1. INTRODUCTION

Evolution has been intensively studied since the publication of *On the Origin of Species* in order to untie the mysteries behind fascinating machineries within living systems. In a broader sense, most of the studies on evolutionary biology have two main goals: (I) To document history of life through evolutionary point of view, and (2) to understand causal mechanisms responsible for the biological diversity that we have on Earth [1, 2]. In addition to these goals, the idea of using pre-existing living systems as cell factories for industrial purposes gained close attention in the last decades [3].

Evolutionary engineering studies, adaptive laboratory evolution in particular, refers to the experiments in which the environmental conditions are altered gradually to obtain adapted populations in the laboratories [4]. Accumulation of mutations obtained due to environmental alterations through generations, where the favored individuals (mainly the ones with increased fitness) are selected to become parents of the next generation, result a population with advantageous traits compared to starting population. Since one of the main challenges in the evolutionary engineering field is being able to analyze experimental to answer fundamental questions, adaptive evolution studies usually require interdisciplinary research in order to answer fundamental questions. What has changed in the cells through generations? What are the genetic basis for adaptation? Can we evolve any organism to any condition? If so, how? These questions and many more are asked everyday, and the corresponding answers potentially give rise to more questions.

The purpose of this thesis is to enlighten metabolic changes in the adaptation of the yeast *S. cerevisiae* using computational methods, specifically genome-scale metabolic models. Evolved yeast strains (such as ethanol tolerant, long-lived, multi-stress resistant strains) are going to be analyzed comparetively to the unevolved strains using intracellular flux distributions obtained with the integration of transcriptomics data. This study will also contribute to the global understanding of metabolic regulations in yeast, and will be further expandable into metabolic engineering studies.

2. THEORETICAL BACKGROUND

2.1. Adaptive Laboratory Evolution

Since the very first laboratory evolution experiment was published in the late 19th century by William Dallinger, technological advancements allowed researchers to employ fully-controlled experiments for strain engineering to achieve desired traits [5]. Challenges in the experimental design of evolutionary studies, such as maintenance of the generations, controlling the environment, and feasibility to perform data analysis make the use of microorganisms in evolutionary studies more suited, especially *Escherichia coli* and *Saccharomyces cerevisiae* given their extensive characterization [6].

Commonly used approaches of adaptive laboratory evolution (ALE) includes chemostat cultures and serial batch or colony transfers (Figure 2.1). Serial transfer experiments in which the initial population is aliquoted and transferred into a new medium are simpler and cheaper to set up, however the possibility of genetic drift is higher due to random sampling of the population. On the other hand, continuous systems such as chemostat experiments in bioreactors have advantages on maintaining constant growth rates and population sizes, but with increasing experimental cost [7].

In addition to the methodological choice, decision of the selection criteria and the time span for the experiments are critical factors in ALE [5]. Growth rates, survivability in stressful environmental conditions and biomass yields are the most common fitness criteria for a population to be selected. Throughout the selection process, reaching >500 generations may take a few weeks or a few months depending on the organism, therefore, detailed planning is essential [9].

Capturing genomic changes in the dynamic evolution process is the main goal of an ALE study, therefore frequent data collection in longitudinal manner is a must to unveil molecular details. Next-generation technologies empower researchers to catch even single-nucleotide mutations through genome sequencing and also provide

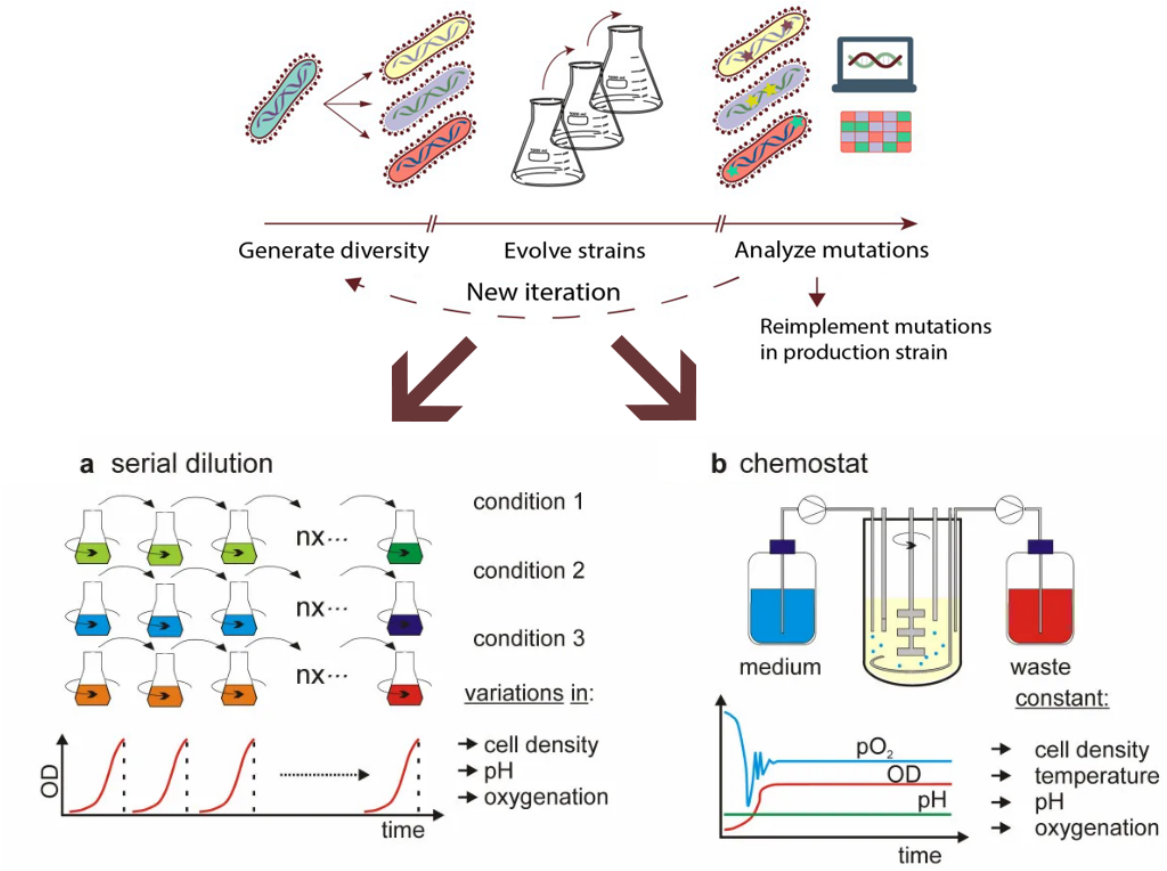

Figure 2.1. ALE methods and workflow. Figure is taken from [5] and [8] and will be redrawn altogether.

an understanding on broader regulatory changes through gene expression profiles [9]. Focus of the studying the differences between evolved and unevolved strains may be on the individual protein level or as a system-level trait. Interpreting the system as a network enables verious ways to investigate causality behind adaptive mechanisms and dynamics of evolution [10, 11].

2.2. 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 [12]. 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 [13]. 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 [14].

A mathematical model of a cell can be approached by two different approaches in either a bottom-up or top-down directionality (Figure 2.2) [15, 16]. 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.

2.2.1. Metabolic Networks

In the context of systems biology, metabolic network reconstructions have become a common interest for the researchers over the past 20 years [17]. 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 [18].

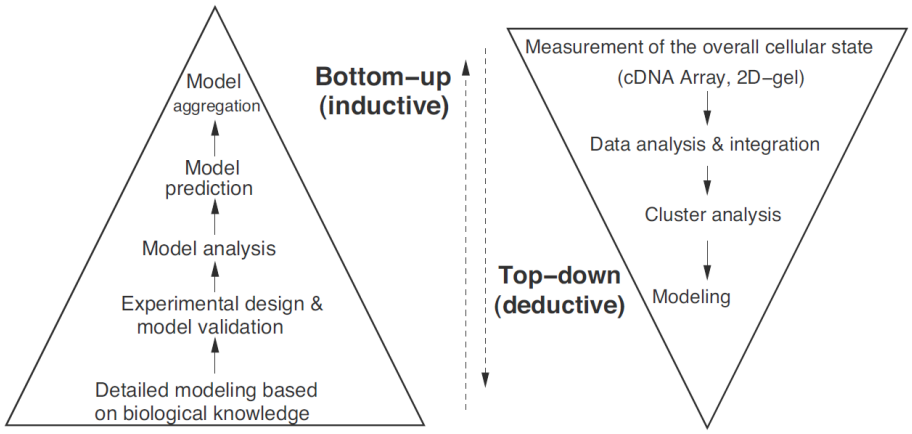

Figure 2.2. Systems biology approaches. Left: Bottom-up approach. Right: Top-down approach. Figure is taken from [14].

Before the improvement of genome sequencing or annotation technologies, initial core metabolic networks were based on the accessible information of biochemical pathways [19] [20]. 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 [21]. 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 [22, 23]. 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 [17, 21] for the reconstruction of biochemical networks described in the Figure 2.3 [24]. 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 [22].

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.

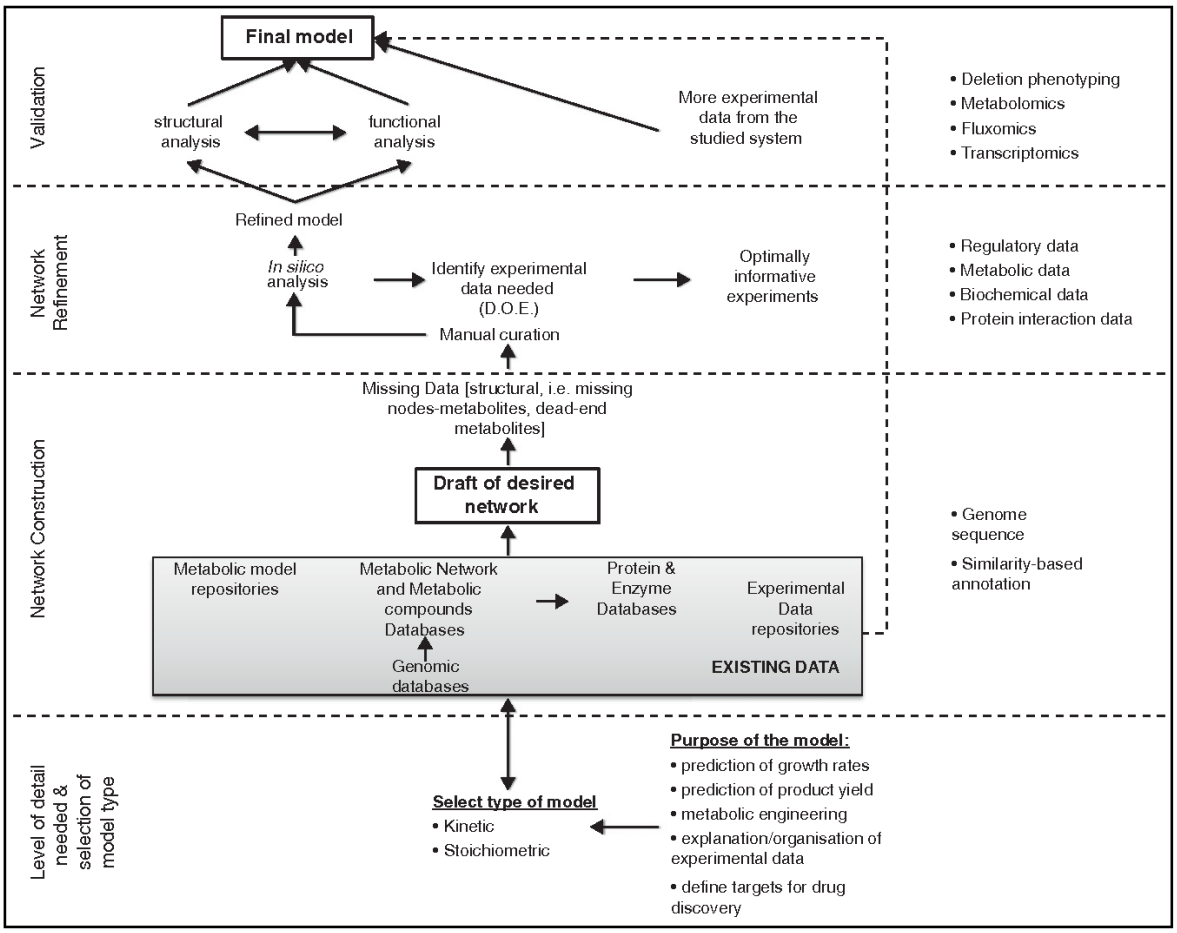

Figure 2.3. Overview of metabolic network reconstruction protocol. Figure is taken from [24].

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 [25]. 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 [26]. 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 [27, 28] 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 [29]. 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 [30]. 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 [31]. Although GSMMs are mainly used in metabolic engineering strategies, other applications both for descriptive and predictive purposes can be found in the literature [32].

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 2.4. Web of Science artice counts on "metabolic model"

2.2.2. Mathematical Representation of Metabolic Networks

Basic applications of metabolic network modelling, although its framework is developed by engineers or mathematicians, is used by biologists with various mathematical backgrounds [33]. In order to speak with one voice, fundamental concepts of mathematical representation of metabolic networks in GSMM reconstructions will be provided in this section.

One can propose a steady-state model for the correlation between the metabolites exist in the network. This model claims that the production and consumption of a metabolites must balance each other [34]. For example, consider the metabolite "A" in the toy network in Figure 2.5. It can be taken into cell with the rate of b1, and can be converted into B or C with the rates of v1 and v2 respectively; at the same time, metabolite C can be converted into A with the rate of v3. Note that enzyme kinetics are not considered and these equations are formed only considering mass balances.

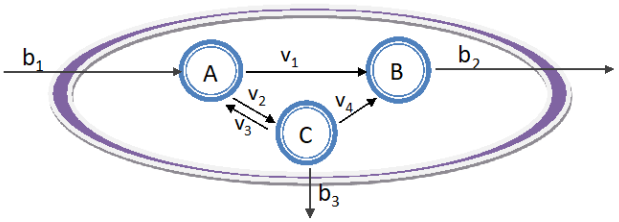

Figure 2.5. Toy network for steady-state assumption (will be redrawn).

If we accept the steady-state assumption, and since we know all the possible reactions passing through metabolite A, we can write following differential equation

$$d(A)/dt = - v1 - v2 + v3 + b1 = 0$$

meaning that production rate of metabolite A is equal to its consumption rate. In other words, there will be no accumulation of metabolite A in the cell. Considering we have 3 metabolites and 7 reactions, this representation of reactions can written as a system of linear equations:

$$\begin{cases} \frac{d[A]}{dt} = -v_1 - v_2 + v_3 + b_1 \\ \frac{d[B]}{dt} = v_1 + v_4 - b_2 \\ \frac{d[C]}{dt} = v_2 - v_3 - v_4 - b_3 \end{cases} \Rightarrow \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$S \cdot v = 0$$

In the above equation, a matrix "S" appears. S matrix consists of stoichiometric coefficients of the reactions. A single column of S matrix, representing a single reaction, provides information about the connections between the metabolites participating in that reaction; and a single row of S matrix, representing a metabolite, provides information about the connection of all the reactions in which that metabolite participates.

Next to S matrix a vector "v" appears. Vector "v" is called the flux vector, and it contains the change in concentration of metabolites. In other words, flux vector represents the flow rates or "fluxes" of each metabolite over time.

The solution space of S matrix, in the steady-state assumption, is generated by the null space base vectors of S. Studying on the null space vectors, we can define dead-end reactions (zero rows which mean reactions cannot carry flux), enzyme subsets (rows of scalar multiplies of each other, possibly chain reactions) and independent components (diagonal block structures in the null space, reactions that are independed from the network). Due to the large number of reactions in biological networks, the system is underdetermined and have multiple solutions in a convex flux cone space, referring to multiple steady-states of the cell.

2.2.3. Constraint-Based Modelling

As explained in the previous section, solution of a metabolic network is not unique, meaning that the model can be found in multiple states of flux distributions. In order to obtain more biologically relevant solutions, v vector can be calculated within a set of constraints [35]. The rationale behind constraining the flux vector is that biological

systems must obey general principles in nature, such as basic rules of chemistry and the laws of thermodynamics.

In terms of physico-chemical properties of a biological system, mass and charge balance must be conserved. These conservation laws are given into mathematical problem as "hard" constraints since these rules are inviolable [36]. Reaction reversibility is a hard constraint, determined by the laws of thermodynamics. Some reactions are irreversible by nature under certain conditions. These reactions are mathematically bounded in the flux vector, so they can have only positive (vi ≥ 0, forward directionality) or negative (vi ≤ 0, backwards directionality) values. Depending on the methodology, by reversing the equations of backwards-directed reactions, and splitting the reversible reactions into two forward and reverse reactions, one can obtain a vector v where each element is greater than or equal to 0. Bunun mathematically favored olmasinin bir sebebi vardi? Stephanopoulos kontrol et.

Since the experiments are carried out in a defined medium, environmental constraints must be added in silico simulations as well. These constraints include nutrient availability (such as carbon, nitrogen or oxygen sources), the pH value, temperature, or any other experiment-specific conditions. Additionally, spatial constraints can be given to the system to limit substrate and enzyme availability for a specific compartment, especially for the metabolites which are transferred between compartments by facilitated diffusion.

More recently, regulatory constraints have started to be used to simulate experimental observations on cellular regulatory mechanisms, such as variable amounts of gene products (transcriptional, translational regulations) and their activities (enzymatic regulations). Integration of omics data, mainly transcriptome data, is used to capture these regulations in silico.

2.3. *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 [37]. It was the first eukaryotic organism whose genome was fully sequenced and annotated [38], and besides its benefits in the industry, it is used as a model system for other eukaryotic cells including humans [39, 40].

2.3.1. 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 [41, 42]. 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 2.6. 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)

2.3.2. Adaptive Evolution Studies on *S. cerevisiae* (Palsson 2015)

2.3.3. 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 [44], and the first commercial platform for oligonucleotide microarray data (Affymetrix) to investigate cellular regulations were reported in 1998 [45]. 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 [46]. More in this section, there will be review on: Genome-scale modeling of yeast: chronology, applications and critical perspectives [47]

2.3.4. Applications of *S. cerevisiae* GSMMs

Literature review on the applications will be added.

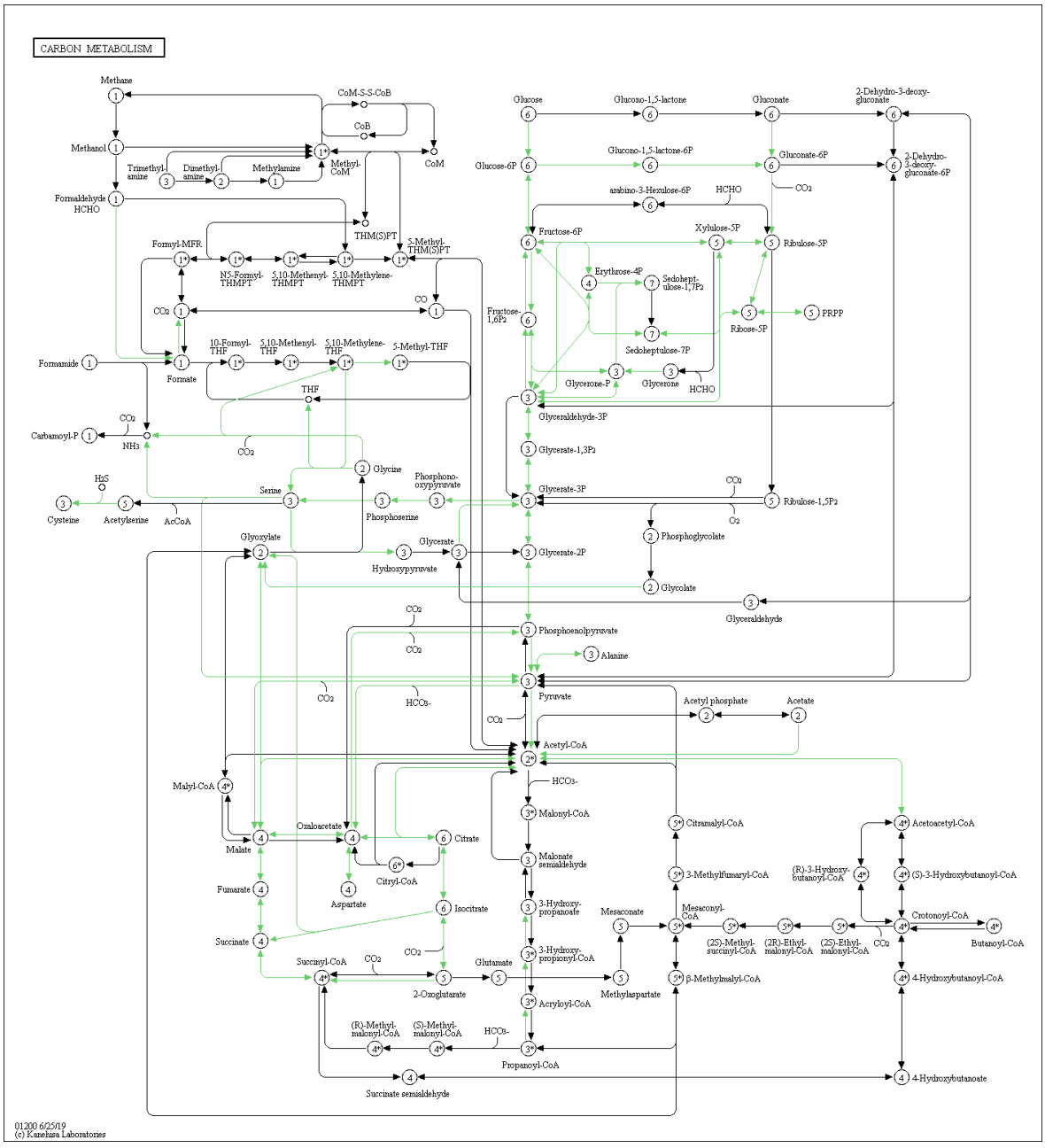

Figure 2.6. Central carbon mechanism of *S. cerevisiae* obtained from KEGG [43].

3. MATERIALS AND METHODS

3.1. 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 [47]. 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 [48]. 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 3.2).

Yeast8 model can be considered as an updated version of Yeast7 [49] with additional corrections based on the annotations available in KEGG and ChEBI, and several gene inclusions from the model iSce926 [50]. 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 3.1.

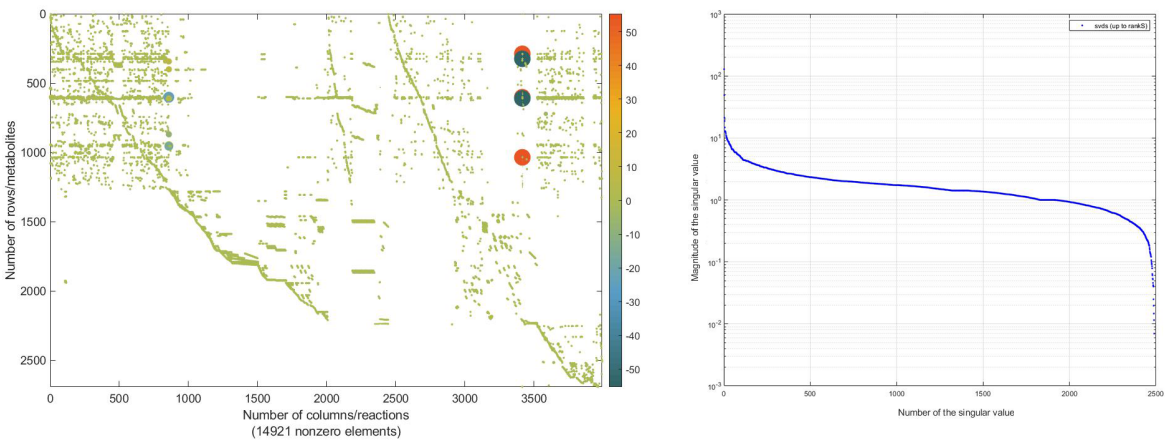

Figure 3.1. Coefficients and singular values of the stoichiometric matrix of Yeast8

All simulations in chapter chapter 3 are done using Yeast8 v8.3.4 model which is hosted in Github (https://github.com/SysBioChalmers/yeast-GEM).

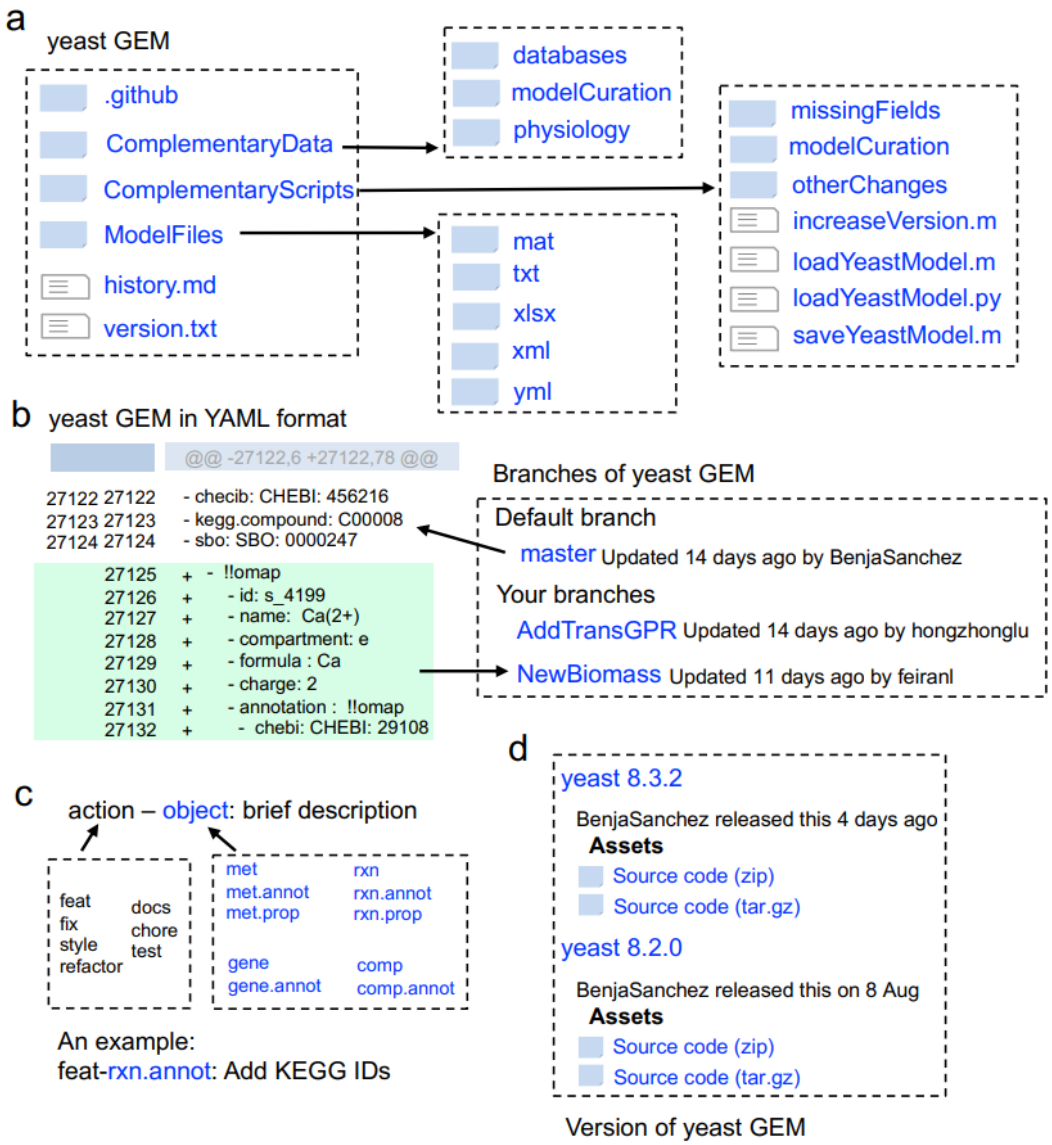

Figure 3.2. Repository of yeast GEM on GitHub. Figure is taken from [48]. will be redrawned in thesis

3.2. 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:

$$S_{m \times n} \cdot v = 0 \tag{3.1}$$

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:

$$\max_v \quad c^T \cdot v \tag{3.2}$$

$$\text{subject to} \quad S_{m \times n} \cdot v = 0 \tag{3.3}$$

$$v_{lb} \leq v \leq v_{ub} \tag{3.4}$$

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 **??**).

3.3. Phenotype Phase Plane Construction

As mentioned in the FBA section, 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 [51]. 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.

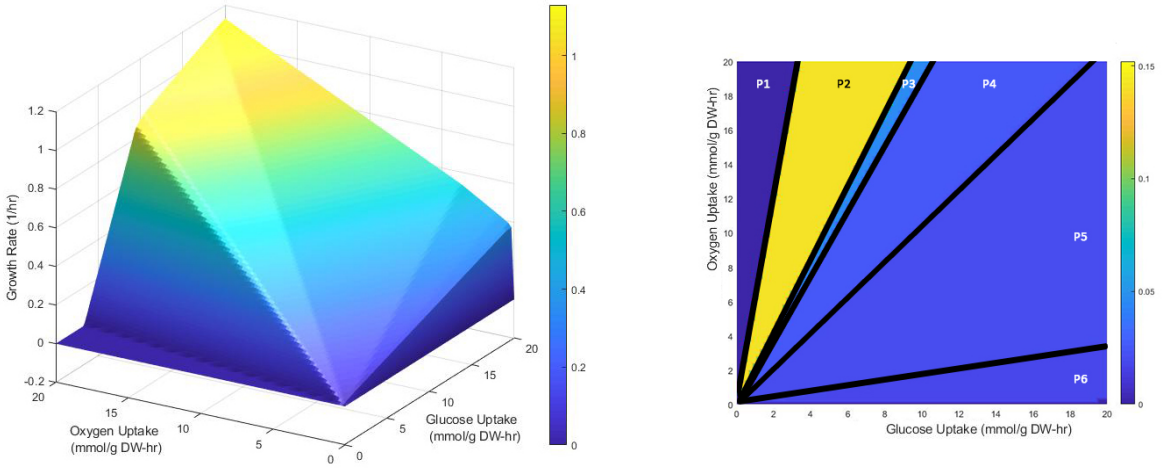

Figure 3.3. Phenotype Phase Plane of Yeast8, corresponding to glucose and oxygen availabilities on the left. Shadow prices of glucose on the right.

3.4. 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 [52], to find alternative optimum states [53], to check flux distributions when growth is not at optimum level [54], and it has many other applications [55].

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

$$\max_v/\min_v \quad v_i \tag{3.5}$$

$$\text{subject to} \quad S_{m\times n}\cdot v = 0 \tag{3.6}$$

$$w^T\cdot v \geq \gamma\cdot Z_0 \tag{3.7}$$

$$v_{lb}\leq v\leq v_{ub} \tag{3.8}$$

where w is the objective function equals to c in the problem 3.3, $Z_0 = w^T\cdot v_0$ describes an optimal solution to the problem 3.3, 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 $0 \leq \gamma < 1$).

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

3.5. Random Sampling of Solution Space

Constraints applied to a model define a solution space, a convex polytope, where every flux distribution is accessible. Random sampling of the solution space is an unbiased tool to explore metabolic models. Mainly, Markov Chain Monte Carlo methods are used to sample this space using algorithms such as (Artificially Centered) Hit-and-Run (HRB) [56, 57] algorithm, and this method has proven to be helpful in the analysis of genome-scale metabolic models [58]. Briefly, the random sampling method collects points that are uniformly distributed in the solution space and calculates the most probable flux value for each reaction.

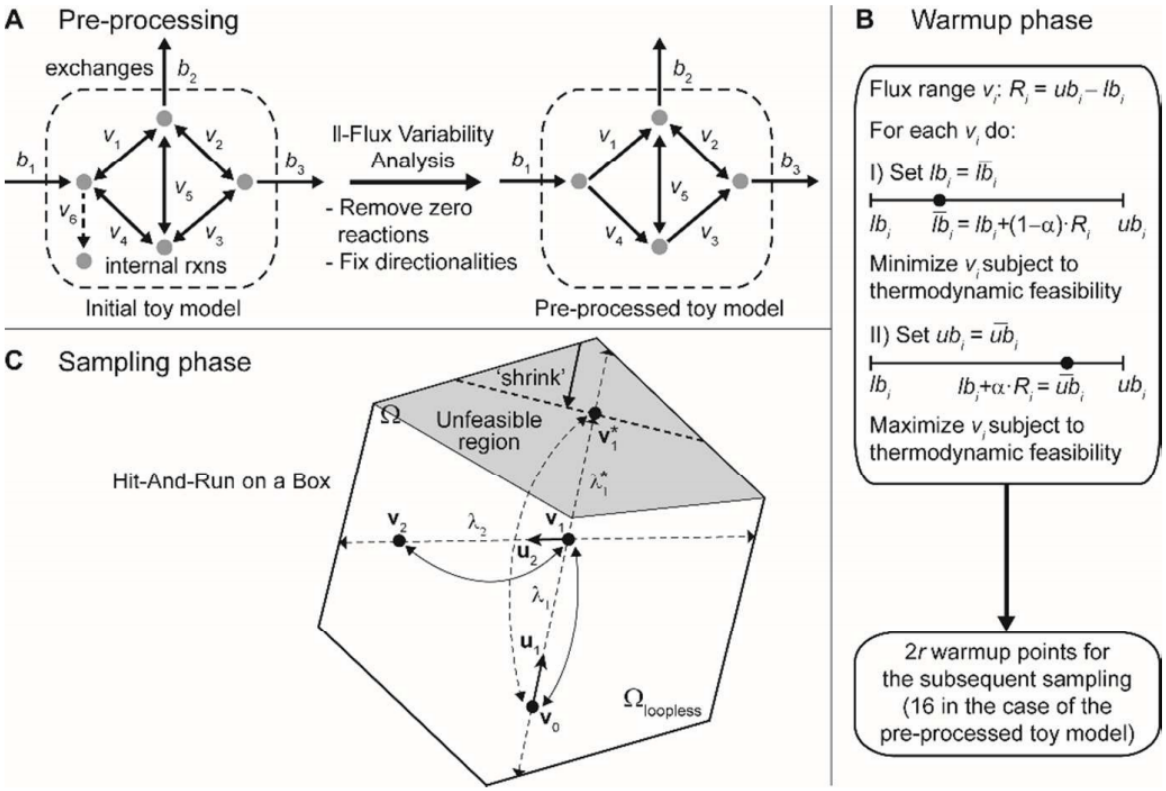

Figure 3.4. Workflow of the ll-ACHRB sampling on a toy model. A) Pre-processing phase, application of loopless-FBA to remove blocked reactions and constraining the directionalities of others. B) Warmup phase, modifying the reaction bounds to more interior space. C) Sampling phase with HRB algorithm. Figure is taken from [57]

Since the computational burden of loopless sampling is high, generated random points in the solution space of Yeast8 includes thermodynamically unfeasble states. Maximum glucose uptake rate was constrained to 1 mmol $gDWh^{-1}$ and total of 5000 points are generated with maximum of 120 secondse alloted for the sampling.

3.6. Expression Data Analysis

GSE numbers, experimental conditions, data normalization, gene enrichment analysis, comparison with essential genes in the model...

3.7. Integration of Expression Data Into Model

Used method and its mathematical explaination on flux bounds...

REFERENCES

1. Futuyma, D. J. and T. R. Meagher, "Evolution, science and society: evolutionary biology and the national research agenda.", *California Journal of Science Education*, Vol. 1, No. 2, pp. 19–32, 2001.

2. Hird, S. M., "Evolutionary biology needs wild microbiomes", *Frontiers in microbiology*, Vol. 8, p. 725, 2017.

3. Nielsen, J. and J. D. Keasling, "Engineering cellular metabolism", *Cell*, Vol. 164, No. 6, pp. 1185–1197, 2016.

4. Garland, T. and M. R. Rose, *Experimental evolution: concepts, methods, and applications of selection experiments*, University of California Press Berkeley, CA, 2009.

5. Dragosits, M. and D. Mattanovich, "Adaptive laboratory evolution–principles and applications for biotechnology", *Microbial cell factories*, Vol. 12, No. 1, p. 64, 2013.

6. McDonald, M. J., "Microbial Experimental Evolution–a proving ground for evolutionary theory and a tool for discovery", *EMBO reports*, Vol. 20, No. 8, 2019.

7. Winkler, J., L. H. Reyes and K. C. Kao, "Adaptive laboratory evolution for strain engineering", *Systems Metabolic Engineering*, pp. 211–222, Springer, 2013.

8. Shepelin, D., A. S. L. Hansen, R. Lennen, H. Luo and M. J. Herrgård, "Selecting the best: evolutionary engineering of chemical production in microbes", *Genes*, Vol. 9, No. 5, p. 249, 2018.

9. Conrad, T. M., N. E. Lewis and B. Ø. Palsson, "Microbial laboratory evolution in the era of genome-scale science", *Molecular systems biology*, Vol. 7, No. 1, 2011.

10. Soyer, O. S. and M. A. O'Malley, "Evolutionary systems biology: what it is and why it matters", *BioEssays*, Vol. 35, No. 8, pp. 696–705, 2013.

11. Long, C. P. and M. R. Antoniewicz, "How adaptive evolution reshapes metabolism to improve fitness: recent advances and future outlook", *Current opinion in chemical engineering*, Vol. 22, pp. 209–215, 2018.

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

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

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

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

16. 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.

17. 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.

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

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

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

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

32. 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.

33. 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.

34. Reimers, A.-M. and A. C. Reimers, "The steady-state assumption in oscillating and growing systems", *Journal of theoretical biology*, Vol. 406, pp. 176–186, 2016.

35. 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.

36. Price, N. D., J. L. Reed and B. Ø. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints", *Nature Reviews Microbiology*, Vol. 2, No. 11, p. 886, 2004.

37. 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.

38. 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.

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

40. Botstein, D., S. A. Chervitz and M. Cherry, "Yeast as a model organism", *Science*, Vol. 277, No. 5330, pp. 1259–1260, 1997.

41. 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.

42. 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.

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

44. 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.

45. 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.

46. 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.

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

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

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

56. Kiatsupaibul, S., R. L. Smith and Z. B. Zabinsky, "An analysis of a variation of hit-and-run for uniform sampling from general regions", *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, Vol. 21, No. 3, p. 16, 2011.

57. Saa, P. A. and L. K. Nielsen, "ll-ACHRB: a scalable algorithm for sampling the feasible solution space of metabolic networks", *Bioinformatics*, Vol. 32, No. 15, pp. 2330–2337, 2016.

58. Schellenberger, J. and B. Ø. Palsson, "Use of randomized sampling for analysis of metabolic networks", *Journal of biological chemistry*, Vol. 284, No. 9, pp. 5457–5461, 2009.

APPENDIX A: APPLICATION

The appendices start here. After references section.