# THESIS TITLE

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

# 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.

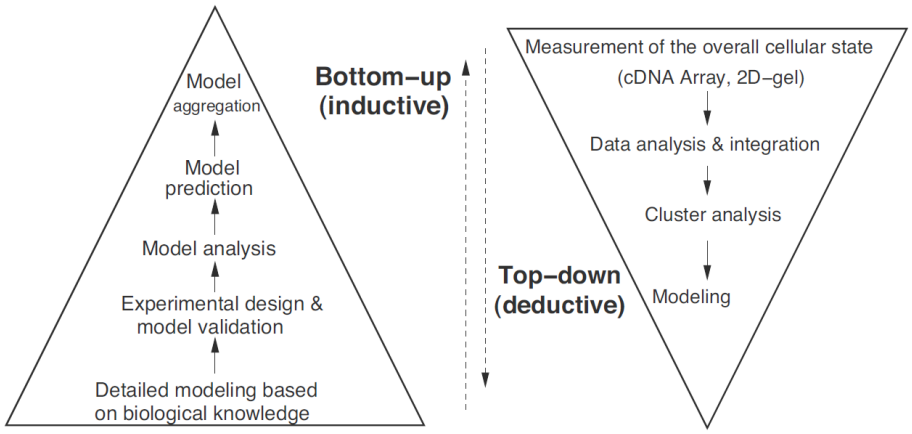


Figure 1.1. Systems biology approaches. Left: Bottom-up approach. Right: Top-down approach. Figure is taken from [3].

### 1.1.1. Metabolic Networks

In the context of systems biology, metabolic network reconstructions have become a common denominator over the past 10 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].

A generally applicable protocol is defined by the Palsson group [6, 8] for the reconstruction of biochemical networks described in the Figure 1.2 [9]). Briefly, genomic data for the biochemical reactions of an organism are identified from the databases, such as NCBI, DDBJ and EMBL-EBI. 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, it is translated into a mathematical model and validated by the experimental data.

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

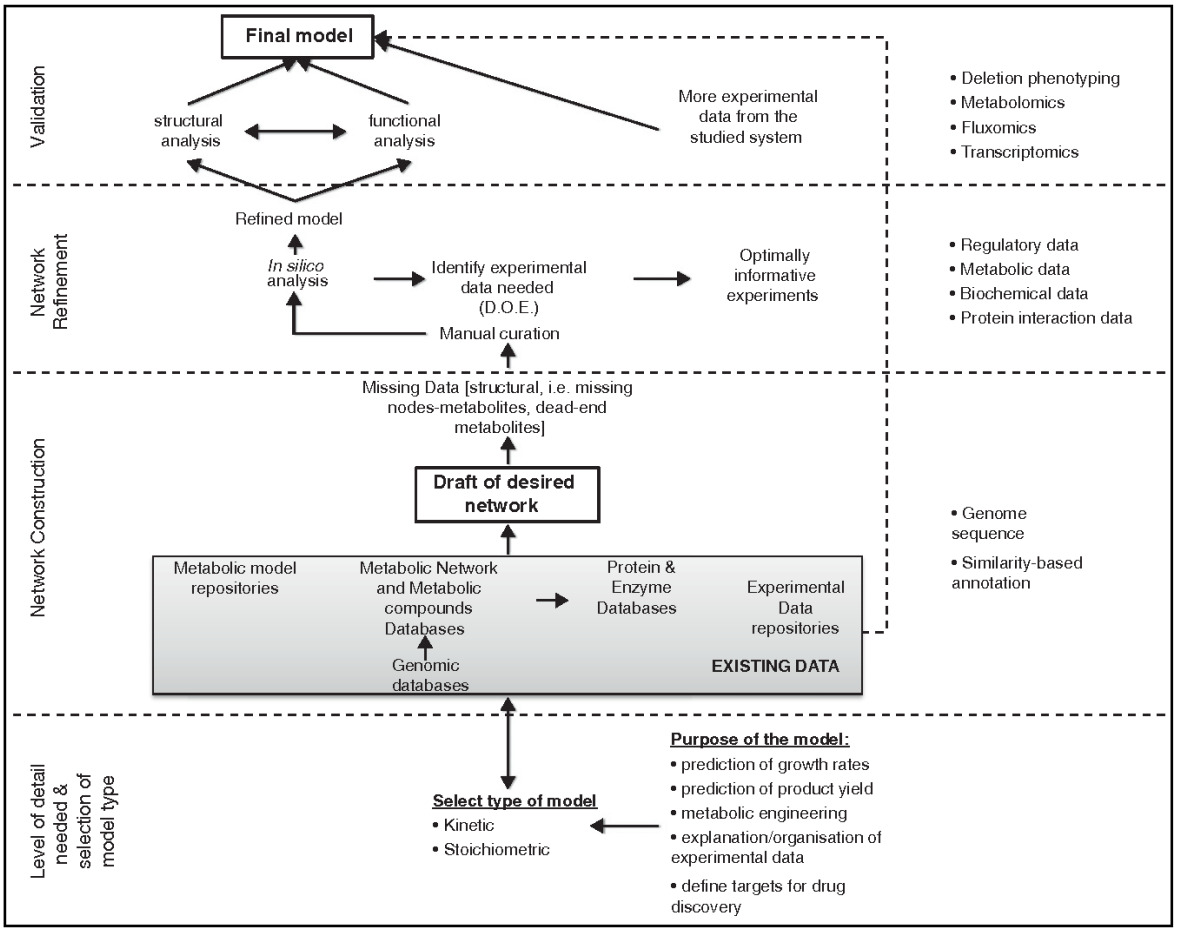


Figure 1.2. Overview of metabolic network reconstruction protocol. Figure is taken from [9].

proteins allowing to investigate the relationships between genotype and phenotype is achieved [10].

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 [11, 12] Because of the lack of kinetic parameters, structural metabolic modeling has been widely used for analyzing cellular metabolism at a steady-state assumption.

### 1.1.2. Mathematical Representation of Metabolic Networks

There will be a review on "Mathematical Framework Behind the Reconstruction and Analysis of Genome Scale Metabolic Models" [13])

In a stoichiometric matrix, each column is a reaction in the network and each of the rows is associated with a component or metabolite expressed by its stoichiometric coefficient, being negative if it is a reactant from the perspective of the reaction...

### 1.1.3. Flux Balance Analysis

There will be a review on "What is flux balance analysis?" [7]

## 1.2. *Saccharomyces cerevisiae*

### 1.2.1. Industrial importance of *S. cerevisiae*

### 1.2.2. Metabolic Models of *S. cerevisiae*

### 1.2.3. Applications of *S. cerevisiae* GSMMs

## 1.3. Significance of Thesis

# 2. MATERIALS AND METHODS

## 2.1. Experimental Data Preparation

### 2.1.1. Data Acquisition

Extracellular metabolomics data is obtained from Cakar's Lab [14]. Missing sentence: What they did was... Briefly, ethyl methane sulfonate (EMS) mutagenesis was performed 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 2.1 and Table 2.2.

### 2.1.2. Determination of Rates

As the slope of the curve of $\ln OD_{600}$ as a function of time gives the growth rates, natural logarithm of $OD_{600}$ values were calculated to obtain specific growth rates by using the equation 2.1.

$$\mu = \frac{\Delta \ln OD_{600}}{\Delta t} \tag{2.1}$$

In order to determine uptake and secretion rates of the metabolites, the steady-state assumption is applied in 3 hours intervals as they are the shortest measured intervals. 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).

Table 2.1. 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 |

Table 2.2. 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 |

Measurement of the cell dry weight at the 3rd hour was crucial for the steady-state assumption, however this data was missing. Curve trend of the OD600 values is used as a guide to estimate cell dry weight (Figure 2.1) Missing method: Estimation approach, similar to regression/curve fitting etc.... Calculated flux values can be found in the Table 2.3.

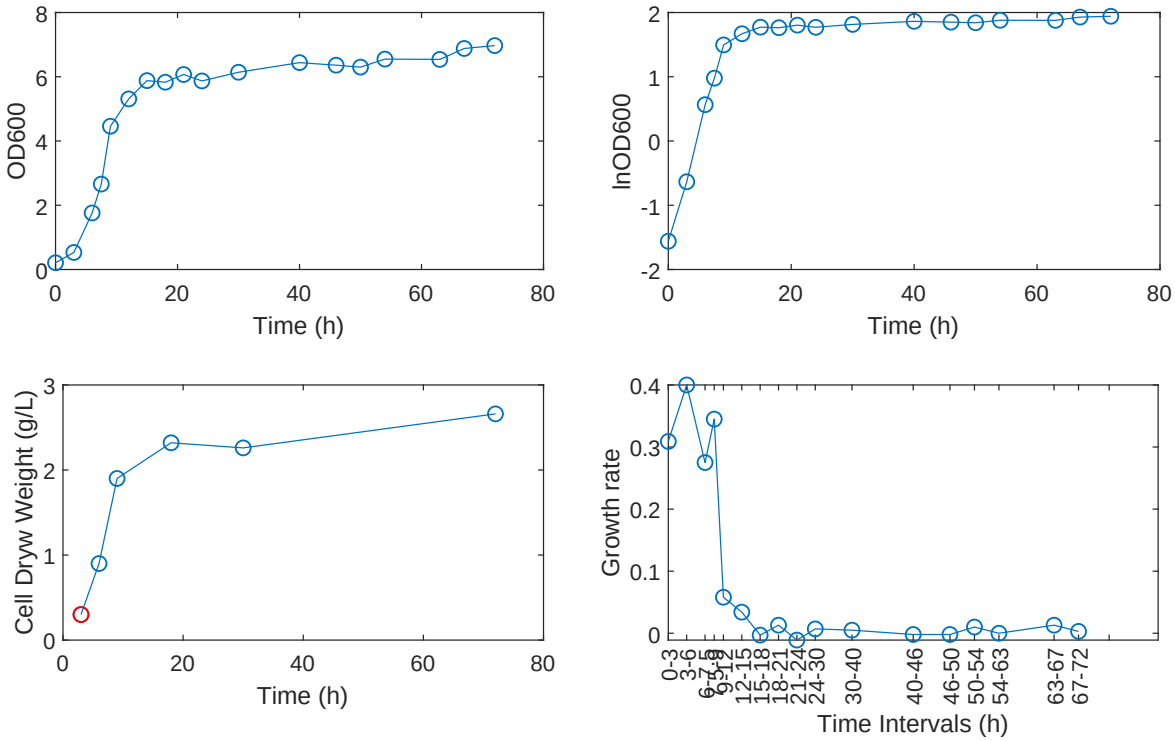


Figure 2.1. $OD_{600}$, $lnOD_{600}$, cell dry weights and growth rates graphs. Estimated missing cell dry weight data is shown in red color.

## 2.2. Model Selection

iAN50, a stoichiometric model of intermediary metabolism including glycolysis, the pentose phosphate pathway, anaerobic excretion, citric acid cycle (TCA cycle), oxidative phosphorylation, and uptake pathways for galactose, ethanol and acetate is used to simulate [15].

Table 2.3. 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 |

## 2.3. Flux Balance Analysis

## 2.4. Simulation of Batch Conditions

# 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. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

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