO3</i>	1	down
<i>PROt2rpp</i>	1	up
<i>SUCOAS</i>	1	up
<i>ZN2tp</i>	1	down

5.3.1.6 Reaching other phenotypes

Another interesting aspect of physiology, and one that we have the ability to investigate through the iRBA formulation, is the capability of the cell to operate in various different physiological states, and the underlying regulatory mechanisms leading to each of them. A prime example of these states and their importance was discussed in Chapter 3 regarding the construction of kinetic models. In a metabolic model, the capabilities of the organism to operate in various physiological states can be captured by the extensive sampling of the flux and concentration distributions, preceding the selection of the representative steady state profiles. Even within one Flux Directionality Profile (FDP), very different flux profiles can be attained in the sampled solution space. For this study, we generated 50'000 flux profile samples, as described in the Materials and Methods section. PCA was used as a tool to navigate this sample population and select some “extreme” physiological states amongst them. The first component used to decompose the sample space resulted in a coverage of 70.27% of the total space variance. Across this component, the fluxes that contributed most of the space variance were found to be mostly related to intermembrane transports between the periplasm and cytosol, such as L- or D-carnitine and formate, and to oxidative phosphorylation. We subsequently selected the minimum and maximum corresponding samples for this component, leading to the selection of two

“extreme” steady state flux profiles for which we could perform our analysis. For this analysis we used only the reference RBA model. The bounds for the flux, concentration, and enzymatic activity deviations were set to 10-, 2-, and 5-fold, respectively, and the number of enzymatic adjustments was set to 25.

For both cases of extreme target states, the reactions that exhibited the largest Euclidean distance between their achieved and objective deviation were almost identical. More specifically, 11 out of the 15 reactions that had the largest distance for each of the two cases were common. Most of these reactions belong to the oxidative phosphorylation pathway or are transports of species across the inner or outer membranes of the cell. Interestingly, the reactions that contributed most to the space variance as calculated by PCA, could not be well-reached in either of the two cases. Concerning the calculated enzymatic adjustments, six out of the 25 enzymes were common for the two cases, with similar values of deviation predicted. These results can be found in supplementary Table S5.1.

We observed that the flux deviations that were not well-reached for both the cases of targeted optimal growth state and targeted extreme states were in the same parts of the metabolic network. This led us to the hypothesis that the network parameters cannot exert much control over these reactions. By looking at the computed CCs corresponding to these fluxes, this hypothesis was confirmed for the majority of them. Another explanation for these results could be that the imposed variable bounds are too tight, and the target deviation for these reactions cannot be reached due to these constraints. However, these bounds were defined based on physiologically relevant assumptions, thus the previously stated hypothesis that the target state is only reachable through multiple iterations of genetic and metabolic restructuring in the network is strengthened.

5.3.2 Part 2 – Mutant identification and mapping

5.3.2.1 Definition of the problem: identification of mutants

In *in vitro* studies, the flux and concentration distributions within the cell cannot be fully measured. It is quite common, especially in health-related disciplines, that conclusions regarding undesirable mutations need to be drawn from limited data. iRBA can be used effectively to identify underlying changes in enzyme expression levels, given experimental measurements of flux and/or concentration discrepancies between the diseased and healthy states of an organism. When omics data are available, there is an associated variability for each reported value, stemming from the multiplicity of individuals and/or measurements, experimental conditions, and instrument inaccuracy. There are two possible options to account for this variability in the iRBA formulation. The first option is to consider only the mean (or median) value of the measurements, and proceed with the quadratic formulation as demonstrated in the above sections. The second option is to define additional constraints that capture the data variability, as well as another appropriate objective function. In the interest of demonstrating the method and its advantages, we chose to proceed this study using the second option.

5.3.2.2 Experimental data for Δ lcb3 mutant strains

The data that we used in this study were lipidomics measurements from both wild-type and mutant yeast cells, provided by the Riezman lab (University of Geneva). These concentrations corresponded to 28 lipid species, and the wild-type data were already used to curate and precondition the redLips network prior to building the kinetic model, as described in Chapter 3. The mutant data originate from a Δ lcb3 mutant strain. LCB3, a long-chain base-1-phosphate phosphatase, has been known to play a part in the regulation of ceramides and long-chain base phosphates. Our goal was to use these data in the iRBA formulation in order to shed light to the underlying enzymatic perturbations that are associated with this mutation. We thus incorporated the lipidomics to the network, using the sets of constraints that were defined in the Materials and Methods section.

5.3.2.3 Selection of variable bounds and objective function

The variable bounds that were chosen for this study were 100-fold for the flux deviations and TFA&2-fold for the concentration deviations. Concerning the enzymatic activity deviations in this study, we imposed non-symmetrical bounds. Since we were examining a case involving gene knockouts, a large downregulation of the enzymes needed to be permitted, but not the corresponding upregulation. Thus, the lower and upper bounds were set to 1000-fold and two-fold, respectively.

Additional to the metabolomic constraints, we needed to define an appropriate objective function for the optimization problem. Since we were interested to study the intracellular responses to stress factors such as gene deletions, we made the physiologically relevant assumption that the cell will strive to compensate for the induced perturbation by introducing the least possible regulatory expression changes. To this end, we selected the minimization of the number of manipulations over the network as the optimization objective.

5.3.2.4 Mapping the mutant phenotype to enzymatic changes

Using the reference kinetic parameter set, the solution of the above defined MILP formulation results in the identification of 171 enzymatic activity changes as the optimal minimum number of adjustments needed to satisfy the imposed bounds and constraints. The enzymatic activity of LCB3 is indeed predicted to be downregulated, serving as validation of the constructed network. However, we identified few alternative solutions where the activity of LCB3 remained unchanged. This stemmed from the associated uncertainty that comes with models of this size and complexity, and shows that in order to produce accurate hypotheses even more data would need to be incorporated in the model as constraints. Nevertheless, we chose to analyze a case where LCB3 was downregulated at the maximal computed value, and investigate the effect that it had over the rest of the network. To this end, we constrained LCB3 to the value mentioned above, and performed a variability analysis over all the enzyme activity deviations of the network. We observed that the vast majority of the enzyme deviations could be both up- and down-regulated, leading to a very large number of combinations and alternative solutions, and further strengthening the argument that

more constraints might need to be introduced to the model. However, there was a number of enzymes that had to be deregulated in a certain direction. More specifically, there were nine enzymes that were predicted to be always downregulated, additional to LCB3. These enzymes were the following:

- sphingoid long-chain base kinase (LCB4), which synthesizes sphingolipid backbone long-chain base phosphates,
- dihydrosphingosine phosphate lyase (DPL1), which degrades sphingolipid long-chain base phosphates,
- fatty aldehyde dehydrogenase (HFD1), which converts fatty acid aldehydes into fatty acids,
- fatty acid elongase (ELO3), which synthesizes very long chain fatty acids from acyl-CoA primers,
- peroxisomal ABC transporter (PXA1), which imports long-chain fatty acids into peroxisomes,
- phospholipid methyltransferase (OPI3), which catalyzes the last two steps in PC biosynthesis,
- lysophospholipid acyltransferase (ALE1), which attaches fatty acyl chains to monolysophospholipid species,
- serine hydroxymethyltransferase (SHM1 or GHMT2r), which converts serine and tetrahydrofolate to glycine and methylenetetrahydrofolate respectively, and
- trehalase (ATH1/NTH1 or TREH), which degrades trehalose to form two units of glucose.

Some of these enzymes are directly related to the biosynthesis or degradation of sphingolipids, such as LCB4 and DPL1. LCB4 is in essence the reverse counterpart of LCB3; the former phosphorylates the sphingolipid backbones using ATP as a phosphate donor, while the latter dephosphorylates them through hydrolysis. Since in the RBA and iRBA formulations the reactions are not allowed to change their net directionality, it is logical that if one part of the phosphorylation cycle would be impaired, the other half would be as well. DPL1 performs the degradation of the long-chain base phosphate species into ethanolamine phosphate and fatty aldehydes or fatty acids. Since the synthesis and degradation of the phosphorylated species would

be damaged in the mutant strain, the function of DPL1 would also be reduced. Accordingly, HFD1, an enzyme that utilizes fatty aldehydes as substrate, was also predicted to have a lower activity in the network. Seeing as the regulation of the production of sphingolipid backbones would be heavily compromised, the regulation of other sphingolipid building blocks would be equally impaired, as proven by the ELO3 reduction of activity. ELO3 is the enzyme responsible for the synthesis of very long chain fatty acids (C24:0 and C26:0), which are the ones used for the synthesis of complex sphingolipids. The activity of the PXA1 transporter of these species to the peroxisomes was also predicted to be impaired. Furthermore, the lipidomics that were used to construct the mutant concentration deviation constraints clearly showed a deregulation of phospholipid concentration levels and ratios. Two enzymes that are part of the phospholipid biosynthetic pathways were identified in the solution: OPI3, responsible for the synthesis of PC from PE through the *de novo* biosynthetic pathway, and ALE1, a key component of membrane phospholipid remodeling. Finally, the conversion of serine to glycine through SHM1 (GHMT2r in model) and the degradation of trehalose to glucose through ATH1/NTH1 (TREH in the model), were also predicted to be compromised. Serine is a major building block in the synthesis of long chain bases, participating in the first committed step of the sphingolipid biosynthetic pathway, which is the condensation of serine with a long chain acyl-CoA to form 3-ketosphinganine catalyzed by LCB1.

This study revealed some of the underlying enzymatic adjustments that would occur in a state of mutation. Using mutant lipidomics, the iRBA formulation could predict correctly the downregulation of the enzyme corresponding to the deleted gene in these mutants, validating its predictive capabilities. Using this method of analysis, the computed genetic and metabolic restructuring of the cell could be mapped to specific mutations and provide potential biomarkers for disease diagnosis.

5.4 Conclusions

In conclusion, the iRBA reformulation of the RBA framework is a powerful tool in the mapping and identification of underlying enzymatic adjustments in an evolutionary, survival and mutant context. iRBA uses a quadratic objective function instead of a linear approximation, and to our knowledge, this type of approach has never been applied in reverse engineering studies of metabolism of this scale. We used a previously published large-scale kinetic model of *E. coli* to identify the enzymatic restructuring of the cell between states of suboptimal and optimal growth rates. We selected a number of kinetic parameter sets among the generated population by utilizing the PCA method, in order to efficiently represent and quantify the associated uncertainty and flexibility of the kinetic model. We identified specific parts of the metabolic network which could not be efficiently restructured to achieve the target state, given the physiological bounds and constraints that were imposed to the model. Additionally, we examined the adjustments that would lead to some extreme flux profile states, which were identified through the use of extensive sampling of the flux solution space of the original model and PCA. Finally, we used a large-scale kinetic model of *S. cerevisiae* that focuses on the lipid metabolism to define a workflow for the identification of cell mutations. We integrated a number of mutant lipidomic measurements to the model, along with the biologically relevant bounds and constraints which were defined in the original RBA formulation. We showed that the model could predict correctly the gene deletion responsible for the mutation, and explored the underlying genetic and metabolic restructuring. We believe that this formulation will provide a useful means toward the elucidation of evolutionary mechanisms that lead to desirable traits, as well as advance the mapping of altered metabolomes to enzymatic perturbations, which could be applied in disease diagnosis and personalized medicine.

Chapter 6 – Conclusions and Future Perspectives

In Chapter 1 we examined the current state of the metabolic engineering and systems biology fields. We identified key issues that hinder the study of cellular regulation, with particular emphasis on the lipid metabolism. Genome-scale metabolic models, although being the most accurate organism representations available, possess a high degree of complexity and introduce large uncertainty in studies that focus on specific parts of the metabolic network. GEM reductions are usually done in an *ad hoc* manner to produce context-specific models, and cannot serve multiple studies since they are limited to specific datasets. We additionally discussed the importance of kinetic models in the elucidation of regulatory mechanisms and cellular responses to perturbations. The mentioned complexity of GEMs combined with the lack of information about detailed enzymatic mechanisms and kinetic parameters turn the construction of consistent large-scale kinetic models into a challenging task. Furthermore, we assessed the present-day methodologies concerning metabolic design. Although cell factories are widely used in industrial manufacturing, most computational tools fail to account for physiological and technical limitations when proposing metabolic engineering strategies. Concerning the field of medicine, the discovery of potential biomarkers and the mapping of various phenotypes to genetic mutations are significant steps in treatment and diagnosis. Mutations are also a vital part of evolution. Directed evolution aims to selectively preserve desirable traits through multiple cell generations, mimicking natural selection. However, state-of-the-art computational methodologies fail to examine the enzymatic and metabolic restructuring that occurs at the onset and progression of mutation.

In Chapter 2 we presented the construction of redLips, a novel metabolic model that focuses on the lipid metabolism. This model captures the complexity of lipid metabolism by consistently assembling the known lipid pathways while avoiding the pitfall of excessive size and detail that GEMs possess. We first gathered all the available knowledge on lipid metabolic pathways, and integrated them into a well-known GEM of *S. cerevisiae*. We subsequently reduced the integrated model around the lipid-related network subsystems using the redGEM and lumpGEM frameworks. redLips was curated with thermodynamic feasibility constraints and is consistent with the organism biochemistry. Additionally, it can accommodate the integration of lipidomics measurements, applied both as flux and concentration bounds. We believe that in the future redLips could be modified and used as a concise platform for studying lipid metabolism across different species, and a valuable tool for health or industry related research. Yeast has a very high homology with the mammalian genome, and could prove advantageous for the testing of hypotheses and treatments. Yeast is inexpensive to cultivate and modify, and scientifically and ethically easier to use for this type of studies. We also trust that this model will continue to accommodate future discoveries through the incorporation of new reactions and species. Additionally, redLips provides a coherent base in order to link cell signaling routes to the lipid pathways. Signaling cascades have a big part in the cell's regulatory mechanisms. Membrane homeostasis has been closely linked to phosphatidylinositol phosphate species, which are already part of the redLips network.

In Chapter 3 we used redLips as a basis to construct a large-scale kinetic model of the lipid metabolism. This model is to our knowledge the largest and most detailed kinetic representation of the lipid network to date. The enzymes which are linked to lipid pathways are usually multifunctional and promiscuous, which gives rise to enzymatic coupling. The notion of apparent inhibitors was used to allow the definition and assignment of appropriate enzymatic mechanisms. We used lipidomics data to curate the model, by taking into account the compartmentalization of species and the relative volume of each cellular compartment. We generated populations of kinetic parameter sets using the ORACLE framework, and calculated the corresponding MCA-based control coefficients. We demonstrated the importance of the enzymatic coupling consideration in the model's predictions, compared to the corresponding

predictions when using simpler mechanisms. Finally, we validated the model's predictions against experimental observations, and discussed emerging regulatory interactions. We believe that the systematic workflow presented in this chapter can be used as a guideline in the construction of consistent kinetic models of metabolism. In the future, even more complex kinetic mechanisms can be utilized in this model, such as the action of inhibitors and activators in multiple reactions. Kinetic parameters such as Michaelis constants are slowly becoming available, and their incorporation in the models could lead to a significantly reduced uncertainty. Furthermore, given the already generated parameter sets, the resulting system of ordinary differential equations can be built and integrated over time in order to study transient system behaviors.

In Chapter 4 we presented Response Balance Analysis (RBA), a constraint-based framework that uses Metabolic Control Analysis (MCA) notions and control coefficients (CCs) to facilitate the consistent and sophisticated design of metabolic engineering strategies. To our knowledge, this type of approach has never been applied to large-scale kinetic models of metabolism. We discussed and highlighted the integration of biologically relevant bounds and constraints, which ensure that the proposed strategies are consistent with the network capabilities and limitations thereof. We additionally showed that multiple viable metabolic engineering strategies can be readily derived for large-scale metabolic networks. We believe that this formulation is a concise and refined tool for computational metabolic design, due to its simplicity and modularity. RBA incorporates constraints regarding not only physiology, but also genetic engineering limitations, which could greatly benefit the design of cell factories. In the future, RBA can continuously be enhanced through the introduction of ever-growing omics data. Additionally, as the community's knowledge grows, new metabolic models will emerge and specialized constraints and objectives will become relevant. As new methods for kinetic modeling are constantly developed and kinetic parameters are measured and reported, the dynamic predictions and corresponding CCs will be able to capture even more accurately the interactions of cellular components.

In Chapter 5 we reformulated the RBA framework that was developed in chapter 4. This method, called inverse RBA (iRBA), aims to facilitate the mapping and

identification of underlying enzymatic adjustments in an evolutionary, survival and mutant context. We examined and discussed the enzymatic and metabolic restructuring of an *E. coli* cell between different states of growth. Furthermore, we used the large-scale kinetic model of lipid metabolism that was developed in Chapter 3 to define a workflow for the identification of underlying cell mutations and the mapping of phenotypes. We integrated lipidomic measurements to the model, along with the biologically relevant bounds and constraints which were defined in the original RBA formulation. We believe that this formulation can provide useful insights for reverse engineering problems, as well as a step forward in the mapping of altered metabolomes to enzymatic perturbations. In the future, genetic mapping could prove essential in biomedicine. The scientific interest in the identification of biomarkers is rapidly growing, with multiple applications in early diagnosis and treatment. Personalized medicine could also benefit from a tool like iRBA; each day millions of people take medication that will not benefit them (Schork 2015). Genetic and environmental factors can vary quite a lot across the human population and lead to different metabolic patterns and phenotypes, including reactions to drugs. iRBA can take into account the individuality of each patient by tracing these phenotypes to genetic causalities. The mapping of these individualities combined with computational design and simulation of potential treatments could shape future advancements and perspectives in the field.

Overall, this thesis has contributed to the study of lipid metabolism through the construction and validation of consistent stoichiometric and kinetic models. Although these models describe the metabolism of *S. cerevisiae*, they can accommodate studies in other organisms due to the high homology of eukaryotic genomes concerning the lipid pathways. Moreover, a framework was established for the consistent design of metabolic engineering strategies in sustainability and health, considering user-defined limitations on both the biological and engineering aspects of the design. Each of these tools can incorporate various types of omics data, strengthening the bond between computations and experiments, as is the interdisciplinary nature of systems biology.

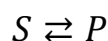
Appendix A - Kinetic mechanism expressions

In all of the following expressions S and P denote substrates and products respectively, C_s and C_p denote small molecules acting as compensating substrates and products, respectively, without binding to the active site of the enzyme, I denotes apparent inhibitors, V_{max} denotes the maximum rate of the reaction in the respective direction (f signifies forward and b signifies backward), K_m and K_i denote the Michaelis constants for the reactants and inhibitors respectively, and k denotes the rate constant in the respective direction.

A.1 Michaelis-Menten kinetics

A.1.1 Uni-Uni

For a reaction in the form:

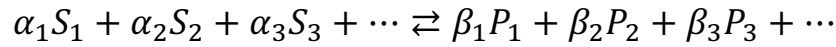


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S]}{K_{m,S}} - V_{max_b} \frac{[P]}{K_{m,P}}}{1 + \frac{[S]}{K_{m,S}} + \frac{[P]}{K_{m,P}}}$$

A.1.2 Irreversible (Michaelis-Menten like kinetics)

For a reaction with nS substrates and nP products in the form:



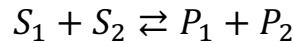
The rate expression reads:

$$v_{net} = V_{max_f} \prod_{i=1}^{nS} \left(\frac{\frac{[S_i]}{K_{m,S_i}}}{1 + \frac{[S_i]}{K_{m,S_i}}} \right)$$

A.2 Generalized Reversible Hill kinetics with Hill coefficient $h = 1$

A.2.1 Bi-Bi

For a reaction in the form:

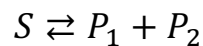


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S_1]}{K_{m,S_1}} \frac{[S_2]}{K_{m,S_2}} - V_{max_b} \frac{[P_1]}{K_{m,P_1}} \frac{[P_2]}{K_{m,P_2}}}{\left(1 + \frac{[S_1]}{K_{m,S_1}} + \frac{[P_1]}{K_{m,P_1}}\right) \left(1 + \frac{[S_2]}{K_{m,S_2}} + \frac{[P_2]}{K_{m,P_2}}\right)}$$

A.2.2 Uni-Bi

For a reaction in the form:

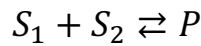


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S]}{K_{m,S}} - V_{max_b} \frac{[P_1]}{K_{m,P_1}} \frac{[P_2]}{K_{m,P_2}}}{\frac{[S]}{K_{m,S}} + \left(1 + \frac{[P_1]}{K_{m,P_1}}\right) \left(1 + \frac{[P_2]}{K_{m,P_2}}\right)}$$

A.2.3 Bi-Uni

For a reaction in the form:



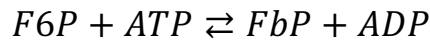
The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S_1]}{K_{m,S_1}} \frac{[S_2]}{K_{m,S_2}} - V_{max_b} \frac{[P]}{K_{m,P}}}{\frac{[P]}{K_{m,P}} + \left(1 + \frac{[S_1]}{K_{m,S_1}}\right) \left(1 + \frac{[S_2]}{K_{m,S_2}}\right)}$$

A.3 Generalized Reversible Hill kinetics with Hill coefficient $h = 4$

A.3.1 PFK specific mechanism

For the reaction:



where AMP acts as an allosteric inhibitor and ADP as an allosteric activator,

The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[F6P]}{K_{m,F6P}} \frac{[ATP]}{K_{m,ATP}} \left(\frac{[F6P]}{K_{m,F6P}} + \frac{[FbP]}{K_{m,FbP}}\right)^3 \left(\frac{[ATP]}{K_{m,ATP}} + \frac{[ADP]}{K_{m,ADP}}\right)^3}{D} - \frac{V_{max_b} \frac{[FbP]}{K_{m,FbP}} \frac{[ADP]}{K_{m,ADP}} \left(\frac{[F6P]}{K_{m,F6P}} + \frac{[FbP]}{K_{m,FbP}}\right)^3 \left(\frac{[ATP]}{K_{m,ATP}} + \frac{[ADP]}{K_{m,ADP}}\right)^3}{D}$$

where

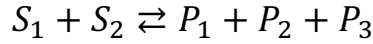
D

$$\begin{aligned}
 &= \left(\frac{[F6P]}{K_{m,F6P}} + \frac{[FbP]}{K_{m,FbP}} \right)^4 \left(\frac{[ATP]}{K_{m,ATP}} + \frac{[ADP]}{K_{m,ADP}} \right)^4 \\
 &+ \frac{\left\{ 1 + \left(\frac{[AMP]}{K_{m,AMP}} \right)^4 \right\} \left\{ 1 + \left(\frac{[ATP]}{K_{m,ATP}} \right)^4 \right\}}{\left\{ 1 + 10^4 \left(\frac{[AMP]}{K_{m,AMP}} \right)^4 \right\} \left\{ 1 + 10^{-4} \left(\frac{[ATP]}{K_{m,ATP}} \right)^4 \right\}} \\
 &+ \frac{\left\{ 1 + 10^2 \left(\frac{[AMP]}{K_{m,AMP}} \right)^4 \right\} \left\{ 1 + 10^{-2} \left(\frac{[ATP]}{K_{m,ATP}} \right)^4 \right\} \left\{ \left(\frac{[F6P]}{K_{m,F6P}} + \frac{[FbP]}{K_{m,FbP}} \right)^4 + \left(\frac{[ATP]}{K_{m,ATP}} + \frac{[ADP]}{K_{m,ADP}} \right)^4 \right\}}{\left\{ 1 + 10^4 \left(\frac{[AMP]}{K_{m,AMP}} \right)^4 \right\} \left\{ 1 + 10^{-4} \left(\frac{[ATP]}{K_{m,ATP}} \right)^4 \right\}}
 \end{aligned}$$

A.4 Convenience kinetics

A.4.1 Bi-Ter

For a reaction in the form:

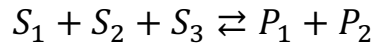


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S_1]}{K_{m,S_1}} \frac{[S_2]}{K_{m,S_2}} - V_{max_b} \frac{[P_1]}{K_{m,P_1}} \frac{[P_2]}{K_{m,P_2}} \frac{[P_3]}{K_{m,P_3}}}{\left(1 + \frac{[S_1]}{K_{m,S_1}} \right) \left(1 + \frac{[S_2]}{K_{m,S_2}} \right) + \left(1 + \frac{[P_1]}{K_{m,P_1}} \right) \left(1 + \frac{[P_2]}{K_{m,P_2}} \right) \left(1 + \frac{[P_3]}{K_{m,P_3}} \right) - 1}$$

A.4.2 Ter-Bi

For a reaction in the form:

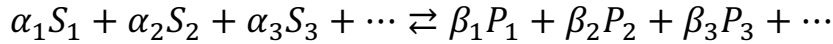


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \frac{[S_1]}{K_{m,S_1}} \frac{[S_2]}{K_{m,S_2}} \frac{[S_3]}{K_{m,S_3}} - V_{max_b} \frac{[P_1]}{K_{m,P_1}} \frac{[P_2]}{K_{m,P_2}}}{\left(1 + \frac{[S_1]}{K_{m,S_1}}\right) \left(1 + \frac{[S_2]}{K_{m,S_2}}\right) \left(1 + \frac{[S_3]}{K_{m,S_3}}\right) + \left(1 + \frac{[P_1]}{K_{m,P_1}}\right) \left(1 + \frac{[P_2]}{K_{m,P_2}}\right) - 1}$$

A.4.3 Stoichiometric coefficient larger than unity

For a reaction with nS substrates and nP products in the form:

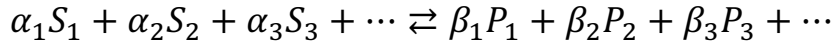


The rate expression reads:

$$v_{net} = \frac{V_{max_f} \prod_{i=1}^{nS} \left(\frac{[S_i]}{K_{m,S_i}}\right)^{\alpha_i} - V_{max_b} \prod_{j=1}^{nP} \left(\frac{[P_j]}{K_{m,P_j}}\right)^{\beta_j}}{\prod_{i=1}^{nS} \left\{ \sum_{n=0}^{\alpha_i} \left(\frac{[S_i]}{K_{m,S_i}}\right)^n \right\} + \prod_{j=1}^{nP} \left\{ \sum_{m=0}^{\beta_j} \left(\frac{[P_j]}{K_{m,P_j}}\right)^m \right\} - 1}$$

A.4.4 Containing apparent inhibition terms

For a reaction with nS substrates, nP products and nI apparent inhibitors in the form:



The rate expression reads:

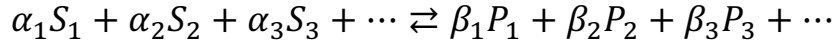
$$v_{net} = \frac{V_{max_f} \prod_{i=1}^{nS} \left(\frac{[S_i]}{K_{m,S_i}}\right)^{\alpha_i} - V_{max_b} \prod_{j=1}^{nP} \left(\frac{[P_j]}{K_{m,P_j}}\right)^{\beta_j}}{\prod_{i=1}^{nS} \left\{ \sum_{n=0}^{\alpha_i} \left(\frac{[S_i]}{K_{m,S_i}}\right)^n \right\} + \prod_{j=1}^{nP} \left\{ \sum_{m=0}^{\beta_j} \left(\frac{[P_j]}{K_{m,P_j}}\right)^m \right\} - 1 + \sum_{k=1}^{nI} \frac{[I_k]}{K_{i,I_k}}}$$

A.5 Mass-action kinetics

Note: Case of participating compensating small molecules

A.5.1 Diffusion of species or chemical reaction

For a reaction with nS small molecule substrates and nP small molecule products in the form:



The rate expression reads:

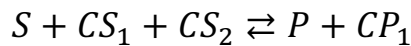
$$v_{net} = k_f \prod_{i=1}^{nS} [S_i]^{\alpha_i} - k_b \prod_{j=1}^{nP} [P_j]^{\beta_j}$$

Note: Case of participating compensating small molecules

If in any of the above cases there are nCS compensating substrates and nCP compensating products, the backward rate part of the numerator needs to be multiplied by the term:

$$\frac{\prod_{p=1}^{nCP} [CP_p]}{\prod_{q=1}^{nCS} [CS_q]}$$

In example, for a reaction in the form:



The uni-uni Michaelis-Menten rate expression will read:

$$v_{net} = \frac{V_{max_f} \frac{[S]}{K_{m,S}} - V_{max_b} \frac{[P]}{K_{m,P}} \frac{[CP_1]}{[CS_1][CS_2]}}{1 + \frac{[S]}{K_{m,S}} + \frac{[P]}{K_{m,P}}}$$

Bibliography

- Almquist J, Cvijovic M, Hatzimanikatis V *et al.* Kinetic models in industrial biotechnology - Improving cell factory performance. *Metabolic Engineering* 2014;**24**: 38-60.
- Alvarez-Vasquez F, Riezman H, Hannun YA *et al.* Mathematical Modeling and Validation of the Ergosterol Pathway in *Saccharomyces cerevisiae*. *Plos One* 2011;**6**.
- Alvarez-Vasquez F, Sims KJ, Cowart LA *et al.* Simulation and validation of modelled sphingolipid metabolism in *Saccharomyces cerevisiae*. *Nature* 2005;**433**: 425-30.
- Alvarez-Vasquez F, Sims KJ, Hannun YA *et al.* Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. *Journal of Theoretical Biology* 2004;**226**: 265-91.
- Alvarez-Vasquez F, Sims KJ, Voit EO *et al.* Coordination of the dynamics of yeast sphingolipid metabolism during the diauxic shift. *Theor Biol Med Model* 2007;**4**.
- Alves AC, Ribeiro D, Nunes C *et al.* Biophysics in cancer: The relevance of drug-membrane interaction studies. *Bba-Biomembranes* 2016;**1858**: 2231-44.
- Andreozzi S, Miskovic L, Hatzimanikatis V. iSCHRUNK–In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks. *Metabolic engineering* 2016;**33**: 158-68.
- Archibald JM. Endosymbiosis and Eukaryotic Cell Evolution. *Curr Biol* 2015;**25**: R911-R21.
- Ataman M. Navigating and Managing the Complexity of Genome Scale Metabolic Networks for Studies in Cellular Physiology and Industrial Biotechnology *Chemistry and Chemical Engineering* volume PhD. Lausanne: EPFL, 2016, 178.
- Ataman M, Gardiol DFH, Fengos G *et al.* redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *Plos Computational Biology* 2017;**13**.
- Ataman M, Hatzimanikatis V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Current opinion in biotechnology* 2015;**36**: 176-82.
- Ataman M, Hatzimanikatis V. lumpGEM: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. *Plos Computational Biology* 2017;**13**.
- Aung HW, Henry SA, Walker LP. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind Biotechnol (New Rochelle N Y)* 2013;**9**: 215-28.
- Baile MG, Lu YW, Claypool SM. The topology and regulation of cardiolipin biosynthesis and remodeling in yeast. *Chem Phys Lipids* 2014;**179**: 25-31.
- Bailey JE. Toward a science of metabolic engineering. *Science* 1991;**252**: 1668-75.
- Beard DA, Liang S-d, Qian H. Energy balance for analysis of complex metabolic networks. *Biophysical journal* 2002;**83**: 79-86.
- Benner SA, Hoshika S, Sukeda M *et al.* Synthetic biology for improved personalized medicine. *Nucleic Acids Symp Ser (Oxf)* 2008, DOI 10.1093/nass/nrn123: 243-4.

- Bhattacharya BS, Sweby PK, Minihane AM *et al.* A mathematical model of the sterol regulatory element binding protein 2 cholesterol biosynthesis pathway. *Journal of Theoretical Biology* 2014;**349**: 150-62.
- Bonarius HPJ, Schmid G, Tramper J. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology* 1997;**15**: 308-14.
- Bowden AC. Metabolic control analysis in biotechnology and medicine. *Nat Biotechnol* 1999;**17**: 641-3.
- Bradbury MW. Lipid metabolism and liver inflammation. I. Hepatic fatty acid uptake: Possible role in steatosis. *Am J Physiol-Gastr L* 2006;**290**: G194-G8.
- Cakir T, Hendriks MMWB, Westerhuis JA *et al.* Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics* 2009;**5**: 318-29.
- Carbonell P, Lecoindre G, Faulon JL. Origins of Specificity and Promiscuity in Metabolic Networks. *Journal of Biological Chemistry* 2011;**286**: 43994-4004.
- Carman GM, Henry SA. Phospholipid Biosynthesis in Yeast. *Annual Review of Biochemistry* 1989;**58**: 635-69.
- Carman GM, Henry SA. Phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae* and interrelationship with other metabolic processes. *Prog Lipid Res* 1999;**38**: 361-99.
- Carman GM, Zeimetz GM. Regulation of phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 1996;**271**: 13293-6.
- Cascante M, Boros LG, Comin-Anduix B *et al.* Metabolic control analysis in drug discovery and disease. *Nat Biotechnol* 2002;**20**: 243-9.
- Chakrabarti A, Miskovic L, Soh KC *et al.* Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology journal* 2013;**8**: 1043-57.
- Cherry JM, Adler C, Ball C *et al.* SGD: *Saccharomyces Genome Database*. *Nucleic Acids Res* 1998;**26**: 73-9.
- Christensen B, Nielsen J. Metabolic network analysis. A powerful tool in metabolic engineering. *Adv Biochem Eng Biotechnol* 2000;**66**: 209-31.
- Cohen AE, Venkatachalam V. Bringing bioelectricity to light. *Annu Rev Biophys* 2014;**43**: 211-32.
- Cook DJ, Nielsen J. Genome-scale metabolic models applied to human health and disease. *Wiley Interdiscip Rev Syst Biol Med* 2017;**9**.
- Costa RS, Hartmann A, Vinga S. Kinetic modeling of cell metabolism for microbial production. *J Biotechnol* 2016;**219**: 126-41.
- da Silveira Dos Santos AX, Riezman I, Aguilera-Romero MA *et al.* Systematic lipidomic analysis of yeast protein kinase and phosphatase mutants reveals novel insights into regulation of lipid homeostasis. *Mol Biol Cell* 2014;**25**: 3234-46.
- Dai Z, Yang S, Xu L *et al.* Identification of Cancer-associated metabolic vulnerabilities by modeling multi-objective optimality in metabolism. *Cell Commun Signal* 2019;**17**: 124.
- Daum G, Tuller G, Nemeč T *et al.* Systematic analysis of yeast strains with possible defects in lipid metabolism. *Yeast* 1999;**15**: 601-14.
- de Kroon AIPM. Lipidomics in research on yeast membrane lipid homeostasis. *Bba-Mol Cell Biol L* 2017;**1862**: 797-9.
- DeBerardinis RJ, Thompson CB. Cellular Metabolism and Disease: What Do Metabolic Outliers Teach Us? *Cell* 2012;**148**: 1132-44.
- Dennis EA. Lipidomics joins the omics evolution. *Proc Natl Acad Sci U S A* 2009;**106**: 2089-90.
- Despres JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature* 2006;**444**: 881-7.
- Dickinson JR, Schweizer M. *Metabolism and Molecular Physiology of Saccharomyces Cerevisiae*: CRC Press, 2004.
- Downes CP, Gray A, Lucocq JM. Probing phosphoinositide functions in signaling and membrane trafficking. *Trends in Cell Biology* 2005;**15**: 259-68.

- Emanuelsson O, Brunak S, von Heijne G *et al.* Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* 2007;**2**: 953-71.
- Fahy E, Subramaniam S, Brown HA *et al.* A comprehensive classification system for lipids. *J Lipid Res* 2005;**46**: 839-61.
- Fahy E, Subramaniam S, Murphy RC *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 2009;**50**: S9-S14.
- Farquhar JW, Gross RC, Wagner RM *et al.* Validation of an Incompletely Coupled 2-Compartment Nonrecycling Catenary Model for Turnover of Liver and Plasma Triglyceride in Man. *J Lipid Res* 1965;**6**: 119-&.
- Feist AM, Henry CS, Reed JL *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* 2007;**3**.
- Fell DA. Increasing the flux in metabolic pathways: A metabolic control analysis perspective. *Biotechnology and Bioengineering* 1998;**58**: 121-4.
- Fell DA, Sauro HM. Metabolic control and its analysis: additional relationships between elasticities and control coefficients. *European Journal of Biochemistry* 1985;**148**: 555-61.
- Forster J, Famili I, Fu P *et al.* Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 2003;**13**: 244-53.
- Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol* 2005;**13**: 550-8.
- Gardner JJ, Hodge BMS, Boyle NR. Multiscale Multiobjective Systems Analysis (MiMoSA): an advanced metabolic modeling framework for complex systems. *Scientific Reports* 2019;**9**.
- Goffeau A. Four years of post-genomic life with 6,000 yeast genes. *Febs Lett* 2000;**480**: 37-41.
- Goffeau A, Barrell BG, Bussey H *et al.* Life with 6000 genes. *Science* 1996;**274**: 546, 63-7.
- Gombert AK, Nielsen J. Mathematical modelling of metabolism. *Current Opinion in Biotechnology* 2000;**11**: 180-6.
- Greenspan RJ. Opinion - The flexible genome. *Nat Rev Genet* 2001;**2**: 383-7.
- Gu ZM, Valianpour F, Chen SL *et al.* Aberrant cardiolipin metabolism in the yeast *taz1* mutant: a model for Barth syndrome. *Mol Microbiol* 2004;**51**: 149-58.
- Guan XL, Souza CM, Pichler H *et al.* Functional Interactions between Sphingolipids and Sterols in Biological Membranes Regulating Cell Physiology. *Mol Biol Cell* 2009;**20**: 2083-95.
- Gupta S, Maurya MR, Merrill AH, Jr. *et al.* Integration of lipidomics and transcriptomics data towards a systems biology model of sphingolipid metabolism. *BMC Syst Biol* 2011;**5**: 26.
- Hadadi N, Cher Soh K, Seijo M *et al.* A computational framework for integration of lipidomics data into metabolic pathways. *Metab Eng* 2014;**23**: 1-8.
- Hameri T, Fengos G, Ataman M *et al.* Kinetic models of metabolism that consider alternative steady-state solutions of intracellular fluxes and concentrations. *Metabolic engineering* 2019;**52**: 29-41.
- Hameri T, Fengos G, Hatzimanikatis V. The effects of model complexity and size on metabolic flux distribution and control. Case study in *E. coli*. *bioRxiv* 2019, DOI 10.1101/666859: 666859.
- Han XL. Lipidomics for studying metabolism. *Nat Rev Endocrinol* 2016;**12**: 668-79.
- Harayama T, Riezman H. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* 2018;**19**: 281-96.
- Hatzimanikatis V, Bailey JE. MCA has more to say. *Journal of Theoretical Biology* 1996;**182**: 233-42.
- Hatzimanikatis V, Bailey JE. Effects of spatiotemporal variations on metabolic control: Approximate analysis using (log)linear kinetic models. *Biotechnology and Bioengineering* 1997;**54**: 91-104.
- Hatzimanikatis V, Floudas CA, Bailey JE. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE Journal* 1996a;**42**: 1277-92.

- Hatzimanikatis V, Floudas CA, Bailey JE. Optimization of regulatory architectures in metabolic reaction networks. *Biotechnology and bioengineering* 1996b;**52**: 485-500.
- He F, Murabito E, Westerhoff HV. Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering. *J R Soc Interface* 2016;**13**.
- Heavner BD, Price ND. Comparative Analysis of Yeast Metabolic Network Models Highlights Progress, Opportunities for Metabolic Reconstruction. *PLoS Comput Biol* 2015;**11**: e1004530.
- Heckmann D, Zielinski DC, Palsson BO. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. *Nature Communications* 2018;**9**.
- Heindel JJ, Blumberg B, Cave M *et al*. Metabolism disrupting chemicals and metabolic disorders. *Reproductive Toxicology* 2017;**68**: 3-33.
- Heinrich R, Rapoport TA. A Linear Steady-State Treatment of Enzymatic Chains. *European Journal of Biochemistry* 1974;**42**: 89-95.
- Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. *Biophysical Journal* 2007;**92**: 1792-805.
- Ho P, Chen YY. Mammalian synthetic biology in the age of genome editing and personalized medicine. *Curr Opin Chem Biol* 2017;**40**: 57-64.
- Holthuis JC, Menon AK. Lipid landscapes and pipelines in membrane homeostasis. *Nature* 2014;**510**: 48-57.
- Hotamisligil GS. Inflammation and metabolic disorders. *Nature* 2006;**444**: 860-7.
- Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. *Nature* 2017;**542**: 177-85.
- Hubner K, Schwager T, Winkler K *et al*. Computational lipidology: Predicting lipoprotein density profiles in human blood plasma. *Plos Computational Biology* 2008;**4**.
- Ivanova PT, Milne SB, Myers DS *et al*. Lipidomics: a mass spectrometry based systems level analysis of cellular lipids. *Curr Opin Chem Biol* 2009;**13**: 526-31.
- Jain KK. Synthetic biology and personalized medicine. *Med Princ Pract* 2013;**22**: 209-19.
- Jensen RA. Enzyme Recruitment in Evolution of New Function. *Annu Rev Microbiol* 1976;**30**: 409-25.
- Joyce AR, Palsson BO. The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology* 2006;**7**: 198-210.
- Kacser H, Burns J. The control of flux *Symp Soc Exp Biol* volume 27, 1973a, 65-104.
- Kacser H, Burns JA, Fell DA. The Control of Flux. *Biochem Soc T* 1995;**23**: 341-66.
- Kacser Ha, Burns J. The control of flux *Symp Soc Exp Biol* volume 27, 1973b, 65-104.
- Kato M, Wickner W. Ergosterol is required for the Sec18/ATP-dependent priming step of homotypic vacuole fusion. *Embo J* 2001;**20**: 4035-40.
- Khan MI, Shin JH, Kim JD. The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories* 2018;**17**.
- Khersonsky O, Tawfik DS. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu Rev Biochem* 2010;**79**: 471-505.
- Khodayari A, Maranas CD. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nature Communications* 2016;**7**.
- Kholodenko BN, Sauro HM, Westerhoff HV. Control by Enzymes, Coenzymes and Conserved Moieties - a Generalization of the Connectivity Theorem of Metabolic Control Analysis. *European Journal of Biochemistry* 1994;**225**: 179-86.
- King ZA, Drager A, Ebrahim A *et al*. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol* 2015;**11**: e1004321.
- Klose C, Surma MA, Gerl MJ *et al*. Flexibility of a Eukaryotic Lipidome - Insights from Yeast Lipidomics. *Plos One* 2012;**7**.
- Klug L, Daum G. Yeast lipid metabolism at a glance. *Fems Yeast Research* 2014;**14**: 369-88.

- Knoblauch H, Schuster H, Luft JC *et al.* A pathway model of lipid metabolism to predict the effect of genetic variability on lipid levels. *J Mol Med-Jmm* 2000;**78**: 507-15.
- Kontush A, Chapman MJ. Lipidomics as a Tool for the Study of Lipoprotein Metabolism. *Curr Atheroscler Rep* 2010;**12**: 194-201.
- Kosicek M, Hecimovic S. Phospholipids and Alzheimer's Disease: Alterations, Mechanisms and Potential Biomarkers. *Int J Mol Sci* 2013;**14**: 1310-22.
- Krauss M, Haucke V. Phosphoinositide-metabolizing enzymes at the interface between membrane traffic and cell signalling. *Embo Rep* 2007;**8**: 241-6.
- Lerman JA, Hyduke DR, Latif H *et al.* In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications* 2012;**3**.
- Lewis NE, Hixson KK, Conrad TM *et al.* Omic data from evolved E-coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* 2010;**6**.
- Link H, Christodoulou D, Sauer U. Advancing metabolic models with kinetic information. *Current Opinion in Biotechnology* 2014;**29**: 8-14.
- Liu GD, Chen Y, Faergeman NJ *et al.* Elimination of the last reactions in ergosterol biosynthesis alters the resistance of *Saccharomyces cerevisiae* to multiple stresses. *Fems Yeast Research* 2017;**17**.
- Liu JK, O'Brien EJ, Lerman JA *et al.* Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC systems biology* 2014;**8**: 110.
- Long MR, Ong WK, Reed JL. Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology* 2015;**34**: 135-41.
- Lopes H, Rocha I. Genome-scale modeling of yeast: chronology, applications and critical perspectives. *Fems Yeast Research* 2017;**17**.
- Lu HZ, Li FR, Sanchez BJ *et al.* A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature Communications* 2019;**10**.
- Lutz S. Beyond directed evolution-semi-rational protein engineering and design. *Current Opinion in Biotechnology* 2010;**21**: 734-43.
- Lydic TA, Goo YH. Lipidomics unveils the complexity of the lipidome in metabolic diseases. *Clin Transl Med* 2018;**7**: 4.
- Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering* 2003;**5**: 264-76.
- Marchesi JR, Adams DH, Fava F *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* 2016;**65**: 330-9.
- Markgraf DF, Al-Hasani H, Lehr S. Lipidomics-Reshaping the Analysis and Perception of Type 2 Diabetes. *Int J Mol Sci* 2016;**17**.
- Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 1975;**405**: 442-51.
- Mavrovouniotis ML. Group Contributions for Estimating Standard Gibbs Energies of Formation of Biochemical-Compounds in Aqueous-Solution. *Biotechnology and Bioengineering* 1990;**36**: 1070-82.
- Mavrovouniotis ML. Estimation of Standard Gibbs Energy Changes of Biotransformations. *Journal of Biological Chemistry* 1991;**266**: 14440-5.
- McCloskey D, Gangoiti JA, King ZA *et al.* A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in *E. coli* K-12 MG1655 that is biochemically and thermodynamically consistent. *Biotechnology and bioengineering* 2014;**111**: 803-15.
- Mewes HW, Albermann K, Bahr M *et al.* Overview of the yeast genome. *Nature* 1997;**387**: 7-8.
- Miskovic L, Beal J, Moret M *et al.* Uncertainty reduction in biochemical kinetic models: Enforcing desired model properties. *PLoS Comput Biol* 2019a;**15**: e1007242.
- Miskovic L, Hatzimanikatis V. Production of biofuels and biochemicals: in need of an ORACLE. *Trends in Biotechnology* 2010;**28**: 391-7.

- Miskovic L, Tokic M, Fengos G *et al.* Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. *Current Opinion in Biotechnology* 2015;**36**: 1-8.
- Miskovic L, Tokic M, Savoglidis G *et al.* Control Theory Concepts for Modeling Uncertainty in Enzyme Kinetics of Biochemical Networks. *Industrial & Engineering Chemistry Research* 2019b;**58**: 13544-54.
- Mo ML, Palsson BO, Herrgard MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *Bmc Systems Biology* 2009;**3**.
- Moreno-Sanchez R, Saavedra E, Rodriguez-Enriquez S *et al.* Metabolic control analysis: A tool for designing strategies to manipulate metabolic pathways. *J Biomed Biotechnol* 2008, DOI Artn 597913
10.1155/2008/597913.
- Mountain HA, Bystrom AS, Larsen JT *et al.* Four major transcriptional responses in the methionine/threonine biosynthetic pathway of *Saccharomyces cerevisiae*. *Yeast* 1991;**7**: 781-803.
- Natter K, Kohlwein SD. Yeast and cancer cells - common principles in lipid metabolism. *Bba-Mol Cell Biol L* 2013;**1831**: 314-26.
- Nielsen J. Systems biology of lipid metabolism: From yeast to human. *Febs Lett* 2009;**583**: 3905-13.
- Nishiguchi H, Hiasa N, Uebayashi K *et al.* Transomics data-driven, ensemble kinetic modeling for system-level understanding and engineering of the cyanobacteria central metabolism. *Metab Eng* 2019;**52**: 273-83.
- Nookaew I, Jewett MC, Meechai A *et al.* The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2008;**2**: 71.
- Norman KL, Kumar A. Mutant power: using mutant allele collections for yeast functional genomics. *Brief Funct Genomics* 2016;**15**: 75-84.
- O'Brien EJ, Lerman JA, Chang RL *et al.* Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology* 2013;**9**.
- O'Brien PJ, Herschlag D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 1999;**6**: R91-R105.
- Ogretmen B, Hannun YA. Biologically active sphingolipids in cancer pathogenesis and treatment. *Nat Rev Cancer* 2004;**4**: 604-16.
- Orij R, Postmus J, Ter Beek A *et al.* In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in *Saccharomyces cerevisiae* reveals a relation between intracellular pH and growth. *Microbiology* 2009;**155**: 268-78.
- Orij R, Urbanus ML, Vizeacoumar FJ *et al.* Genome-wide analysis of intracellular pH reveals quantitative control of cell division rate by pH(c) in *Saccharomyces cerevisiae*. *Genome Biol* 2012;**13**.
- Orth JD, Conrad TM, Na J *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol Syst Biol* 2011a;**7**.
- Orth JD, Conrad TM, Na J *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular systems biology* 2011b;**7**.
- Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nature Biotechnology* 2010;**28**: 245-8.
- Osterlund T, Nookaew I, Bordel S *et al.* Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 2013;**7**: 36.
- Osterlund T, Nookaew I, Nielsen J. Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Biotechnology Advances* 2012;**30**: 979-88.
- Pandey V, Hadadi N, Hatzimanikatis V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *Plos Computational Biology* 2019;**15**.

- Paroutis P, Touret N, Grinstein S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology (Bethesda)* 2004;**19**: 207-15.
- Petranovic D, Tyo K, Vemuri GN *et al.* Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism. *Fems Yeast Research* 2010;**10**: 1046-59.
- Preston RA, Murphy RF, Jones EW. Assay of vacuolar pH in yeast and identification of acidification-defective mutants. *Proc Natl Acad Sci U S A* 1989;**86**: 7027-31.
- Price ND, Papin JA, Schilling CH *et al.* Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology* 2003;**21**: 162-9.
- Ragauskas AJ, Williams CK, Davison BH *et al.* The path forward for biofuels and biomaterials. *Science* 2006;**311**: 484-9.
- Rani V, Deep G, Singh RK *et al.* Oxidative stress and metabolic disorders: Pathogenesis and therapeutic strategies. *Life Sci* 2016;**148**: 183-93.
- Reder C. Metabolic control theory: a structural approach. *Journal of theoretical biology* 1988a;**135**: 175-201.
- Reder C. Metabolic Control-Theory - a Structural Approach. *Journal of Theoretical Biology* 1988b;**135**: 175-201.
- Rodionova MV, Poudyal RS, Tiwari I *et al.* Biofuel production: Challenges and opportunities. *Int J Hydrogen Energy* 2017;**42**: 8450-61.
- Saa PA, Nielsen LK. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology advances* 2017.
- Salvy P, Fengos G, Ataman M *et al.* pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. *Bioinformatics* 2019;**35**: 167-9.
- Salvy P, Hatzimanikatis V. ETFL: A formulation for flux balance models accounting for expression, thermodynamics, and resource allocation constraints. *bioRxiv* 2019, DOI 10.1101/590992: 590992.
- Sanchez BJ, Li F, Kerkhoven EJ *et al.* SLIMer: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. *BMC Syst Biol* 2019;**13**: 4.
- Sanchez BJ, Nielsen J. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr Biol-Uk* 2015;**7**: 846-58.
- Sánchez BJ, Zhang C, Nilsson A *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular systems biology* 2017;**13**: 935.
- Santiago JA, Potashkin JA. Shared dysregulated pathways lead to Parkinson's disease and diabetes. *Trends Mol Med* 2013;**19**: 176-86.
- Santos AXS, Riezman H. Yeast as a model system for studying lipid homeostasis and function. *Febs Lett* 2012;**586**: 2858-67.
- Sarkar D, Maranas CD. Engineering microbial chemical factories using metabolic models. *BMC Chemical Engineering* 2019;**1**: 22.
- Savageau MA. Biochemical Systems Analysis .1. Some Mathematical Properties of Rate Law for Component Enzymatic Reactions. *Journal of Theoretical Biology* 1969;**25**: 365-&.
- Savoglidis G, dos Santos AXD, Riezman I *et al.* A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism. *Metabolic Engineering* 2016;**37**: 46-62.
- Schlame M, Greenberg ML. Biosynthesis, remodeling and turnover of mitochondrial cardiolipin. *Bba-Mol Cell Biol L* 2017;**1862**: 3-7.
- Schneider R, Brugger B, Sandhoff R *et al.* Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of the lipid molecular species composition of yeast subcellular membranes reveals acyl chain-based sorting/remodeling of distinct molecular species en route to the plasma membrane. *J Cell Biol* 1999;**146**: 741-54.
- Schork NJ. Time for one-person trials. *Nature* 2015;**520**: 609-11.
- Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology* 2007;**3**.

- Schuster S. Use and Limitations of Modular Metabolic Control Analysis in Medicine and Biotechnology. *Metabolic Engineering* 1999;**1**: 232-42.
- Schutzhold V, Hahn J, Tummeler K *et al.* Computational Modeling of Lipid Metabolism in Yeast. *Front Mol Biosci* 2016;**3**.
- Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *P Natl Acad Sci USA* 2002;**99**: 15112-7.
- Singh D, Lercher MJ. Network reduction methods for genome-scale metabolic models. *Cellular and Molecular Life Sciences* 2019, DOI 10.1007/s00018-019-03383-z.
- Soh KC, Hatzimanikatis V. Constraining the Flux Space Using Thermodynamics and Integration of Metabolomics Data *Metabolic Flux Analysis*: Springer, 2014, 49-63.
- Soh KS, Miskovic L, Hatzimanikatis V. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Research* 2012;**12**: 129-43.
- Strutz J, Martin J, Greene J *et al.* Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain. *Current Opinion in Biotechnology* 2019;**59**: 24-30.
- Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* 2010;**5**: 93-121.
- Thomas D, Barbey R, Henry D *et al.* Physiological analysis of mutants of *Saccharomyces cerevisiae* impaired in sulphate assimilation. *J Gen Microbiol* 1992;**138**: 2021-8.
- Tidhar R, Futerman AH. The complexity of sphingolipid biosynthesis in the endoplasmic reticulum. *Bba-Mol Cell Res* 2013;**1833**: 2511-8.
- Tokic M, Miskovic L, Hatzimanikatis V. Large-scale kinetic metabolic models of *Pseudomonas putida* for a consistent design of metabolic engineering strategies. *bioRxiv* 2019, DOI 10.1101/569152: 569152.
- Tomar N, De RK. Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene* 2013;**521**: 1-14.
- Tran LM, Rizk ML, Liao JC. Ensemble Modeling of Metabolic Networks. *Biophysical Journal* 2008;**95**: 5606-17.
- van Eunen K, Simons SMJ, Gerding A *et al.* Biochemical Competition Makes Fatty-Acid beta-Oxidation Vulnerable to Substrate Overload. *Plos Computational Biology* 2013;**9**.
- Van Urk H, Mak PR, Scheffers WA *et al.* Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess. *Yeast* 1988;**4**: 283-91.
- Varma A, Palsson BO. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic by-Product Secretion in Wild-Type *Escherichia-Coli* W3110. *Applied and Environmental Microbiology* 1994;**60**: 3724-31.
- Velez Rueda AJ, Palopoli N, Zacarias M *et al.* ProtMiscuity: a database of promiscuous proteins. *Database (Oxford)* 2019;**2019**.
- Volke DC, Rohwer J, Fischer R *et al.* Investigation of the methylerythritol 4-phosphate pathway for microbial terpenoid production through metabolic control analysis. *Microbial Cell Factories* 2019;**18**.
- Vriens K, Christen S, Parik S *et al.* Evidence for an alternative fatty acid desaturation pathway increasing cancer plasticity. *Nature* 2019;**566**: 403-+.
- Wang L, Birol I, Hatzimanikatis V. Metabolic Control Analysis under Uncertainty: Framework Development and Case Studies. *Biophysical Journal* 2004;**87**: 3750-63.
- Wang L, Dash S, Ng CY *et al.* A review of computational tools for design and reconstruction of metabolic pathways. *Synth Syst Biotechno* 2017;**2**: 243-52.
- Wang LQ, Hatzimanikatis V. Metabolic engineering under uncertainty - II: Analysis of yeast metabolism. *Metabolic Engineering* 2006a;**8**: 142-59.
- Wang LQ, Hatzimanikatis V. Metabolic engineering under uncertainty. I: Framework development. *Metabolic Engineering* 2006b;**8**: 133-41.

- Ward VCA, Chatzivasileiou AO, Stephanopoulos G. Cell free biosynthesis of isoprenoids from isopentenol. *Biotechnol Bioeng* 2019;**116**: 3269-81.
- Watterson S, Guerriero ML, Blanc M *et al*. A model of flux regulation in the cholesterol biosynthesis pathway: Immune mediated graduated flux reduction versus statin-like led stepped flux reduction. *Biochimie* 2013;**95**: 613-21.
- Weng JK, Philippe RN, Noel JP. The Rise of Chemodiversity in Plants. *Science* 2012;**336**: 1667-70.
- Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov* 2005;**4**: 594-610.
- Yang L, Yurkovich JT, Lloyd CJ *et al*. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Scientific Reports* 2016;**6**.
- Zamboni N, Kümmel A, Heinemann M. anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data. *BMC bioinformatics* 2008;**9**: 199.
- Zaugg J, Gumulya Y, Gillam EM *et al*. Computational tools for directed evolution: a comparison of prospective and retrospective strategies. *Methods Mol Biol* 2014;**1179**: 315-33.
- Zhang YHP. Production of Biocommodities and Bioelectricity by Cell-Free Synthetic Enzymatic Pathway Biotransformations: Challenges and Opportunities. *Biotechnology and Bioengineering* 2010;**105**: 663-77.
- Zomorodi AR, Maranas CD. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *Bmc Systems Biology* 2010;**4**.



CURRICULUM VITAE

SOFIA TSOUKA

Allée des Cèdres 7
1022 Chavannes-près-Renens
Tel.: +41 76 690 60 91
Email: sofia.tsouka@epfl.ch

EDUCATION

- 2016 – 2020** **Doctor of Science**, EPFL, Switzerland.
Doctoral School of Chemistry and Chemical Engineering
- 2013 – 2015** **Master of Science**, University of Patras, Greece.
Process Simulation, Optimization and Control
- 2007 – 2013** **Diploma of Chemical Engineering**, University of Patras, Greece.

PROFESSIONAL EXPERIENCE

- 2016 – 2020** **Doctoral Assistant**, EPFL, Switzerland.
Laboratory of Computational Systems Biotechnology under the supervision of Prof. Vassily Hatzimanikatis. Development of new computational methods for strain design and integration of experimental data in metabolic models.
- 2013 – 2015** **Research Assistant**, University of Patras, Greece.
Master thesis: Stress-gradient induced migration of polymers in complex geometries.
- 2011 – 2013** **Research Assistant**, University of Patras, Greece.
Diploma thesis: Stress-gradient induced migration of polymers in corrugated channels.

TEACHING EXPERIENCE

- 2016 – 2019** **Teaching Assistant**, EPFL, Switzerland.
620 hours of teaching undergraduate and graduate courses: Introduction to Chemical Engineering, Biochemical Engineering, Fluid Mechanics and Transport Phenomena, and Transport Phenomena I & II.
- 2018** **Invited Tutor**, International Max Planck Research School for Cell, Developmental and Systems Biology, Germany.
Dresden Summer School in Systems Biology, organized by the Center for Systems Biology Dresden.
- 2013 – 2015** **Teaching Assistant**, University of Patras, Greece.
100 hours of teaching undergraduate courses: Heat Transfer, Fluid Dynamics, and Numerical Methods.

PERSONAL SKILLS

- Languages** • Greek, native • English, fully fluent (C2) • French, limited (B1)
- Software** Proficient in programming in MATLAB and Fortran 90/95.
Proficient in MS Office (Word, PowerPoint, Excel) and Adobe Suite.
Proficient in working in a UNIX environment.
Proficient in version control tools (GIT).
Basic knowledge in Python and R.
- Methods** Big data analysis, biostatistics and bioinformatics.
Integration of -omics data in computational models.
Reconstruction and analysis of metabolic models.
Formulation of constraint-based optimization problems.
- Management** Experience in mentoring Bachelor and Master students.
Passionate about interdisciplinary collaborations and continued learning opportunities.

AWARDS

- 2015** 1st LIMMAT Excellency award from the Limmat Stiftung Foundation for the performance during postgraduate studies.

PUBLICATIONS

redLips: a comprehensive mechanistic model of the lipid metabolic network of yeast, S. Tsouka and V. Hatzimanikatis, FEMS Yeast Research, *in press*

Stress-Gradient Induced Migration of Polymers in Thin Films Flowing Over Smoothly Corrugated Surfaces, S. Tsouka, Y. Dimakopoulos and J. Tsamopoulos, Journal of NonNewtonian Fluid Mechanics, Journal of Non-Newtonian Fluid Mechanics 228, 79-95, 2016

Hemodynamics in stenotic vessels of small diameter under steady state conditions: Effect of viscoelasticity and migration of red blood cells, Y. Dimakopoulos, G. Kelesidis, S. Tsouka, G.C. Georgiou and J. Tsamopoulos, Biorheology 52, 183–210, 2015

Stress-gradient Induced Migration of Polymers in Corrugated Channels, S. Tsouka, Y. Dimakopoulos, V. Mavrantzas and J. Tsamopoulos, Journal of Rheology 58(4), 911-947, 2014

CONFERENCES (SELECTED TALKS)

- 2018** **Mapping the diverse phenotypic states of yeast lipid metabolism to enzymes using mechanistic kinetic models**, S. Tsouka and V. Hatzimanikatis, 59th International Conference on the Bioscience of Lipids, Helsinki (Finland).
- 2014** **Polymer Stress-Gradient Induced Migration Phenomena in the Thin Film Flow over a Corrugated Topography**, S. Tsouka, Y. Dimakopoulos and J. Tsamopoulos, 9th National Conference on Fluid Flow Phenomena, Athens (Greece).
- 2014** **Stress-Gradient Induced Migration Phenomena in Thin Film Flow over Topography**, S. Tsouka, Y. Dimakopoulos, V. Mavrantzas and J. Tsamopoulos, WCCM XI – ECCM V – ECFD VI, Barcelona (Spain).

CONFERENCES (SELECTED POSTERS)

- 2018** **Identification of the Enzymes Responsible for Diverse Phenotypic States of Yeast Lipid Metabolism Using Comprehensive Mechanistic Models**, S. Tsouka and V. Hatzimanikatis, 5th Conference on Constraint Based Reconstruction and Analysis, Seattle (USA).
- 2017** **Identification of the Enzymes that Determine the Robustness of Lipid Profiles in Yeast**, S. Tsouka, G. Savoglidis, H. Riezman and V. Hatzimanikatis, 58th International Conference on the Bioscience of Lipids, Zürich (Switzerland).
- 2017** **A Kinetic Model of the Lipid Network in Yeast**, S. Tsouka, G. Savoglidis and V. Hatzimanikatis, 3rd International SystemsX.ch Conference on Systems Biology, Zürich (Switzerland).

REFERENCES

Available upon request