Computational investigation of adaptive evolution in *S. cerevisiae* by using genome-scale metabolic models

by

Handan Çetin

B.S., Molecular Biology and Genetics, Canakkale Onsekiz Mart University, 2018

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computational Science and Engineering
Boğaziçi University

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Systems Biology

With the increasing availability of the computational tools and the development of high throughput techniques in the omics field, systems biology has shown a strong emergence in the last few years as a key multi-disciplinary field for integrating the multi-layer complexity of biological systems, particularly in the areas of transcriptomics, proteomics, metabolomics and fluxomics [1]. This amount of available data allows researchers to investigate molecular cell processes in a large scales, applying theoretical, experimental and computational methods.

Biological systems based on complex interactions between various molecular components. The relations between these components are often obey nonlinear kinetics, for example, most of the reactions are regulated by one or more feedback or feed-forward loops with incomprehensible behaviours. When considered, cell structure and compartmentalization are also often introduce complexities to the unexpected behavior of the entire biological system [2]. Mathematical modeling with these factors taken into consideration is used as a general approach to encompass existing knowledge in biological systems, and to gather information by analyzing these models to acquire a better understanding [3].

A mathematical model of a cell can be approached by two different approaches in either a bottom-up or top-down directionality (Figure 1.1) [4, 5]. Top-down approach is an experimental oriented approach, it starts from the whole picture and aims to characterize biological mechanisms closer to the smaller parts and their interactions in the network. In the bottom-up approach, collected data from biological knowledge is used as a starting point, a subsystem is generated to deduce the functional properties of smaller points in the network. Combination of the pathway level models (bottom-up) into a model for the entire system level (top-down) is the ultimate goal in the systems biology therefore these approaches are complementary.

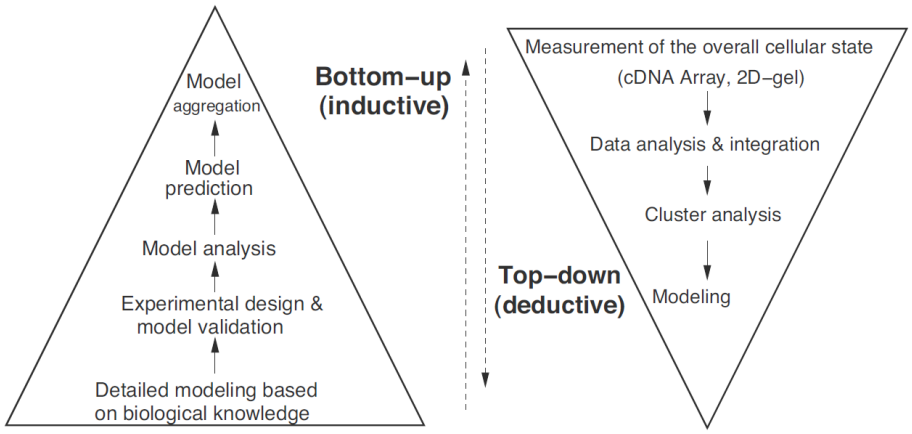

Figure 1.1. Systems biology approaches. Left: Bottom-up approach. Right: Top-down approach. Figure is taken from [3].

1.1.1. Metabolic Networks

In the context of systems biology, metabolic network reconstructions have become a common interest for the researchers over the past 20 years [6]. Organism-specific metabolic network analyses allow scientists to design experiments and even obtain beforehand predictions. These networks are the main sources of the mathematical models which can simulate metabolic fluxes reflecting the experimantal reality [7].

Before the improvement of genome sequencing or annotation technologies, initial core metabolic networks were based on the accessible information of biochemical pathways [8] [9]. In the last decade, larger genome-scale metabolic models (GSMMs) have been able to be developed rapidly with the help of databases for annotated genomes, providing information on substrates and products of each enzyme and each bioreaction [10]. Growing biochemical databases provide automatization processes for the metabolic network reconstructions. As a result, genome-scale metabolic networks are available today for almost all organisms with an annotated genome available in the literature [11, 12]. From the first genome-scale metabolic model of *Escherichia coli* to other organisms, the steps are required for GSMM development remained the same regardless of the biological diversity.

A generally applicable protocol is defined by the Palsson group [6, 10] for the reconstruction of biochemical networks described in the Figure 1.2 [13]. Briefly, genomic data for the biochemical reactions of an organism are identified from the databases, such as NCBI, DDBJ and EMBL-EBI. Extraction and processing of the gene-protein-reaction relationship (GPR) of the genomic data results a draft reconstruction. GPR associations in the draft model should be reviewed by the researchers and manually curated if the identifying process is achieved with the help of automated computational algorithms [11]. Since the genomic data is the least representative of the biological phenotypes, available transcriptomic, proteomic, metabolomic and/or subcellular localization data are also used to further curate the model. Once the final metabolic network is reconstructed with bibliographic information, it is translated into a mathematical model.

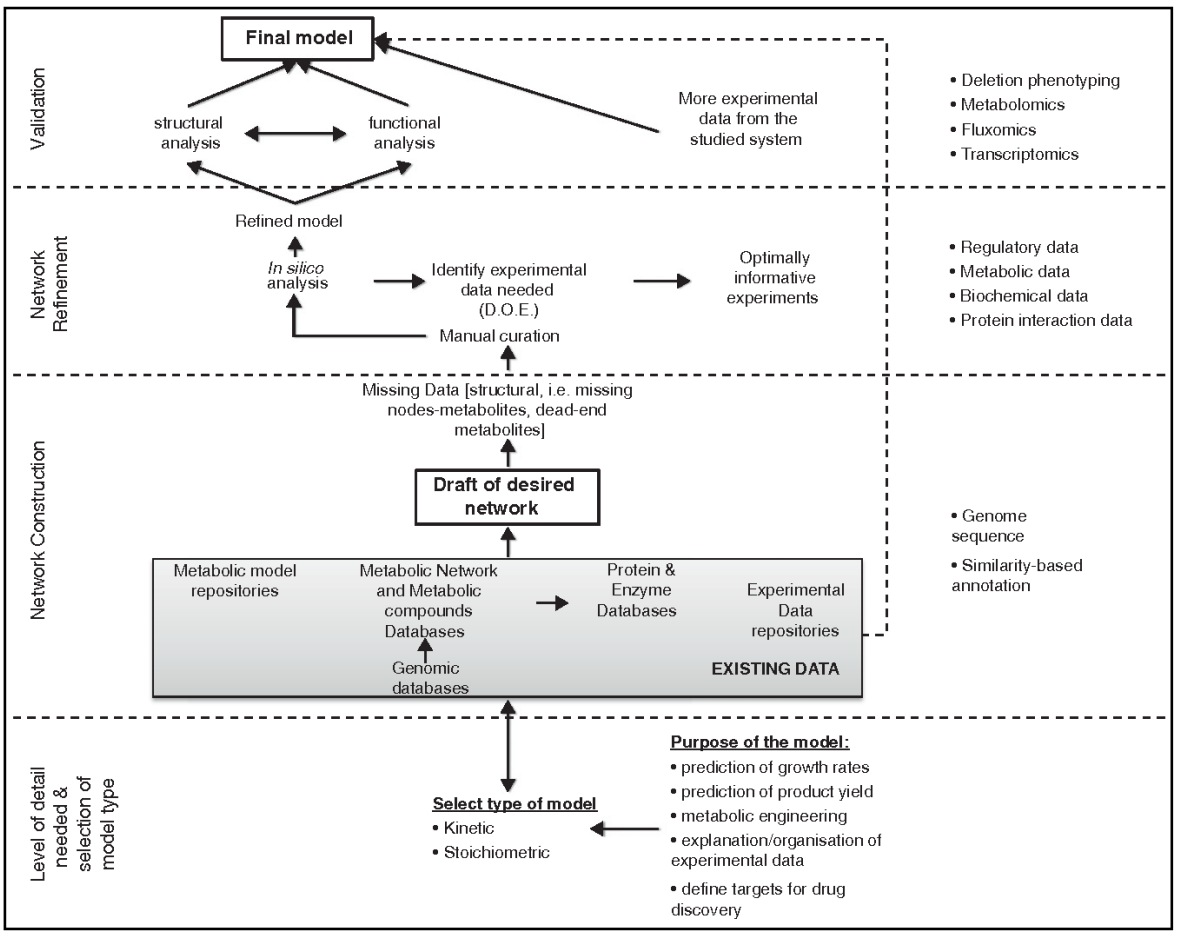

Figure 1.2. Overview of metabolic network reconstruction protocol. Figure is taken from [13].

Once a metabolic network is reconstructed, a rational link between a genome sequence, the proteins encoded in the genome, and the reactions catalyzed by the proteins allowing to investigate the relationships between genotype and phenotype is achieved [14]. As the final step, GSMM needs to be validated by the new experimental data sets. GSMM validation process for various experimental conditions require detailed cultivation data from experiments. For example, information on the biomass composition of the specific organism leads more accurate biomass equation in the model, that is one of the key factors in the GSMM optimization and validation [15]. Even tough multiple steps in the GSMM reconstruction can be achieved with the automated softwares available, it is usually necessary to curate the obtained model manually.

Approaches for analyzing metabolic networks are mainly categorized as dynamic or structural approaches. Even though the former is promising more realistic approach, its implementation in the literature is obstructed due to the unavailability of kinetic parameters for the majority of enzymes within a metabolic network [16, 17] Because of the lack of kinetic parameters, structural metabolic modeling has been widely used for analyzing cellular metabolism at a steady-state assumption as a kind of snapshots taken at specific times.

GSMMs are one of the most useful tools in systems biology, especially in metabolic engineering studies [18]. In 1998, with the publication of *Metabolic Engineering: Principles and Methodologies*, the term metabolic engineering is defined as the optimization of natural processes within cells to increase the production of certain substances [19]. Hence, studies of metabolic engineering can be considered as genetic engineering in strain development. However, while metabolic engineering manipulates strains by altering flux distributions in the pathways; genetic engineering modificates specific genes, proteins and/or enzymes of interest [20]. Although GSMMs are mainly used in metabolic engineering strategies, other applications both for descriptive and predictive purposes can be found in the literature [21].

The ultimate goal of the GSMM reconstruction is to predict flux distribution profiles as close *in silico* as they are *in vivo*. Hence, GSMMs are in continuous research

to improve predictability of organism-specific models.

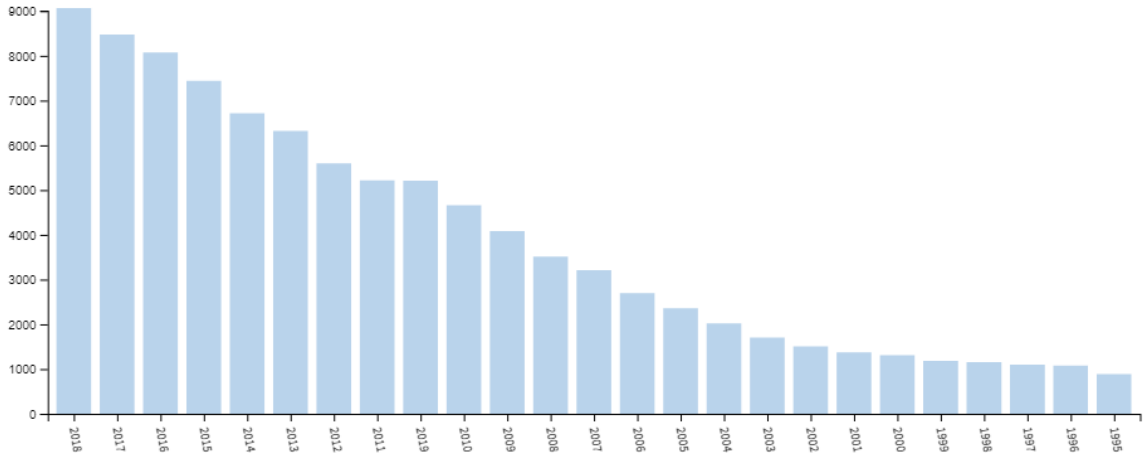

Figure 1.3. Web of Science artice counts on "metabolic model"

1.1.2. Mathematical Representation of Metabolic Networks

In this section, there will be a review on "Mathematical Framework Behind the Reconstruction and Analysis of Genome Scale Metabolic Models" [22].

This section will be detailed as much as possible since biology majors must understand the basics before reading the methods sections. How the biochemical equations are converted into differential equations? Steady state assumption, mass balances? Stoichiometric matrix formulation and its rank, linearly independent vectors, singular values? etc. Subjects will be similar to course ChE529.

1.1.3. Constraint-Based Modeling

In this section, there will be a review on "Bringing genomes to life: the use of genome-scale in silico models" [23]. Topics can be identification of constraints (physicochemical constraints, spatial constraints, environmental constraints, regulatory constraints), list of constraint-based modeling methods (optimal solutions, alternate optima, OptKnock, sampling as an unbiased modeling), linear programming, quadratic

programming ...

1.2. *Saccharomyces cerevisiae*

The species "yeast" includes a range of eukaryotic single-celled microorganisms, although it is commonly used to describe *Saccharomyces cerevisiae*. Also known as the baker's yeast, *S. cerevisiae* is one of the extensively used microorganisms for alcoholic fermentation of beverages, bio-ethanol production, and processing various foods since ancient times [24]. It was the first eukaryotic organism whose genome was fully sequenced and annotated [25], and besides its benefits in the industry, it is used as a model system for other eukaryotic cells including humans [26, 27].

1.2.1. Central Carbon Metabolism of *S. cerevisiae*

From the end of the eighteenth century, mainly after the fermentation is defined as "respiration without oxygen", the metabolism of *S. cerevisiae* has been studied extensively [28, 29]. Its capability to produce ethanol is one of the most characterized microbial processes due to industrial utilization.

The set of anabolic and catabolic reactions in the cell are reffered as the metabolism. A shematic representation of the central carbon metabolism in *S. cerevisiae* can be found in Figure 1.4. Glycolysis, pentose-phosphate pathway (PPP), tricarboxylic acid cycle (TCA) or Krebs cycle, the glyoxylate cycle and the electron transport chain are the main pathways in central carbon metabolism.

In this section, there will be subsections on all biological pathways individually (explaining each in detail -probably referencing Lehninger biochemistry-) especially NAD regulation, fermantation (crabtree effect, industrial applications, bio-ethanol production) etc... Maybe also regulation strategies in cells (feedback/feedforward loops with figures)

1.2.2. Metabolic Models of *S. cerevisiae*

After the first *S. cerevisiae* genome sequence is published, the first cDNA spotted microarray exploring metabolic gene regulation in 1997 [31], and the first commercial platform for oligonucleotide microarray data (Affymetrix) to investigate cellular regulations were reported in 1998 [32]. Existing genome data is integrated with the extensive annotation based on microarray data and biochemical knowledge from literature, leading of the publication of the first GSMM of *S. cerevisiae* in 2013 [33]. More in this section, there will be review on: Genome-scale modeling of yeast: chronology, applications and critical perspectives [34]

1.2.3. Applications of *S. cerevisiae* GSMMs

Literature review on the applications will be added.

1.3. Significance of Thesis

The purpose of this master's thesis is to enlighten molecular mechanisms behind the building tolerance / conditional adaptation in yeast. Intracellular flux distributions of evolved/tolerent strains for various conditions such as caffeine, ethanol, iron, phenylethanol, nickel and sodium chloride in mediums are going to be analyzed comparetively. This study will also contribute to the global understanding of metabolic regulations in the *S. cerevisiae*, and will be further expandable into metabolic engineering studies of evolved strains.

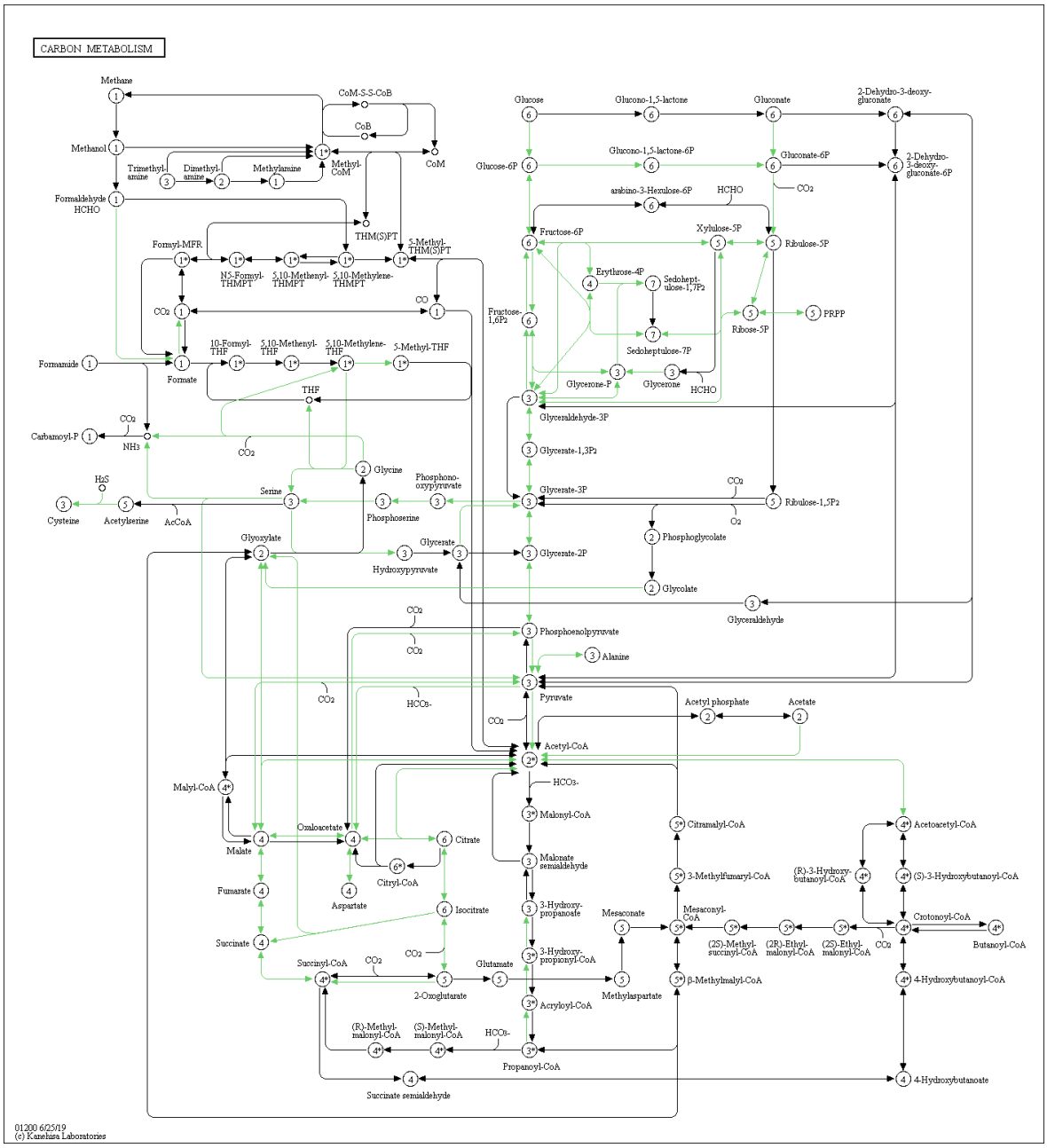

Figure 1.4. Central carbon mechanism of *S. cerevisiae* obtained from KEGG [30].

2. MATERIALS AND METHODS - Yeast8

2.1. Consensus *S. cerevisiae* Metabolic Model

Variety of *S. cerevisiae* genome-scale metabolic models have been used since 2003, and each reconstructed model introduced more manual curations, increasing gene numbers from annotations and better predictions regarding the previous ones [34]. A consensus genome-scale metabolic model of *S. cerevisiae*, Yeast8, is presented in an open-source, version-controlled maintainable way in 2019, claiming that the model can be represented and investigated in a systematic way using Git (https://git-scm.com/) and GitHub (https://github.com/) as a hosting service for the model repository [35]. Systematic way of Yeast8 enables to study simultaneously in collaborative studies, provides record keeping of model changes, version updates, where each version of can be released periodically and accessible all the time (Figure 2.2).

Yeast8 model can be considered as an updated version of Yeast7 [36] with additional corrections based on the annotations available in KEGG and ChEBI, and several gene inclusions from the model iSce926 [37]. Final version of Yeast8, version 8.3.4 released on July 28, has 3991 reactions, 2691 metabolites, 1149 genes, 14 intracellular compartments. Additional statistical analysis results can be seen in Figure 2.1.

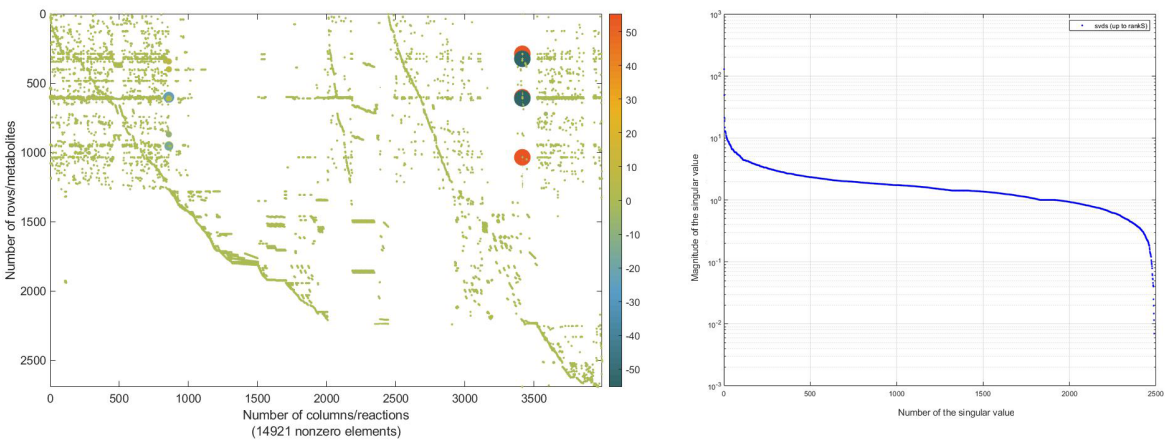

Figure 2.1. Coefficients and singular values of the stoichiometric matrix of Yeast8

All simulations in this chapter are done using Yeast8 v8.3.4 model which is hosted in Github (https://github.com/SysBioChalmers/yeast-GEM).

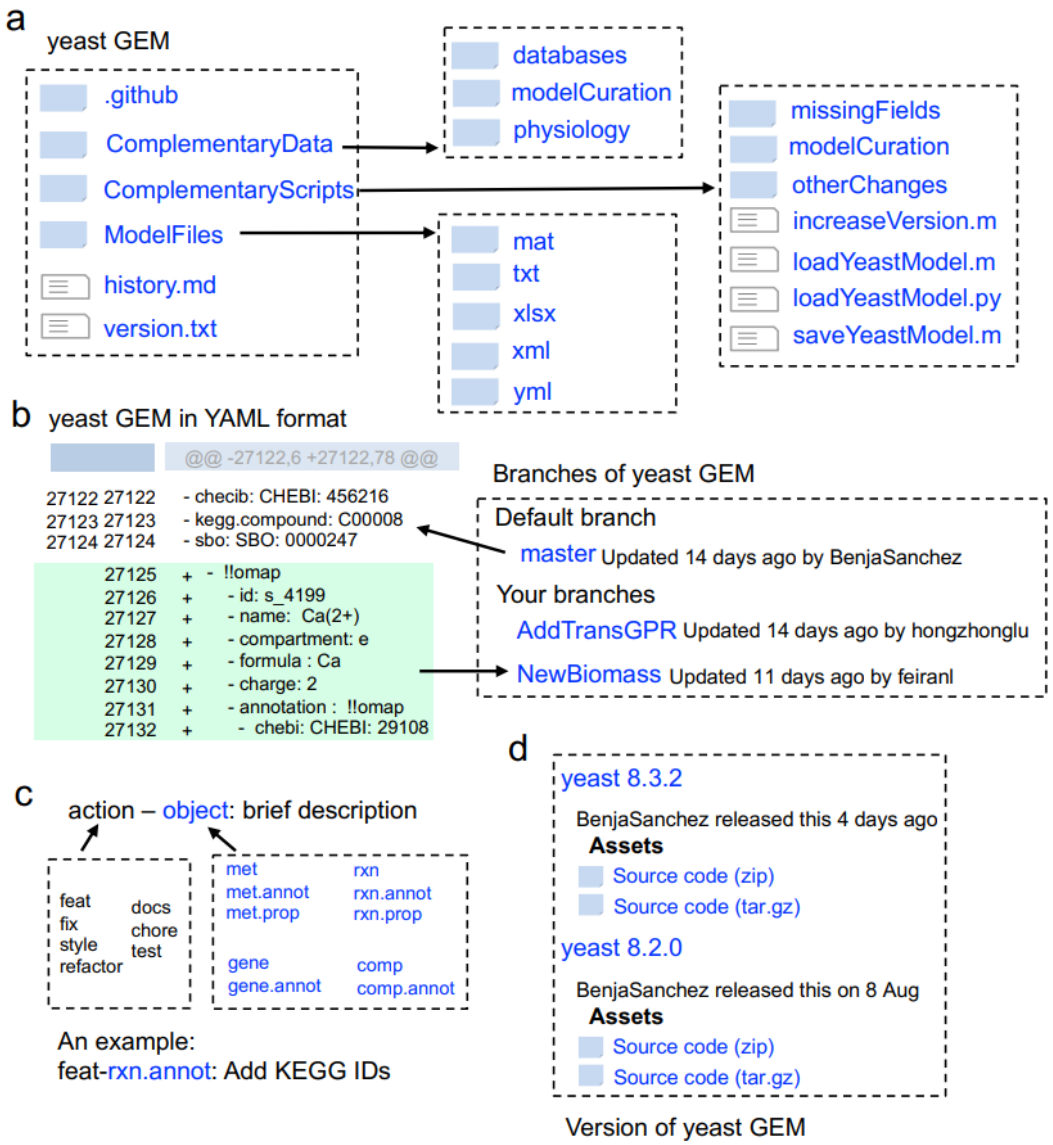

Figure 2.2. Repository of yeast GEM on GitHub. Figure is taken from [35]. will be redrawned in thesis

2.2. Chemostat Simulation: GAM Fitting

In order to make sure the *in-silico* obtained growth rate predictions are in agreement with the physiological kinetic parameters obtained from real experiments, fine adjustment on the energy reactions is a requirement. Since the growth-associated main-

tenance (GAM) and non-growth associated maintenance (NGAM) energy reactions play a determinant role in simulations, fluxes through these reactions must be constrained to a fixed value. Flux of NGAM is constrained to 0.7 mmol $gDWh^{-1}$ for aerobic, and 0 mmol $gDWh^{-1}$ for anaerobic simulations as calculated in the previous studies [38]. For the estimatation of GAM, since it depends on the biomass composition, findings of a chemostat experiment [39] is used as a guide to fit predictions to. Model is simulated iteratively with a range of values for GAM, and the best fit is found at the level of 55.25 mmol $gDWh^{-1}$ (Figure 2.3), GAM flux is constrained accordingly.

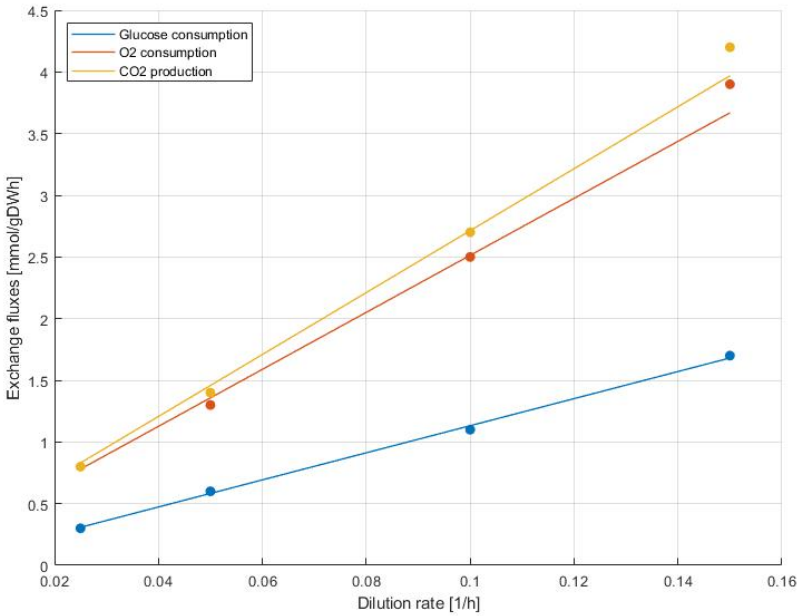

Figure 2.3. Chemostat simulation to re-fit growth associated maintenance.

2.3. Flux Balance Analysis

Flux balance analysis (FBA) assumes that the living cells act as they optimized their lives towards some goal, and as if they were at steady state. To be more clear, steady-state assumption indicates that the metabolites are both produced and consumed at the same rate in a cell, without an accumulation. Therefore, in this system, metabolites are constrained by only the stoichiometric coefficients arised from mass balance of metabolites. As a result of this assumption, FBA solves a set of ordinary

differential equations regarding to the stoichiometric matrix:

$$S_{m\times n} \cdot v = 0 \tag{2.1}$$

where S is the matrix of the stoichiometric reaction coefficients with m number of metabolites (as rows) and n number of reactions (as columns), and v is the vector of all associated reaction fluxes (mmol/gDWh). Because the matrix S usually has more reactions than metabolites (m¡n), the system can result multiple solutions, and being called an underdetermined system. To solve it for an optimal solution, additional constraints are required.

A "growth reaction" is usually included in the reactions of the system to represent the "goal" in the definition of living systems. Growth reactions act as the final consumption of metabolites necessary for the biomass production or cell replication. Additional to the growth, several exchange reactions (uptake or secretion of metabolites from or into extracellular space) are also included. Since the concentrations of extracellular metabolites are measurable experimentall, constraints can be applied to exchange reaction fluxes to shrink solution space. The more constraints introduced into the system, such as reversibility of reactions or known rate values, result smaller solution space. The growth reaction is usually used as an objective function to determine a unique solution from this solution space. The linear problem appears as:

$$\max_v \quad c^T \cdot v \tag{2.2}$$

$$\text{subject to} \quad S_{m\times n} \cdot v = 0 \tag{2.3}$$

$$v_{lb} \leq v \leq v_{ub} \tag{2.4}$$

where c is the objective function vector, v is the vector of fluxes, S is the stoichiometric matrix as above equation. Subscripts lb and ub are the lower and upper boundaries on v. These constraints defines a feasible region of the problem.

In order to simulate batch conditions where minimal yeast medium is used, all the

exchange reactions in the model are blocked first (lower bounds are set to 0). Then, only the exchange reactions of ions that are available to the cells in the experimental design (ammonium, phosphate, sulphate, iron(2+), H+, water, chloride, Mn^{2+}, Zn^{2+}, Mg^{2+}, sodium, $Cu2^{2+}$, Ca^{2+}, potassium) are set free (lower bounds are set to -1000), means that cells can uptake as it needs. While oxygen and glucose uptake rates decreased from 20 mmol $gDWh^{-1}$ and increased to 20 mmol $gDWh^{-1}$, respectively, fluxes of ethanol, acetate, glycerol, formate, succinate secretion reactions with the growth rate is collected (Figure 2.4). I need to show experimental results -preferably from the same article where I got expression data- on the figure to compare with my FBA results

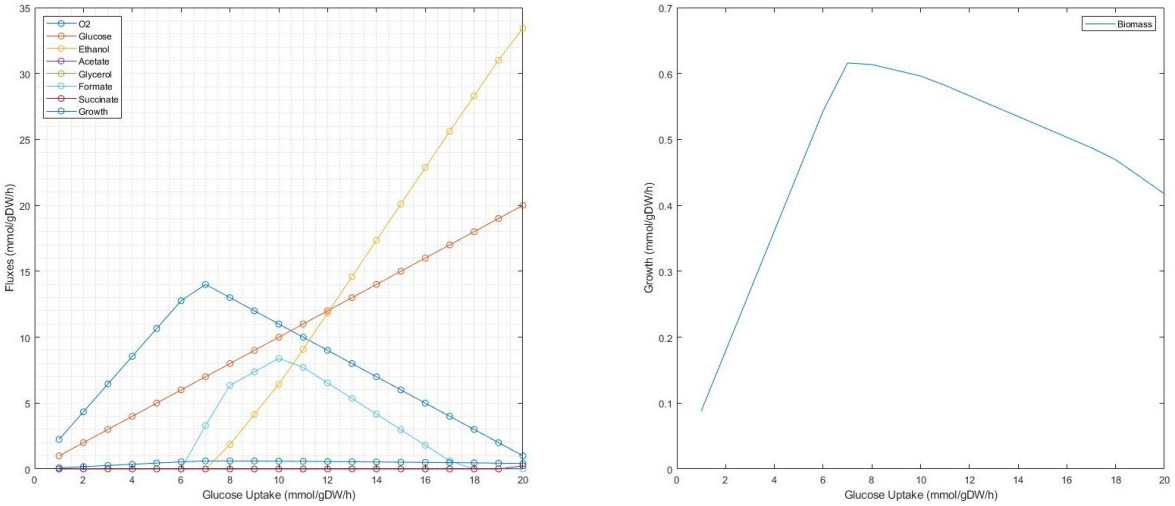

Figure 2.4. Flux balance simulation results where oxygen uptake rate decreased and glucose uptake rate is increased simultaneously. Flux rates of several metabolites on the left, predicted growth rate on the right.

2.4. Flux Variability Analysis

Flux variability analysis (FVA) finds the minimum and maximum available fluxes for each reaction while obeying the provided constraints (for example fixed glucose uptake or growth rate). FVA is mainly used to evaluate the robustness of the model [40], to find alternative optimum states [41], to check flux distributions when growth is not at optimum level [42], and it has many other applications [43].

FVA, similar to FBA, solves two optimization problems for each reaction:

$$\max_v/\min_v \quad v_i \tag{2.5}$$

$$\text{subject to} \quad S_{m\times n} \cdot v = 0 \tag{2.6}$$

$$w^T \cdot v \geq \gamma \cdot Z_0 \tag{2.7}$$

$$v_{lb} \leq v \leq v_{ub} \tag{2.8}$$

where w is the objective function equals to c in the problem 2.3, $Z_0 = w^T \cdot v_0$ describes an optimal solution to the problem 2.3, γ is an indicator to check whether the FVA is done at the optimal state (where objective flux is the same and $\gamma = 1$) or any other state (where $0 \leq \gamma < 1$).

Flux variabilites of Yeast8 reactions are analyzed by solving the linear problem with the objective functions to minimize and maximize all reactions iteratively with tolerance value of 1e-9 using the GUROBI solver. Minimum and maximum available fluxes are collected in the iterative process for each reaction, and results are plotted for glycolysis, pentose phosphate pathway and TCA pathway reactions as error boxes (Figure 2.5). Flux values obtained through the ordinary FBA solution are also shown as a line.

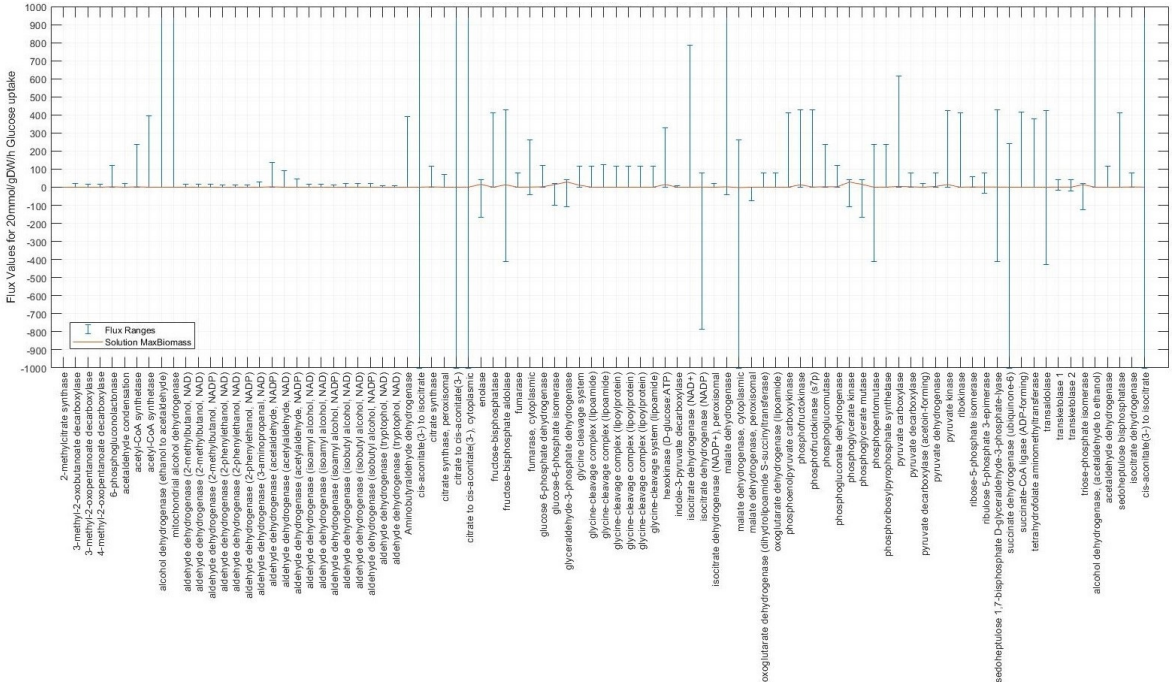

Figure 2.5. Minimum and maximum fluxes of Glycolysis, PPP and TCA reactions.

2.5. Phenotype Phase Plane Construction

As mentioned in the FBA and FVA sections, there is no single solution to the linear problem of the model. Phenotype phase planes (PhPP) are used to describe all feasible metabolic states in a two or three dimentional surfaces, depending on the number of metabolites chosen to see how they affect the objective function [44]. In general, for aerobic models, various levels of glucose and oxygen availability through their uptake reactions are used to generate PhPP surfaces in three dimention with objective function. Fundamentally, PhPP construction refers to a double robustness analysis on the model for selected reactions.

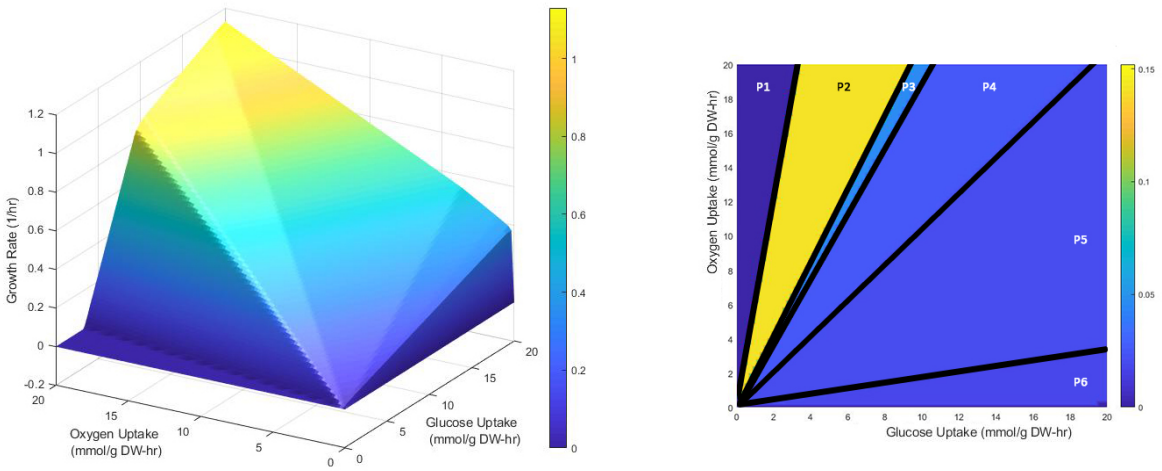

Figure 2.6. Phenotype Phase Plane of Yeast8, corresponding to glucose and oxygen availabilities on the left. Shadow prices of glucose on the right.

In the results section of this, FBA simulations will be discussed with the phases observed in PhPP.

2.6. ACHRB Sampling

2.7. Expression Data Analysis

2.8. Integration of Expression Data Into Model

3. MATERIALS AND METHODS - Petek

3.1. Experimental Data Preparation

3.1.1. Data Acquisition

Extracellular metabolomics data is obtained from Cakar's Lab [45].Briefly, they perform ethyl methane sulfonate (EMS) mutagenesis on the prototrophic *Saccharomyces cerevisiae* strain CEN.PK 113-7D (MATa, MAL2-8c, SUC2) to increase the genetic diversity as an evolutionary engineering selection strategy. Cells were inoculated in 2% Yeast Minimal Media (YMM), and the extracellular concentrations of glucose, ethanol, glycerol and acetate were measured at different time points. OD_{600} values were determined by a spectrophotometer. Additionally, cell dry weight analysis was conducted to determine biomass production. Acquired extracellular metabolite concentrations, OD_{600} values and dry weights of the reference strain (without mutagenesis) were used in this study are collected in Table 3.2 and Table 3.1.

Table 3.1. Measured OD_{600} and cell dry weight values of reference strain.

Time (h)	**OD600**	**ln(OD600)**	**Cell DW (g/L)**
0	0.21	-1.560647748	-
3	0.53	-0.634878272	-
6	1.76	0.565313809	0.9
7.5	2.66	0.978326123	-
9	4.46	1.495148766	1.9
12	5.31	1.669591835	-
15	5.88	1.771556762	-
18	5.83	1.763017	2.32
21	6.07	1.803358605	-
24	5.87	1.769854634	-
30	6.14	1.814824742	2.26
40	6.44	1.86252854	-
46	6.36	1.850028377	-
50	6.3	1.840549633	-
54	6.55	1.87946505	-
63	6.54	1.877937165	-
67	6.88	1.928618652	-
72	6.97	1.941615225	2.66

Table 3.2. Measurements of extracellular concentrations.

Time (h)	Glucose (g/L)	Ethanol (g/L)	Glycerol (g/L)	Acetate (g/L)
0	19.99	0	0	1.08
3	17.98	0.58	0.02	1.24
6	15.85	1.2	0.06	1.16
9	12.21	3.39	0.18	1.37
12	9.18	7.97	0.61	2.45
15	0.4	8.17	0.69	2.46
27	0	8.28	0.76	2.6
46	0	8	0.77	2.45
50	0	6.62	0.64	2.02
54	0	5.74	0.55	1.73
58	0	5.46	0.54	1.74
72	0	3.72	0.49	1.33

3.1.2. Determination of Rates

As the slope in the curve of $\ln OD_{600}$ as a function of time gives the growth rates of cells, natural logarithm of OD_{600} values were calculated to obtain specific growth rates by using the equation 3.1.

$$\mu = \frac{\Delta \ln OD_{600}}{\Delta t} \tag{3.1}$$

In order to determine uptake and secretion rates of the metabolites, the steady-state assumption is applied in three hours intervals as the shortest measured time-points. Missing data on cell dry weights are estimated from the OD_{600} values, and these cell dry weight data is used to calculate fluxes (in the unit of mmol/gDWh). Measurement of the cell dry weight at the 3rd hour was crucial for the steady-state assumption, however

data was not available from the experiments. Curve trend of the OD_{600} plot is used as a guide to estimate cell dry weight (Figure 3.1) Need a method here: Estimation approach, maybe regression or curve fitting?. Calculated flux values can be found in the Table 3.3.

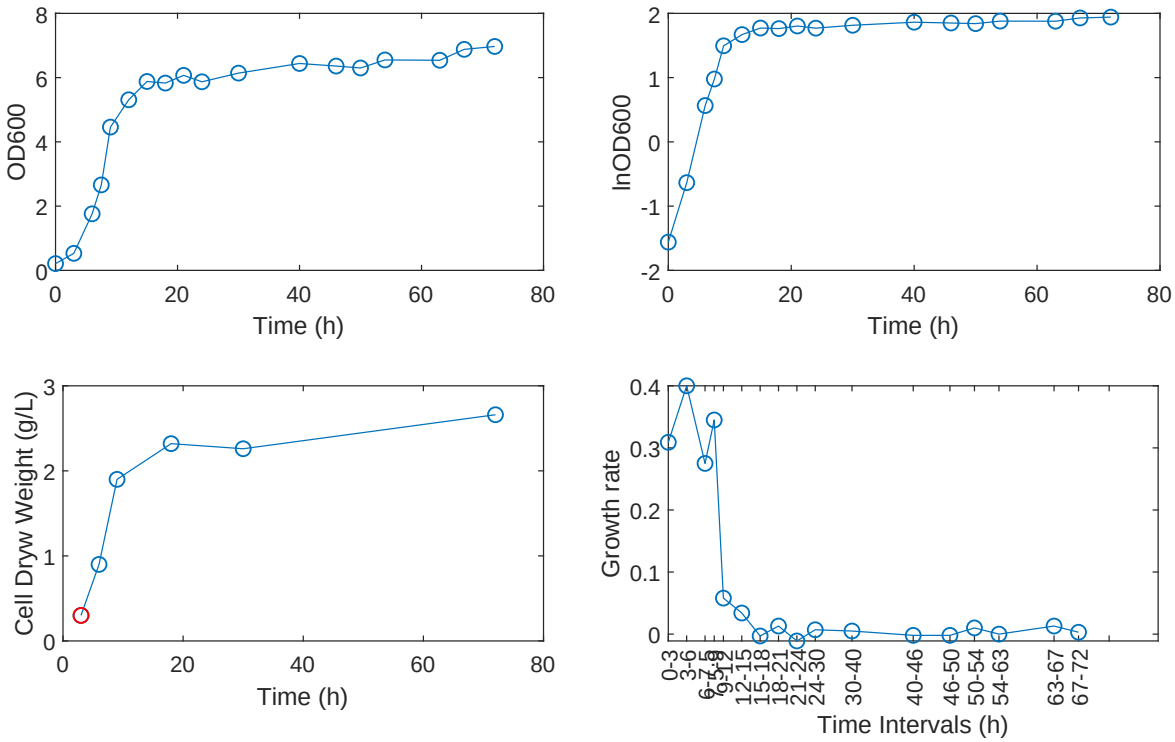

Figure 3.1. OD_{600}, $lnOD_{600}$, cell dry weights and growth rates graphs. Estimated missing cell dry weight data is shown in red color.

Table 3.3. Calculated flux values.

Time	**Metabolite fluxes in mmol/gDWh**				**Growth h-1**
	Glucose	**Ethanol**	**Glycerol**	**Acetate**	**Biomass**
0-3	-12.3963884	13.98837518	0.24131	2.960496	0.30859
3-6	-4.378823762	4.984363569	0.160873	-0.49342	0.400064

3.2. Model Selection

iAN50 [38], a stoichiometric model of intermediary metabolism including glycolysis, the pentose phosphate pathway (PPP), anaerobic excretion, citric acid cycle (TCA cycle), oxidative phosphorylation, and uptake pathways for galactose, ethanol and acetate is used to simulate batch conditions. The model was implemented based on the first GSMM of yeast, iFF708 [33] with the addition of the yeast intermediary metabolism from the RAVEN toolbox [46], and finally curated using the KEGG database [30].

Biomass equation (Eq. 3.2) in the model was left as a function of sub-reactions, for example protein, lipid, DNA synthesis reactions, so that the coefficients of biomass constituents could be optimized for each FBA simulation.

$$\begin{aligned} &0.5185\ \text{Glycogen} + 0.0234\ \text{Trehalose} + 0.8079\ \text{Mannan} + 1.1348\ \text{Glucan} \\ &+ 0.1966\ \text{DNA} + 0.012\ \text{DNA} + 4.14\ \text{Protein} + 0.0269\ \text{Lipid} + 35.3630\ \text{Maintainance} \\ &= \text{BIOMASS} \end{aligned} \quad (3.2)$$

Enzymatic reactions that are found in the model is collected in Table **??**. In contrast to other GSMM's, each reaction was irreversible in the iAN50. This was achieved by splitting each reversible reaction into two seperate reactions in both directions.

Another reason to select iAN50 is that the total masses of enzymes catalyzing reactions in the GSMM were already estimated, and fluxes through these reactions were constrained to the biologic level, using an approach similar to intracellular crowding method using kinetic parameters [47, 48]. To clarify the method briefly, a flux value for each reaction was obtained by applying a standart flux balance analysis, and this value was divided by the maximum *in vitro* activity collected from the enzyme database BRENDA [49], and a saturation factor of 0.5 (half) for simplification. Therefore, the mass of the enzymes required for that particular reaction was estimated and the

constrains are applied to the corresponding enzymatic reactions.

NAME	EQUATION	GENE ASSOCIATION	EC-NUMBER	SUBSYSTEM
Hexokinase	GLC[c] + ATP[c] =>ADP[c] + G6P[c]	YFR053C; YGL253W; YCL040W	2.7.1.1 OR 2.7.1.2	Glycolysis
Glucose-6-phosphate isomerase	G6P[c] <=>F6P[c]	YBR196C	5.3.1.9	Glycolysis
Phosphofructokinase	ATP[c] + F6P[c] =>ADP[c] + F16P[c]	YGR240C; YMR205C	2.7.1.11	Glycolysis
Fructose-1,6-bisphosphatase	F16P[c] =>F6P[c] + PI[c]	YLR377C	3.1.3.11	Glycolysis
Fructose-bisphosphate aldolase	F16P[c] <=>GA3P[c] + DHAP[c]	YKL060C	4.1.2.13	Glycolysis
Triosephosphate isomerase	DHAP[c] <=>GA3P[c]	YDR050C	5.3.1.1	Glycolysis
Triosephosphate dehydrogenase	GA3P[c] + NAD[c] + PI[c] <=>P13G[c] + NADH[c]	YJL052W; YJR009C; YGR192C	1.2.1.12	Glycolysis
Phosphoglycerate kinase	P13G[c] + ADP[c] <=>P3G[c] + ATP[c]	YCR012W	2.7.2.3	Glycolysis
Phosphoglycerate mutase	P3G[c] <=>P2G[c]	YKL152C; YDL021W; YOL056W	5.4.2.11	Glycolysis
Enolase	P2G[c] <=>PEP[c]	YGR254W; YHR174W; YOR393W; YPL281C; YMR323W	4.2.1.11	Glycolysis
Pyruvate kinase	ADP[c] + PEP[c] =>ATP[c] + PYR[c]	YOR347C; YAL038W	2.7.1.40	Glycolysis
Glucose-6-phosphate 1-dehydrogenase	G6P[c] + NADP[c] =>G15L[c] + NADPH[c]	YNL241C	1.1.1.49	Pentose Phosphate
6-phosphogluconolactonase	G15L[c] =>P6G[c]	YNR034W; YCR073W-A; YHR163W; YGR248W	3.1.1.31	Pentose Phosphate
6-phosphogluconate dehydrogenase, decarboxylating	P6G[c] + NADP[c] =>CO2[c] + RU5P[c] + NADPH[c]	YGR256W; YHR183W	1.1.1.44	Pentose Phosphate
ribose 5-phosphate isomerase	RU5P[c] <=>R5P[c]	YOR095C	5.3.1.6	Pentose Phosphate
Ribulose-phosphate 3-epimerase	RU5P[c] <=>X5P[c]	YJL121C	5.1.3.1	Pentose Phosphate
Transketolase	R5P[c] + X5P[c] <=>GA3P[c] + S7P[c]	YBR117C; YPR074C	2.2.1.1	Pentose Phosphate
Transaldolase	GA3P[c] + S7P[c] <=>F6P[c] + E4P[c]	YLR354C	2.2.1.2	Pentose Phosphate
Transketolase	E4P[c] + X5P[c] <=>F6P[c] + GA3P[c]	YBR117C; YPR074C	2.2.1.1	Pentose Phosphate
Pyruvate carboxylase 1	ATP[c] + CO2[c] + PYR[c] =>ADP[c] + OAA[c] + PI[c]	YGL062W; YBR218C	6.4.1.1	TCA
Citrate synthase, mitochondrial	ACCOA[m] + OAA[m] =>CI[m] + COA[m]	YNR001C; YPR001W	2.3.3.1	TCA
Aconitate hydratase, mitochondrial	CI[m] <=>ICI[m]	YLR304C	4.2.1.3	TCA
Isocitrate dehydrogenase	ICI[m] + NAD[m] =>AKG[m] + CO2[m] + NADH[m]	YNL037C; YOR136W	1.1.1.41	TCA
Isocitrate dehydrogenase [NADP], mitochondrial	ICI[m] + NADP[m] =>AKG[m] + CO2[m] + NADPH[m]	YDL066W; YLR174W	1.1.1.42	TCA
Alpha-ketoglutarate dehydrogenase	AKG[m] + NAD[m] + ADP[m] + PI[m] =>CO2[m] + NADH[m] + ATP[m] + SUC[m]	YDR148C; YIL125W	1.2.4.2	TCA
Succinate dehydrogenase complex	FAD[m] + SUC[m] =>FADH2[m] + FUM[m]	YKL148C; YLL041C	1.3.5.1	TCA
Fumarate reductase	FADH2[m] + FUM[m] =>FAD[m] + SUC[m]	YJR051W; YEL047C	1.3.1.6	TCA
Fumarate hydratase	FUM[m] <=>MAL[m]	YPL262W	4.2.1.2	TCA
Malate dehydrogenase	MAL[m] + NAD[m] <=>NADH[m] + OAA[m]	YKL085W	1.1.1.37	TCA
Galactokinase	GAL[c] + ATP[c] =>GALP[c] + ADP[c]		2.7.1.6	Galactose metabolism
UDP-glucose 4-epimerase	GALUDP[c] <=>GLUUDP[c]		5.1.3.2	Galactose metabolism
Galactose-1-phosphate uridylyltransferase	GLUUDP[c] + GALP[c] <=>G1P[c] + GALUDP[c]		2.7.7.12	Galactose metabolism
Phosphoglucomutase-1	G1P[c] <=>G6P[c]		5.4.2.2	Galactose metabolism

NAME	EQUATION	GENE ASSOCIATION	EC-NUMBER	SUBSYSTEM
glycerol-3-phosphate dehydrogenase	DHAP[c] + NADH[c] =>GP[c] + NAD[c]	YDL022W; YOL059W	1.1.1.8	Anaerobic excretion
sn-glycerol-3-phosphate phosphohydrolase	GP[c] =>GLY[c] + PI[c]	YER062C; YIL053W	3.1.3.21	Anaerobic excretion
Pyruvate decarboxylase	PYR[c] =>ACA[c] + CO2[c]	YGR087C; YLR134W; YLR044C	4.1.1.1	Anaerobic excretion
Alcohol dehydrogenase	ACA[c] + NADH[c] <=>ETH[c] + NAD[c]	YGL256W; YMR303C; YOL086C	1.1.1.1	Anaerobic excretion
Aldehyde dehydrogenase	ACA[c] + NADP[c] =>AC[c] + NADPH[c]	YPL061W	1.2.1.3	Anaerobic excretion
Aldehyde dehydrogenase [NAD(P)+] 1	ACA[c] + NAD[c] =>NADH[c] + AC[c]	YMR170C; YMR169C; YOR374W; YOR374W; YER073W	1.2.1.5	Aromatic amino acid biosynthesis
Isocitrate lyase	ICI[m] =>Glyoxylate[m] + SUC[m]	YER065C; YPR006C	4.1.3.1	Anaplerotic reactions
Malate synthase 1, glyoxysomal	ACCOA[m] + Glyoxylate[m] =>COA[m] + MAL[m]	YIR031C; YNL117W	2.3.3.9	Anaplerotic reactions
NADH-ubiquinone oxidoreductase, mitochondrial ("Complex 1")	NADH[m] + Ubiquinone-9[m] =>Ubiquinol[m] + NAD[m]	YML120C; YMR145C	1.6.5.3 (1.6.5.9)	Oxidative Phosphorylation
External NADH-ubiquinone oxidoreductase 2, mitochondrial ("Complex 1")	NADH[c] + Ubiquinone-9[m] =>Ubiquinol[m] + NAD[c]	YDL085W	1.6.5.3 (1.6.5.9)	Oxidative Phosphorylation
Succinate dehydrogenase [ubiquinone] cytochrome b subunit, mitochondrial (Complex II)	FADH2[m] + Ubiquinone-9[m] <=>FAD[m] + Ubiquinol[m]	YKL141W; YDR178W	1.3.5.1	Oxidative Phosphorylation
Cytochrome b-c1 complex subunit Rieske, mitochondrial (complex III)	Ubiquinol[m] + 2 Ferricytochrome_c[m] + 1.5 H[m] =>Ubiquinone-9[m] + 2 Ferrocytochrome_c[m] + 1.5 HMit[m]	YEL024W; YBL045C; YHR001W-A; YPR191W; YFR033C; YDR529C; YJL166W; YGR183C	1.10.2.2	Oxidative Phosphorylation
Cytochrome c oxidase subunit 1 (Complex IV)	Ferrocytochrome_c[m] + 0.25 O2[c] + 1.5 H[m] =>Ferricytochrome_c[m] + 1.5 HMit[m]	YNL052W; YIL111W; YLR395C; Q0045; Q0250; YGL187C; YHR051W; YGL191W; YLR038C; YMR256C; YDL067C	1.9.3.1	Oxidative Phosphorylation
ATP synthase subunit alpha, mitochondrial (Complex V)	ADP[m] + PI[m] + 3 HMit[m] =>ATP[m] + 3 H[m]	YBL099W; Q0080; YPL078C; YDR298C; Q0130; Q0085; YJR121W; YKL016C; YDL004W; YDR322C-A; YPL271W; YDR377W; YPR020W; YBR039W; YLR295C; YML081C-A; YOL077W-A; YML042W	3.6.3.14	Oxidative Phosphorylation
ATP hydrolysis	ATP[c] =>ADP[c] + PI[c]			Other
Pyruvate dehydrogenase complex	COA[m] + NAD[m] + PYR[m] =>ACCOA[m] + CO2[m] + NADH[m]	YER178W; YFL018C; YBR221C	1.2.4.1	Other
Acetyl-coenzyme A synthetase 1	AC[c] + 2 ATP[c] + COA[c] =>ACCOA[c] + 2 ADP[c] + 2 PI[c]	YAL054C; YLR153C	6.2.1.1	Other
Phosphoenolpyruvate carboxykinase	ATP[c] + OAA[c] =>ADP[c] + CO2[c] + PEP[c]	YKR097W	4.1.1.49	Other

3.3. Flux Balance Analysis

Uptake reaction of glucose with the secretion reactions of glycerol and acetate were constrained according to the calculated flux values in Table 3.3, for both time intervals seperately. Since the main goal was to validate model for experimental conditions, ethanol was not constrained in regard to be used as the control metabolite. Experiments were done in fully aerobic conditions, therefore oxygen uptake reaction was set unlimited.

Coefficients of the biomass constituents are defined as the same as the batch conditions in the reference article [38], for the reason that detailed knowledge is not available in the acquired experimental data. Coefficients for the final biomass equation can be found in the Table 3.5.

Table 3.5. Biomass coefficients that are used in the simulation.

Constituent	**Coefficient**
Protein	3.703704
RNA	0.37037
DNA	0.018519
Lipid	0.041667
Glycogen	0.030864
Trehalose	0.029214
Mannan	0
Glucan	2.469136
Maintainance	40

For the linear optimization, an implementation of pFBA from the reference publication was performed [38, 50]. After solving the system using a linear solver with the objective maximizing growth, the solution was used as a constraint. From that point, a second optimization was run to minimize the sum of all other fluxes.

3.4. Visualization of the Model

The GSMM was visualized in Cytoscape [51] using the Fluxviz plug-in [52] and the solution fluxes were mapped onto edges. Network was imported in SBML [53] and the flux distributions were imported in XML format prepared according to the plug-ins guide.

4. RESULTS

4.1. Intracellular Flux Distributions

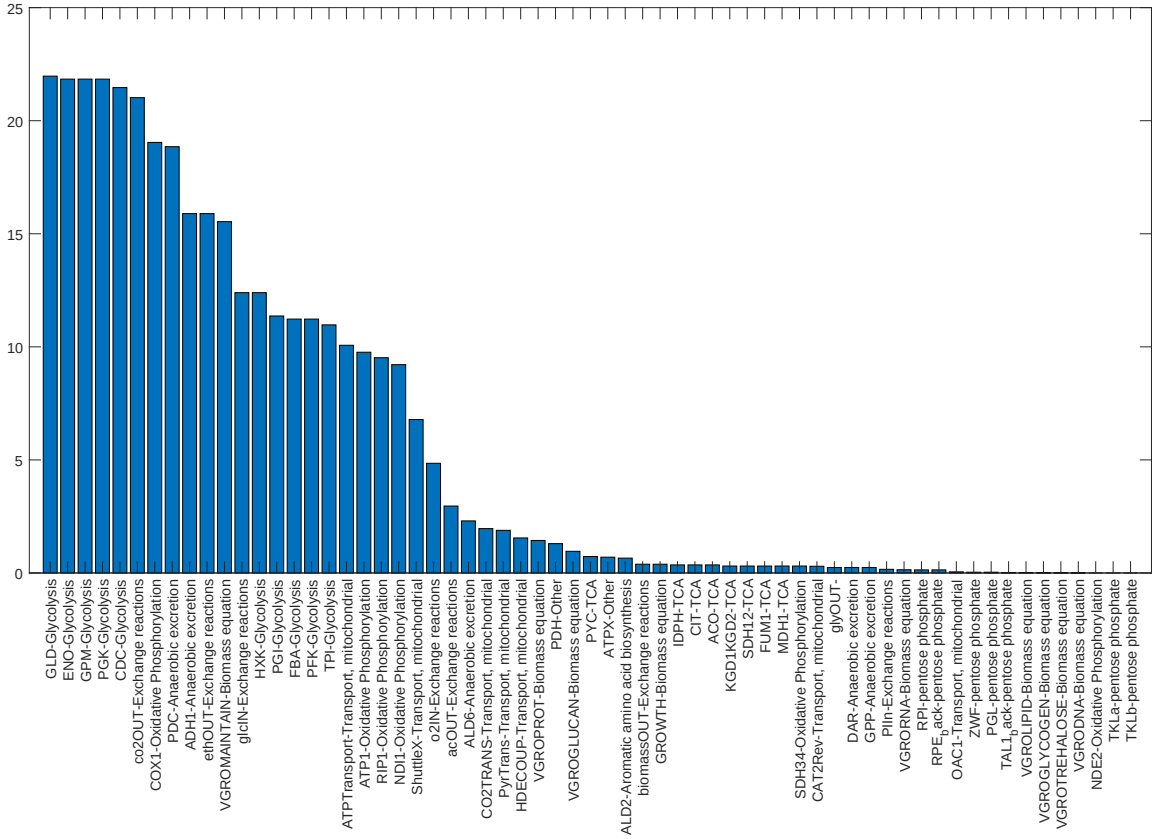

Figure 4.1. Non-zero fluxes in the solution.

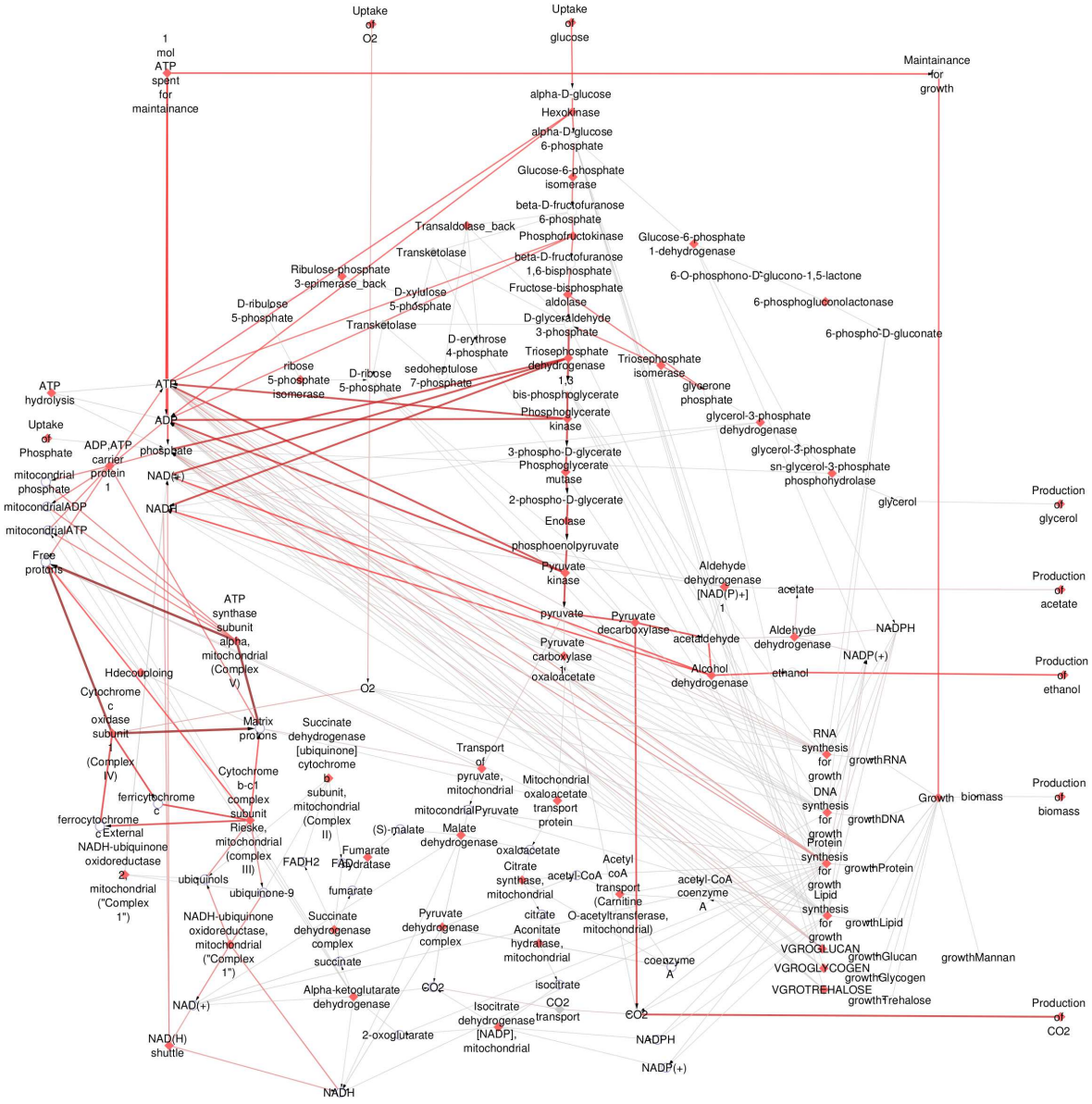

Figure 4.2. Non-zero flux distributions on the map.

5. DISCUSSION

6. EXPANDABLE TOPICS

6.1. Following will be added into introduction

, Toolboxes (such as COBRA, RAVEN or Other Softwares and Algorithms), FBA Methods (pFBA, dFBA, TFBA etc), Omics Data (subtopics for each omics, Model Integration Methods, Transcriptomics Analysis), metabolic control analysis...

REFERENCES

1. Kitano, H., "Systems biology: a brief overview", *science*, Vol. 295, No. 5560, pp. 1662–1664, 2002.

2. Bellouquid, A. and M. Delitala, *Mathematical modeling of complex biological systems*, Springer, 2006.

3. Kremling, A., *Systems biology: mathematical modeling and model analysis*, Chapman and Hall/CRC, 2013.

4. Bruggeman, F. J. and H. V. Westerhoff, "The nature of systems biology", *TRENDS in Microbiology*, Vol. 15, No. 1, pp. 45–50, 2007.

5. Shahzad, K. and J. J Loor, "Application of top-down and bottom-up systems approaches in ruminant physiology and metabolism", *Current Genomics*, Vol. 13, No. 5, pp. 379–394, 2012.

6. Thiele, I. and B. Ø. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction", *Nature protocols*, Vol. 5, No. 1, p. 93, 2010.

7. Orth, J. D., I. Thiele and B. Ø. Palsson, "What is flux balance analysis?", *Nature biotechnology*, Vol. 28, No. 3, p. 245, 2010.

8. Vallino, J. J. and G. Stephanopoulos, "Carbon flux distributions at the glucose 6-phosphate branch point in Corynebacterium glutamicum during lysine overproduction", *Biotechnology Progress*, Vol. 10, No. 3, pp. 327–334, 1994.

9. Varma, A., B. W. Boesch and B. O. Palsson, "Biochemical production capabilities of Escherichia coli", *Biotechnology and bioengineering*, Vol. 42, No. 1, pp. 59–73, 1993.

10. Feist, A. M., M. J. Herrgård, I. Thiele, J. L. Reed and B. Ø. Palsson, "Reconstruction of biochemical networks in microorganisms", *Nature Reviews Microbiology*, Vol. 7, No. 2, p. 129, 2009.

11. Pitkänen, E., P. Jouhten, J. Hou, M. F. Syed, P. Blomberg, J. Kludas, M. Oja, L. Holm, M. Penttilä, J. Rousu *et al.*, "Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species", *PLoS computational biology*, Vol. 10, No. 2, p. e1003465, 2014.

12. Kerkhoven, E. J., P.-J. Lahtvee and J. Nielsen, "Applications of computational modeling in metabolic engineering of yeast", *FEMS Yeast Res*, Vol. 15, No. 1, pp. 1567–1364, 2014.

13. Chen, N., I. J. del Val, S. Kyriakopoulos, K. M. Polizzi and C. Kontoravdi, "Metabolic network reconstruction: advances in in silico interpretation of analytical information", *Current opinion in biotechnology*, Vol. 23, No. 1, pp. 77–82, 2012.

14. Durot, M., P.-Y. Bourguignon and V. Schachter, "Genome-scale models of bacterial metabolism: reconstruction and applications", *FEMS microbiology reviews*, Vol. 33, No. 1, pp. 164–190, 2008.

15. Dikicioglu, D., B. Kırdar and S. G. Oliver, "Biomass composition: the "elephant in the room" of metabolic modelling", *Metabolomics*, Vol. 11, No. 6, pp. 1690–1701, 2015.

16. Machado, D. and M. Herrgård, "Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism", *PLoS computational biology*, Vol. 10, No. 4, p. e1003580, 2014.

17. Ramkrishna, D. and H.-S. Song, "Dynamic models of metabolism: Review of the cybernetic approach", *AIChE Journal*, Vol. 58, No. 4, pp. 986–997, 2012.

18. Kim, T. Y., S. B. Sohn, Y. B. Kim, W. J. Kim and S. Y. Lee, "Recent advances in reconstruction and applications of genome-scale metabolic models", *Current opinion in biotechnology*, Vol. 23, No. 4, pp. 617–623, 2012.

19. Stephanopoulos, G., "Metabolic fluxes and metabolic engineering", *Metabolic engineering*, Vol. 1, No. 1, pp. 1–11, 1999.

20. Stephanopoulos, G., "Synthetic biology and metabolic engineering", *ACS synthetic biology*, Vol. 1, No. 11, pp. 514–525, 2012.

21. Österlund, T., I. Nookaew and J. Nielsen, "Fifteen years of large scale metabolic modeling of yeast: developments and impacts", *Biotechnology advances*, Vol. 30, No. 5, pp. 979–988, 2012.

22. Pinzon, W., H. Vega, J. Gonzalez and A. Pinzon, "Mathematical Framework Behind the Reconstruction and Analysis of Genome Scale Metabolic Models", *Archives of Computational Methods in Engineering*, pp. 1–14, 2018.

23. Thiele, I. and B. Ø. Palsson, "Bringing genomes to life: the use of genome-scale in silico models", *Introduction to Systems Biology*, pp. 14–36, Springer, 2007.

24. Gélinas, P., "Inventions on baker's yeast strains and specialty ingredients", *Recent patents on food, nutrition & agriculture*, Vol. 1, No. 2, pp. 104–132, 2009.

25. Goffeau, A., J. Park, I. T. Paulsen, J.-L. JONNIAUX, T. Dinh, P. Mordant and M. H. SAIER JR, "Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames within the major facilitator superfamily", *Yeast*, Vol. 13, No. 1, pp. 43–54, 1997.

26. Dujon, B., "The yeast genome project: what did we learn?", *Trends in Genetics*, Vol. 12, No. 7, pp. 263–270, 1996.

27. Botstein, D., S. A. Chervitz and M. Cherry, "Yeast as a model organism", *Science*,

Vol. 277, No. 5330, pp. 1259–1260, 1997.

28. Barnett, J. A., "A history of research on yeasts 1: work by chemists and biologists 1789–1850", *Yeast*, Vol. 14, No. 16, pp. 1439–1451, 1998.

29. Barnett, J. A., "A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850–1880", *Yeast*, Vol. 16, No. 8, pp. 755–771, 2000.

30. Kanehisa, M. and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes", *Nucleic acids research*, Vol. 28, No. 1, pp. 27–30, 2000.

31. DeRisi, J. L., V. R. Iyer and P. O. Brown, "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, Vol. 278, No. 5338, pp. 680–686, 1997.

32. Cho, R. J., M. Fromont-Racine, L. Wodicka, B. Feierbach, T. Stearns, P. Legrain, D. J. Lockhart and R. W. Davis, "Parallel analysis of genetic selections using whole genome oligonucleotide arrays", *Proceedings of the National Academy of Sciences*, Vol. 95, No. 7, pp. 3752–3757, 1998.

33. Förster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network", *Genome research*, Vol. 13, No. 2, pp. 244–253, 2003.

34. Lopes, H. and I. Rocha, "Genome-scale modeling of yeast: chronology, applications and critical perspectives", *FEMS yeast research*, Vol. 17, No. 5, 2017.

35. Lu, H., F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P. M. Anton, D. Lappa, C. Lieven *et al.*, "A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism", *Nature communications*, Vol. 10, No. 1, pp. 1–13, 2019.

36. Aung, H. W., S. A. Henry and L. P. Walker, "Revising the representation of fatty

acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism", *Industrial biotechnology*, Vol. 9, No. 4, pp. 215–228, 2013.

37. Chowdhury, R., A. Chowdhury and C. Maranas, "Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models", *Metabolites*, Vol. 5, No. 4, pp. 536–570, 2015.

38. Nilsson, A. and J. Nielsen, "Metabolic trade-offs in yeast are caused by F1F0-ATP synthase", *Scientific reports*, Vol. 6, p. 22264, 2016.

39. Van Hoek, P., J. P. Van Dijken and J. T. Pronk, "Effect of specific growth rate on fermentative capacity of baker's yeast", *Appl. Environ. Microbiol.*, Vol. 64, No. 11, pp. 4226–4233, 1998.

40. Thiele, I., R. M. Fleming, A. Bordbar, J. Schellenberger and B. Ø. Palsson, "Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery", *Biophysical journal*, Vol. 98, No. 10, pp. 2072–2081, 2010.

41. Mahadevan, R. and C. Schilling, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models", *Metabolic engineering*, Vol. 5, No. 4, pp. 264–276, 2003.

42. Reed, J. L. and B. Ø. Palsson, "Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states", *Genome research*, Vol. 14, No. 9, pp. 1797–1805, 2004.

43. Gudmundsson, S. and I. Thiele, "Computationally efficient flux variability analysis", *BMC bioinformatics*, Vol. 11, No. 1, p. 489, 2010.

44. Edwards, J. S., R. Ramakrishna and B. O. Palsson, "Characterizing the metabolic phenotype: a phenotype phase plane analysis", *Biotechnology and bioengineering*,

Vol. 77, No. 1, pp. 27–36, 2002.

45. Arslan, M., C. Holyavkin, H. İ. Kısakesen, A. Topaloğlu, Y. Sürmeli and Z. P. Çakar, "Physiological and transcriptomic analysis of a chronologically long-lived Saccharomyces cerevisiae strain obtained by evolutionary engineering", *Molecular biotechnology*, Vol. 60, No. 7, pp. 468–484, 2018.

46. Agren, R., L. Liu, S. Shoaie, W. Vongsangnak, I. Nookaew and J. Nielsen, "The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum", *PLoS computational biology*, Vol. 9, No. 3, p. e1002980, 2013.

47. Beg, Q. K., A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási and Z. N. Oltvai, "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity", *Proceedings of the National Academy of Sciences*, Vol. 104, No. 31, pp. 12663–12668, 2007.

48. Adadi, R., B. Volkmer, R. Milo, M. Heinemann and T. Shlomi, "Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters", *PLoS computational biology*, Vol. 8, No. 7, p. e1002575, 2012.

49. Schomburg, I., A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote *et al.*, "BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA", *Nucleic acids research*, Vol. 41, No. D1, pp. D764–D772, 2012.

50. Nilsson, A., "solveLinMin: Code of Implementation of pFBA", GitHub, SysBioChalmers, EnzymeConstrainedSmallYeast, sourceCode, solveLinMin, 2019.

51. Cline, M. S., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross *et al.*, "Integration of biological networks and gene expression data using Cytoscape", *Nature protocols*, Vol. 2,

No. 10, p. 2366, 2007.

52. König, M. and H.-G. Holzhütter, "Fluxviz—Cytoscape plug-in for visualization of flux distributions in networks", *Genome Informatics 2010: Genome Informatics Series Vol. 24*, pp. 96–103, World Scientific, 2010.

53. Hucka, M., F. T. Bergmann, A. Dräger, S. Hoops, S. M. Keating, N. Le Novère, C. J. Myers, B. G. Olivier, S. Sahle, J. C. Schaff *et al.*, "The Systems Biology Markup Language (SBML): language specification for level 3 version 2 core", *Journal of integrative bioinformatics*, Vol. 15, No. 1, 2018.