Dynamic modeling and control of the main metabolism in Lactic Acid Bacteria

Bhabuk Koirala

Dissertation submitted to obtain Masters Degree in Biotechnology

Examination Committee

Chairperson: Prof. Dr. Luís Joaquim Pina da Fonseca
Supervisors: Prof. Dr. Isabel Maria de Sá Correia Leite de Almeida
Dr. Rafael Sousa Costa
Members of the Committee: Prof. Dr. Nuno Gonçalo Pereira Mira
Prof. Dr. Susana de Almeida Mendes Vinga Martins

July 2013
Acknowledgments

I would like to thank Rafael Costa, who provided me the technical basis, guidance, supervision, always pointed right directions, was always patient with problems encountered and answered all of my questions tirelessly and Prof. Susana Vinga, who provided me the chance to work in project PNEUMOSYS at INESC-ID in KDBio group, and was providing me with me all possible support, motivation, goals to be set, comments, guidance and supervision. This dissertation is performed under the framework of project PNEUMOSYS (PTDC/SAU-MII/100964/2008). I would like to thank KDBio group in INESC-ID. I would also like to thank Prof. Isabel Sá-Correia and Prof. Juho Rousu, my supervisors for the constant motivation and support throughout the work.

I would like to admit my sincere thanks to the European Union and euSYSBIO program for the scholarship to complete euSYSBIO Masters program, of which this thesis work is part of. I would also like thank my parents, my brother, my class fellows at Aalto University and Instituto Superior Técnico for being a source of continual emotional support.

Thank you Prapti for supporting and motivating me in all bad days and long nights during this masters program and during the course of this thesis in particular. Thank you very much.
Abstract

Lactic acid bacteria (LAB) are widely used in industrial manufacture of fermented foods, such as cheese and buttermilk and regarded as cell factories for production of pharmaceutical and food products. *Lactococcus lactis*, due to its small genome size and simple metabolism, has been considered a model organism for strain design strategies and metabolic engineering. These strain design strategies are applied for production of compounds such as acetoin and 2,3-butanediol. Acetoin is used as additives in food industries and cigarette industries while 2,3-butanediol is extensively being used in manufacture of printing inks, perfumes, plasticizers, foods, and pharmaceuticals. Such strain design strategies have mainly focused on rerouting pyruvate metabolism to produce fermentation end products. These compounds, other than the main product of metabolism, are refereed as secondary metabolites and are often produced in insignificant amounts compared to primary metabolites. The strain design strategies implement the over production of these secondary metabolites compared to the primary metabolite.

Biological network modeling, a fundamental aspect of systems biology, provides a platform to conduct *in silico* experiments with biotechnological and biomedical applications. These models are advantageous in the field of metabolic engineering to design mutant strains with capability of producing biotechnologically relevant products. With a fully detailed kinetic model, time-course simulations, response to different input can be predicted and system controllers can be designed. For *L. lactis*, the dynamic models for the central carbon metabolism have already been constructed. However, these models lack our compound of interest and need to be extended. Here, provided the interaction map of pathway under study and kinetic parameters, a dynamic model that describes the glycolytic pathway in *L. lactis* is constructed using convenience kinetics. This model is now improved and extended by estimating the parameters using *in vivo* Nuclear Magnetic Resonance (NMR) data fitting.

Sensitivity analysis was performed in the reconstructed model for acetoin and butanediol production which suggests that down expressing the enzyme levels for lactate dehydrogenase, phosphofructokinase, pyruvate dehydrogenase causes increased acetoin and 2,3-butanediol production. In addition to these enzyme levels, down expressing enzyme levels for acetoin transportase and alcohol dehydrogenase accounts for enhanced production of
2,3-butanediol. The role of enzymes such as lactate dehydrogenase in over production of acetoin and 2,3-butanediol are also supported by different experimental evidences. With the role of different enzyme levels known for production of specific metabolites, the model can later be used as a tool for metabolic engineering. The constructed model can also be used to predict the phenotype of the bacterium under different environmental and genetic conditions and can be used as a starting point to model other Lactic Acid Bacteria such as *Streptococcus pneumoniae*.

**Keywords**

*Lactococcus lactis*, dynamic modeling, parameter estimation, convenience kinetics, *in vivo* NMR data fitting, sensitivity analysis, optimization.
# Contents

1 Introduction 1
   1.1 Systems Biology .................................................. 2
   1.2 Metabolic Engineering ........................................... 2
   1.3 Lactic Acid Bacteria ............................................. 3
   1.4 Motivation of the work .......................................... 4
   1.5 Thesis Outline .................................................. 5

2 Background 7
   2.1 Biological Networks ............................................. 8
      2.1.1 Signaling networks ........................................ 8
      2.1.2 Gene regulatory networks ................................. 8
      2.1.3 Metabolic Networks ....................................... 8
   2.2 Metabolic models and modeling strategies ................. 9
      2.2.1 Model requirements ....................................... 9
      2.2.2 Stoichiometric models .................................. 10
      2.2.3 Kinetic models ........................................... 11
   2.3 Mechanism based models ....................................... 12
      2.3.1 Michaelis-Menten and similar rate laws ............... 12
         2.3.1.A Convenience kinetics ................................ 13
      2.4 Approximated approaches .................................... 14
      2.5 Dynamic Systems and Simulations ......................... 15
      2.6 Parameter estimation ........................................ 15

3 Methods 17
   3.1 Computational theory, tools and algorithms ............... 18
      3.1.1 Flux Balance Analysis (FBA) ........................... 18
      3.1.2 Implementation ........................................... 19
         3.1.2.A OptFlux ............................................... 19
      3.1.3 cellDesigner ............................................. 19
      3.1.4 COPASI .................................................. 21
      3.1.5 Algorithms for parameter estimation used .......... 22
3.1.5.A Global Optimization ........................................ 22
3.1.5.B Local Optimization ........................................ 24
3.2 Akaike Information Criterion .................................... 24
3.3 Sensitivity Analysis ............................................. 24
  3.3.1 Local parameter sensitivity analysis ....................... 25
  3.3.2 Computational implementation .............................. 25
3.4 Glycolysis in L. lactis within BST framework ............... 26
3.5 Model comparison and extension ............................... 27
  3.5.1 Model comparison ........................................... 27
  3.5.2 Model extension ............................................ 28

4 Results and Discussion ........................................ 29
  4.1 Comparison of stoichiometric and kinetic modeling ........ 30
    4.1.1 Critical Reaction determination ......................... 30
    4.1.2 Maximize ‘biomass’, ‘desired’ and ‘R_ext’ formation .... 30
    4.1.3 Knock-out simulations ..................................... 31
    4.1.4 OptKnock .................................................. 31
    4.1.5 Sensitivity Analysis ....................................... 32
      4.1.5.A Sensitivities of M2_ext ............................... 33
      4.1.5.B Sensitivities of M4_ext ............................... 33
      4.1.5.C Sensitivities of M5_ext ............................... 34
    4.1.6 Sensitivity analysis and Flux Balance Analysis ......... 34
  4.2 Simulation of Glycolysis within BST framework ............. 35
  4.3 Approximated vs. semi-mechanistic kinetics ............... 40
    4.3.1 Model construction and parameter estimation ........... 40
    4.3.2 Model validation ......................................... 41
  4.4 Model extension ............................................. 42
  4.5 Sensitivity analysis in extended model ...................... 46
    4.5.1 Single parameter perturbation ............................ 46
      4.5.1.A Analysis for Acetoin ................................ 46
      4.5.1.B Analysis for Butanediol ............................. 47
    4.5.2 Double parameter perturbation ......................... 48
      4.5.2.A Analysis for Acetoin ................................ 49
      4.5.2.B Analysis for Butanediol ............................. 50

5 Conclusions and Future work .................................. 53

Bibliography .................................................... 57
Appendix A  Rate equations and initial conditions used while translating stoichiometric model to kinetic model  A-1

Appendix B  Network structure, rate equations and parameters for simulation of glycolysis within BST  B-1

Appendix C  Network structure and rate equations for model mimicking glycolysis  C-1

Appendix D  Rate equations and initial conditions used in Approximated vs. Mechanism based modeling approach  D-1

Appendix E  Reactions, rate laws, parameters and species concentration used in extended model  E-1

Appendix F  Calculation of $RD$ values using MATLAB  F-1
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Electron microscopy image of <em>Lactococcus lactis</em></td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Pyruvate branches in <em>Lactococcus lactis</em></td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Network structure to construct stoichiometric model</td>
<td>10</td>
</tr>
<tr>
<td>2.3</td>
<td>Transformation of stoichiometric model to kinetic model</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Snapshot of cellDesigner</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>Snapshot of COPASI</td>
<td>21</td>
</tr>
<tr>
<td>3.3</td>
<td>Graphical illustration of dynamic sensitivity analysis</td>
<td>26</td>
</tr>
<tr>
<td>4.1</td>
<td>Toy network used to compare FBA and sensitivity analysis</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>Sensitivity analysis of toy model for species ‘M2_ext’</td>
<td>33</td>
</tr>
<tr>
<td>4.3</td>
<td>Sensitivity analysis of toy model for species ‘M4_ext’</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>Sensitivity analysis of toy model for species ‘M5_ext’</td>
<td>34</td>
</tr>
<tr>
<td>4.5</td>
<td>Time course dynamics of glycolytic model in <em>L. lactis</em></td>
<td>36</td>
</tr>
<tr>
<td>4.6</td>
<td>Response of simplified glycolytic pathway in <em>L. lactis</em> from figure. C.1</td>
<td>37</td>
</tr>
<tr>
<td>4.7</td>
<td>Response of mimicked glycolytic pathway with one extra species</td>
<td>37</td>
</tr>
<tr>
<td>4.8</td>
<td>Effect of different regulations on PYK</td>
<td>39</td>
</tr>
<tr>
<td>4.9</td>
<td>Fitting of the model with 40 mM glucose impulse at time zero</td>
<td>40</td>
</tr>
<tr>
<td>4.10</td>
<td>Validation of the model with 80 mM glucose impulse at time zero</td>
<td>42</td>
</tr>
<tr>
<td>4.11</td>
<td>Network structure extended and reconstructed</td>
<td>43</td>
</tr>
<tr>
<td>4.12</td>
<td>Fittings of the reconstructed model with 40 mM glucose impulse at time zero</td>
<td>44</td>
</tr>
<tr>
<td>4.13</td>
<td>Validation of the reconstructed model with 80 mM glucose impulse</td>
<td>45</td>
</tr>
<tr>
<td>4.14</td>
<td>Sensitivity analysis of the reconstructed model for acetoin</td>
<td>46</td>
</tr>
<tr>
<td>4.15</td>
<td>Sensitivity analysis of the reconstructed model for butanediol</td>
<td>47</td>
</tr>
<tr>
<td>4.16</td>
<td>Double parameter perturbation for acetoin</td>
<td>49</td>
</tr>
<tr>
<td>4.17</td>
<td>Double parameter perturbation for butanediol (+3%)</td>
<td>50</td>
</tr>
<tr>
<td>B.1</td>
<td>Glycolysis in <em>L. lactis</em></td>
<td>B-2</td>
</tr>
<tr>
<td>C.1</td>
<td>Generic linear feedforward activated pathway</td>
<td>C-2</td>
</tr>
</tbody>
</table>
List of Tables

4.1 Flux distribution in toy network ................................................. 30
4.2 Flux distribution, after knockout, maximizing biomass .................... 31
4.3 Flux distribution, maximizing $R_{ext}$ ........................................ 32
4.4 FBA and sensitivity analysis compared in figure 4.1 ......................... 34
4.5 Parameters changed for FBP and P regulations ............................. 38
4.6 Parameters used in the simulation of ODE’s in equation (D.2) .............. 41
4.7 Parameters used in the simulation of ODE’s in equation (D.1) .............. 41
4.8 RD values for Acetoin with $+3\%$ of single perturbation .................... 48
4.9 RD values for Butanediol with $+3\%$ of single perturbation .................. 48
4.10 Enzyme levels with significant effect for Acetoin and Butanediol ......... 50

A.1 Parameters used to simulate equation (A.1) ................................ A-2

B.1 Parameters used in the simulation of ODEs in equation (B.1) ............. B-4
B.2 Parameters used in the simulation of ODEs in equation (B.1) ............. B-4
C.1 Parameter and species used in the simulation of ODEs in equation (C.1) .. C-2

D.1 Initial concentrations used in the simulation of ODE’s in equation (D.1), (D.2) D-3

E.1 Reaction name and abbreviation ................................................. E-6
E.2 Initial and estimated enzyme levels for system of ODE’s for extended model E-7
E.3 Initial and estimated regulator levels for system of ODE’s for extended model E-7
E.4 Equilibrium constants and hill coefficients ................................... E-8
E.5 Initial concentrations used to simulate extended model ..................... E-8
E.6 Initial and estimated $k_M$ values for the extended model ................... E-9
Abbreviations

LAB  Lactic Acid Bacteria
GMA  Generalized Mass Action
NMR  Nuclear Magnetic Resonance
ODE  Ordinary Differential Equations
PSS  Pseudo Steady State
FBA  Flux Balance Analysis
BST  Biochemical Systems Theory
SBML  Systems Biology Markup Language
DAE  Differential Algebraic Equation
COPASI  COmplex PAthway SImulator
AICc  Akaike Information Criterion
G6P  Glucose 6-phosphate
F6P  Fructose 6-phosphate
FBP  Fructose bi-phosphate
G3P  Glyceraldehyde 3-phosphate
PGA/P3GA  3-phosphoglycerate
BPG  bi-phosphoglycerate
PEP  Phosphoenolpyruvate
PYR  Pyruvate
LAC  Lactate
Co A  Coenzyme-A
AC. Co A  Acetyl Coenzyme-A
Etoh  Ethanol
P/Pi  Orthophosphate
ATP  Adenosine-5′-triphosphate
ADP  Adenosine diphosphate
NAD  Nicotinamide adenine dinucleotide
NADH  Nicotinamide adenine dinucleotide hydride
PTS:glucose  Phosphotransferase: glucose
PGI  Phosphoglucoisomerase
PK  Phosphofructokinase
FBA  Fructose-bisphosphate aldolase
GAPDH  Glyceraldehyde 3-phosphate dehydrogenase
ENO  Enolase
PYK/PK  Pyruvate kinase
LDH  L-lactate dehydrogenase
<table>
<thead>
<tr>
<th>Acetolactate synthase; acetolactate decarboxylase</th>
<th>Acetoin Transportase</th>
<th>Butanedioi dehydrogenase</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>AT</td>
<td>AB</td>
</tr>
<tr>
<td>Acetate kinase</td>
<td>Pyruvate dehydrogenase</td>
<td>Alcot aldehyde dehydrogenase; alcohol dehydrogenase</td>
</tr>
<tr>
<td>MPD Manna nol phosphate dehydrogenase</td>
<td>MP</td>
<td>MT Macki acid phosphate dehydrogenase</td>
</tr>
<tr>
<td>MPD Manna nol phosphate dehydroginase</td>
<td>MP</td>
<td>MT Macki acid phosphate dehydrogenase</td>
</tr>
<tr>
<td>pts: Manna nol phosphotransferase: Manna nol</td>
<td>PT</td>
<td>Phosphate transportase</td>
</tr>
<tr>
<td>MPD Manna nol phosphate dehydrogenase</td>
<td>MP</td>
<td>MT Macki acid phosphate dehydrogenase</td>
</tr>
<tr>
<td>pts: Manna nol phosphotransferase: Manna nol</td>
<td>PT</td>
<td>Phosphate transportase</td>
</tr>
</tbody>
</table>
# Introduction

Contents

1.1 Systems Biology ........................................... 2
1.2 Metabolic Engineering .................................... 2
1.3 Lactic Acid Bacteria .................................... 3
1.4 Motivation of the work .................................. 4
1.5 Thesis Outline .......................................... 5
Chapter 1. Introduction

1.1 Systems Biology

Systems Biology is a recent field of study, where it focuses on complex interactions that occurs within a system such as cell. Systems Biology has set a new paradigm as it looks these systems as a whole unit rather than breaking them into parts as done by classical biology. Living systems are dynamic and complex, and their behavior may be hard to predict from the properties of individual parts. Cell is composed of number of interacting components such as genes, proteins and metabolites which form complex biological networks.

The behaviour of cell depends not only on the topology but also in the dynamics of the network. To study them, quantitative measurements of groups of interacting components are obtained via high throughput technologies such as genomics, transcriptomics, proteomics and metabolomics [1], [2]. These high throughput measurement technologies allows the quantification of the molecular species and construction of a interaction map. With a interaction map, data regarding the interacting species and the interaction, an in silico model to drive experiments can be constructed. Systems biology has been growing both in academia and industry. The in silico approach provides a fast and inexpensive way to run experiments and test new hypothesis. Though it does not replaces the laboratory experiments, these are particularly used to guide the experimental design saving resources and time [3]. Computational models can be used to understand biological systems via simulations that predicts the cell behavior and redirect the behavior for a required manipulation, typically in the fields such as metabolic engineering. These models are also used as a framework for studying disease mechanism and drug discovery [4], [5].

1.2 Metabolic Engineering

The utilization of micro-organisms began centuries ago, where these cell factories were being used for the production of alcoholic beverages. Currently, metabolic engineering has a widespread application in synthesis of industrially relevant compounds. Industrial biotechnology copes with cost competitive, environmental friendly, and sustainable alternatives to existing chemical-based production process [6], [7]. These industrially relevant compounds may refer to pharmaceutical drugs, vitamins, amino acids etc. [8], [9], [10]. Biotechnological products to be cost competitive, maximum conversion of substrate to its product is necessary. Metabolic engineering provides a platform to modify the cellular metabolism and optimization of desired metabolites [11]. Traditionally this was being done using mutagenesis [12] or directed evolution [13]. These strategies have proven to be useful for penicillin production [14], however does not accounts for genetic changes that occurs in cell. Metabolic engineering uses different strain design strategies to redirect the metabolic flux towards the desired output. Metabolic engineering elucidates which gene to be over-expressed or knocked-out for high enzyme or altered enzyme activity for expression of
targeted metabolites.

To design a microbial strain for maximum production of targeted metabolite, a metabolic model capable of predicting the metabolic phenotype is necessary. These mathematical models are based upon Generalized Mass Action kinetics (GMA), Biochemical Systems Theory (BST) and convenience kinetics which are used to analyze the parameters controlling the metabolic flux. However, the difficulty of obtaining kinetic data have decreased the use, and popularity of these models favoring genome-scale stoichiometric models [15]. These stoichiometric models allows simulation of the metabolic pathway under steady-state conditions, using constraint-based methods like Flux Balance Analysis (FBA) [16].

1.3 Lactic Acid Bacteria

Lactic Acid Bacteria (LAB) are gram-positive, non-sporeforming cocci, coccobacilli or rods. They ferment glucose primarily to lactic acid or lactate, CO₂ and ethanol. All LAB grow anaerobically, but unlike most anaerobes, they can also grow in the presence of O₂ as "aerotolerant anaerobes". Although many genera of bacteria produce lactic acid as a primary or secondary end-product of fermentation, the term Lactic Acid Bacteria is conventionally reserved for genera in the order Lactobacillales, which includes Lactobacillus, Leuconostoc, Pediococcus, Lactococcus and Streptococcus, in addition to Carnobacterium, Enterococcus, Oenococcus, Tetragenococcus, Vagococcus, and Weisella. Complete genomic sequence of Bacillus subtilis, Lactobacillus plantarum and Lactococcus lactis have been determined while that of others LAB are partially analyzed [17], [18].

LAB are restricted to environments in which sugars are present since they obtain energy only via sugar metabolism. Most are free-living or live in beneficial or harmless associations with animals, although some are opportunistic pathogens. They are found in milk and milk products and in decaying plant materials. They are present as normal flora of humans in the oral cavity, the intestinal tract and the vagina, where they play a beneficial role. A few LAB are pathogenic for animals, mainly some members of the genus Streptococcus. In humans, Streptococcus pyogenes is a major cause of disease (strep throat, pneumonia, and other pyogenic infections, scarlet fever and other toxemias), Streptococcus pneumoniae causes lobar pneumonia, otitis media and meningitis. [17], [19].

Lactic acid bacteria are among the most important groups of microorganisms used in food fermentation. Taste and texture of food are contributed by LAB and also food spoilage is prevented by LAB by producing growth inhibiting substance and large amount of lactic acid. As agents of fermentation, LAB are involved in making yogurt, cheese, cultured butter, sour cream, sausage, cucumber pickles, olives and sauerkraut, but some species may
spoil beer, wine and processed meats. Recent interest focuses on production of bio-refinery products, such as stereoisomers of lactic acid, ethanol and other acids. Several LAB are also used in probiotic products that provides health benefit by production of specific metabolites [17], [18].

*Lactococcus lactis*

*Lactococcus* is a genus of LAB with five major species formerly classified as Group N streptococci. The type species for the genus is *Lactococcus lactis*, which has two subspecies, *lactis* and *cremoris*. Lactococci differ from other lactic acid bacteria by their pH, salt and temperature tolerances for growth. *Lactococcus lactis* is critical for manufacturing cheeses such as cheddar, cottage, cream, camembert, roquefort and brie, as well as other dairy products like cultured butter, buttermilk, sour cream and kefir. The bacterium can be used in single strain starter cultures, or in mixed strain cultures with other lactic acid bacteria such as *Lactobacillus* and *Streptococcus*. Under normal growth conditions, *L. lactis* strictly follows homolactic fermentation where sugars get completely converted to L-isomer of lactate [17], [19], [18] [20].

![Figure 1.1: Electron microscopy image of Lactococcus lactis figure form](image)

1.4 Motivation of the work

Mathematical modeling and strain design strategies could be applied in *L. lactis* for the production of metabolites such as acetoin and 2,3-butanediol. Acetoin is used as additives in food and cigarette industries while 2,3-butanediol is extensively being used in manufacture of printing inks, perfumes, plasticizers and pharmaceuticals [22]. These compounds are often produced in insignificant amounts compared to the main product of glucose metabolism. The strain design strategies implement the over production of these secondary metabolites or metabolites of interest compared to the primary metabolites.
With the advances in system biology, dynamic models of *L. lactis* that can describe the phenotype of the organism has been constructed. Hoefnagel et al, 2002 [23] devised a kinetic model describing the pyruvate branches in *L. lactis*. A genome scale model of *L. lactis* was given by Oliveira et al, 2005 [24]. Regulation of glycolysis in *L. lactis* was explained by Voit et al, 2006 [25]. The central carbon metabolism in *L. lactis* have been devised by Levering et al, 2012 [26]. However, these models lack the metabolites of our interest or are incomplete glycolytic models describing glucose metabolism till lactate production only. The glycolytic enzymes that possess a confound effect on acetoin and butanediol production in *L. lactis* have not yet been elucidated, although there are experimental evidence of increased production of acetoin caused by single gene knockouts [27].

The main goal of this thesis is to reconstruct a metabolic model of *L. lactis* which constitutes our metabolites of interest, evaluate the predictions of the model by comparing it with the experimental data available and know the reactions that have a confound effect on production of our metabolite of interest (acetoin and butanediol). This information is regarded valuable in terms of cellular biochemistry or metabolism, which is not readily available to the biologists and thus the model can later guide the design of experiment to produce mutant strains of *L. lactis*.

Modeling begins with a network structure of the system under study and the required formulations which are the reaction rates or stoichiometry, kinetic parameters and initial conditions used within it. These formulations when put together, make up a model with help of tools such as cellDesigner [28] and COPASI [29] described in chapter 3. A metabolic model of anaerobic glycolysis in *L. lactis* is reconstructed with known modeling strategies and network structure. The kinetic parameters are obtained by fitting the model to data. After checking the consistency of its prediction with experimental data, a script to perform sensitivity analysis of the model for acetoin and butanediol production was developed in MATLAB®.

### 1.5 Thesis Outline

Chapter 1 (current chapter) gives a brief introduction to the field of Systems Biology, its relation with metabolic engineering, *Lactococcus lactis* and motivation of this work.

Chapter 2 gives a brief overview on metabolic networks and modeling strategies used.

Chapter 3 presents the methodology followed for the reconstruction of metabolic model, parameter estimation in kinetic models, optimization strategy in stoichiometric models and analysis of kinetic models. Section 4.3 presents the work that was presented as a poster.
Chapter 1. Introduction

on the second edition of Bioinformatics Open Days event which was held at University of Minho, Gualtar campus (Braga, Portugal) on 14\textsuperscript{th} and 15\textsuperscript{th} March 2013.

Chapter\textsuperscript{4} is the results and discussion derived from this work.

Chapter\textsuperscript{5} is the conclusions and discusses future work that can be done and ends the thesis.
2 Background

Contents

2.1 Biological Networks ........................................... 8
2.2 Metabolic models and modeling strategies ...................... 9
2.3 Mechanism based models ...................................... 12
2.4 Approximated approaches .................................. 14
2.5 Dynamic Systems and Simulations .......................... 15
2.6 Parameter estimation ....................................... 15
Understanding the mechanism of complex system or cell, has proven to be advantageous in several research areas such as drug development and industrially relevant compounds production. Using mathematical models of cellular metabolism, systems biology has provided a benchmark to optimize a model for desired phenotype and predict manipulations such as gene knockouts, over expressions and under expressions. The interconnection between different cellular processes such as genetic regulation and metabolism reflects the holistic approach in systems biology in replacement of traditional methods. Though the cellular processes are studied individually, the behavior of the cell is given by the network-level [29], [30].

2.1 Biological Networks

Cell houses a variety of components that interact with each other in different number of ways. These networks can be classified according to their biological functions, that are signaling, gene regulatory and metabolic network [31].

2.1.1 Signaling networks

The representation of signal transduction between cells where a cell receives and responds to a stimuli by other cell is known as signaling network. Apoptosis, cell division are well known mechanisms effected by cell signaling [32].

2.1.2 Gene regulatory networks

Gene regulation controls the expression of different genes. In a cell, different functions are active in different stages of life for instance, when adapting in a new environment. The gene is transcribed into mRNA, followed by translation to proteins. The transcription is controlled by different inhibitors and activators which are also encoded by some other genes, subject to regulation forming a complex network of regulation, known as gene regulatory networks [33].

2.1.3 Metabolic Networks

Metabolism involves a mechanism where biochemical enzyme-catalyzed reactions produces different metabolites thus fulfilling the nutrients and growth requirement of a cell. Metabolic network are further divided into pathways where different pathways are connected to each other via some intermediate metabolite. The enzymes that catalyse these reactions are encoded by genes which are regulated in cell that directs appropriate pathway for utilization of available substrate which may not be same in different conditions [34].
2.2 Metabolic models and modeling strategies

A representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in a usable form is referred to as a mathematical model [35]. Dynamic models are mathematical models, represented by differential equations. In other words, a model that accounts for the element of time is called dynamic model in contrast to static models, which does not takes account into time [36].

The enzyme catalysed conversions of metabolites can be represented mathematically via a reaction kinetics. These reaction kinetics vary in different ways defining the modeling system or approach. With law of mass action, the rate of change of these metabolites are obtained thus forming a system of Ordinary Differential Equations (ODE). When the detail information about the kinetics are not known, the stoichiometry of the metabolite is utilized to obtain a reaction flux in steady state conditions [37]. These quantitative representations of metabolic networks are known as metabolic models.

2.2.1 Model requirements

Biological interaction and processes are complex and non-linear. A mathematical model should be able to capture non-linearities and a priori that does not excludes relevant biological phenomena are required. The model should be time dependent. With few molecules
in the network, biological systems may have stochastic features. In this case, the fundamental laws of kinetics and thermodynamics are not directly applicable and the biological behaviour becomes difficult to predict. Thus, in addition to grasping a deterministic phenomenon, the mathematical model should also be able to capture stochastic behaviours. Biological reactions rarely happen in a homogeneous environment and are often restricted to surfaces, channels, organelles or compartments. This feature is sometimes important, and thus the ability of handling spatial processes is necessary for a comprehensive mathematical analysis. Biological phenomena evolve over distinctly different time scales and are controlled from different levels of organization. Moreover, discrete event affects continuous trends for instance sudden activation of gene and by events that occurred in the past and cause a delayed effect. No theoretical or computational framework exists to deal with all these aspects [38].

### 2.2.2 Stoichiometric models

Stoichiometric properties of a model is considered time invariant in contrast to kinetic properties. Stoichiometry describes a diagram of the pathway and describes which fluxes enter or exit the system. This diagram is then translated into a matrix equation. The stoichiometric matrix resulted consists of positive, negative or zero elements that tells which metabolites are converted into which other metabolites. Zero elements in the matrix means that a metabolite and reaction is unrelated. The sign represents the material flow direction and indicates whether the reaction increases or decreases the concentration of a given metabolite. For instance, if one substrate molecule gives two product molecules, the respective element in stoichiometric matrix or the gain in product is +2.

A differential equation, formulated as:

\[
\frac{dM}{dt} = S \cdot v
\]

(2.1)

where \( M \) is a vector of metabolite concentrations, \( S \) is a stoichiometric matrix and \( v \) is a vector of fluxes.

Let us consider following network and devise stoichiometric matrix \( S \) and flux vector \( v \).

**Figure 2.2:** Network structure to construct stoichiometric model, figure adapted from [37]
2.2. Metabolic models and modeling strategies

With the network structure and stoichiometries (here the stoichiometry of each metabolite is 1), the next step is to construct a stoichiometric matrix and replace \( S \) in equation (2.1).

\[
\frac{dM}{dt} = \begin{bmatrix}
-1 & -1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & -1 \\
0 & 1 & -1 & 0 & 0 & -1 \\
\end{bmatrix} \begin{bmatrix}
v_1 \\
v_2 \\
v_3 \\
v_4 \\
b_1 \\
b_2 \\
\end{bmatrix}
\]

(2.2)

Stoichiometric models have a wide applicability in determination of flux rate. Mostly, these models are studied for metabolic system in a steady state, where the material flow into the pool equals the material flow out of the pool and all flux rates are constant. With this assumption, the rate of change of metabolite is zero in equation (2.1) and the differential equation now turns into a linear algebraic equation.

Stoichiometric models are sometimes studied under pseudo-steady state (PSS) assumption, where the concentrations of the analyte rapidly adjust to new levels [36]. Under this assumption, it is reasonable to neglect the instantaneous changes of metabolites and set the rate of change to zero. These models are then used to optimize the network for desired metabolite production using Flux Balance Analysis (FBA) discussed in methods chapter, section 3.1.1. When complete time course of metabolites are available, the flux distribution at each time point can be determined with or without PSS assumption [39], [40].

2.2.3 Kinetic models

If detailed information about the kinetics of the metabolic reactions in the pathway is available, it is possible to describe its dynamics by incorporating kinetic features in flux representation \( v \) of the general stoichiometric model in equation (2.1). The crucial step in combining the stoichiometric properties with kinetic feature is to search for appropriate function forms to represent the flux quantities \( v_i^+ \) and \( v_i^- \) in equation (2.3), which later translates into representation of vector \( v \) in equation (2.1).

\[
X_i = v_i^+ - v_i^- = v_i^+(X_{1i}, \ldots, X_{ni}) - v_i^-(X_{1i} \ldots X_{ni}), \quad i = 1, \ldots, n
\]

(2.3)

Here, \( X_i \) denotes the concentration or amount of a variable or variable pool and \( n \) is the total number of time dependent variables in the system. \( v_i^+ \) and \( v_i^- \) are the reaction fluxes entering or leaving the system. It is known that stochastic, spatial and time scale effects exist in biological systems and it is possible to use approximations, that simplify the model. If the approximations are valid, then they give rise to simplified system representations based on ODEs, of which the generic format is given by equation (2.3) [38].
Figure 2.3: By adding kinetic information in a stoichiometric model, it is transformed into a kinetic model \[38\]

In figure 2.3, the kinetic information is incorporated to transform a stoichiometric model to kinetic model. The types of modeling approaches for kinetic modeling: mechanism based modeling approach and approximated approaches are discussed in section 2.3.

2.3 Mechanism based models

2.3.1 Michaelis-Menten and similar rate laws

Michaelis-Menten rate law is based on the concept that a substrate and an enzyme form a transient complex, which has potential to form product or dissociate back to enzyme and substrate. The enzymatic reaction is given as \[41\]

\[ E + S \xrightleftharpoons{k_i}{k_r} ES \xrightarrow{k_{cat}} E + P \]

Here, \( k_f, k_r, \) and \( k_{cat} \) denote the rate constants. Under certain assumptions such that the enzyme concentration being much less than the substrate concentration; the rate of product formation is given by

\[
\frac{dp}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{2.4}
\]

\( V_{max} \) is the velocity of reaction and \( K_m \) is Michaelis-menten constant and both are known as reaction parameters.

The modeling of enzymatic reaction in this approach is simplified considerably under quasi-steady state assumption, which states that the intermediate complex does not change appreciably over time \[38\]. Even they are straight-forward to set up, complete descriptions of complex mechanisms may become massive if several reactions are involved \[42\]. Thus, parameter estimation requires experimental data and the simplicity of the model vanishes as the number of reactions increases. Other similar rate laws include convenience kinetics which is widely used in biological modeling.
2.3. Mechanism based models

2.3.1. Convenience kinetics

Convenience kinetics is a generalised form of Michaelis-Menten kinetics that covers all possible stoichiometries, enzyme regulation and can be derived from random order mechanism. Considering a reaction,

\[ \alpha_1 A_1 + \alpha_2 A_2 \rightleftharpoons \beta_1 B_1 + \beta_2 B_2 \]

the kinetic equation for the rate of the reaction given by convenience kinetics is as follows:

\[
v(a, b) = E \cdot \frac{k_{cat}^+ \prod_i \tilde{a}_i^{\alpha_i} - k_{cat}^- \prod_j \tilde{b}_j^{\beta_j}}{(1 + \tilde{a}_i + \cdots + \tilde{a}_i^{\alpha_i}) (1 + \tilde{b}_j + \cdots + \tilde{b}_j^{\beta_j}) - 1}
\]  

where \( E \) is the enzyme concentration, \( k_{cat}^+ \) and \( k_{cat}^- \) are turnover rates. \( \tilde{a}_i = a_i / k_{Ma} \) and \( \tilde{b}_j = b_j / k_{Mb} \). \( k_{Ma} \) and \( k_{Mb} \) are Michaelis-Menten constants. In an enzyme catalysed reaction, these \( k_{Ma} \) and \( k_{Mb} \) are the dissociation constants for a reactant to bind with an enzyme. \( \alpha_i \) and \( \beta_j \) are the stoichiometric coefficients in the reaction. The reaction velocity is also controlled by the presence of activators and inhibitors:

For an activator, \( h_A(d, k^A) = \frac{d}{k^A + d} \) (2.6)

For an inhibitor, \( h_I(d, k^I) = \frac{k^I}{k^I + d} \)

Here, \( k^A \) and \( k^I \) are the activation constants and inhibition constants respectively. \( d \) is the concentration of the regulator. The turnover rates of the reactions are replaced equilibrium constant \( K^{eq} \) which is given by:

\[
K^{eq} = \prod_i (c_i^{eq})^N
\]  

(2.7)

Where, \( c_i^{eq} \) is the vector of metabolite concentration in the equilibrium state and \( N \) is the vector of stoichiometries. By setting equation (2.5) to zero, Haldane relationship is obtained for convenience kinetics that reads as:

\[
K^{eq} = \frac{\prod_j b_j^{\beta_j}}{\prod_i a_i^{\alpha_i}}
\]  

(2.8)

The significance of \( k^M \) is such that it is the substrate concentration at which the reaction rate is half of its maximum value. In other words, if an enzyme has a small value of \( k^M \), it achieves its maximum catalytic efficiency at low substrate concentrations. The smaller the
value, the more efficient is the catalysis. The value depends upon temperature and pH of
the reaction conditions and also the substrate.

From equations (2.5) - (2.8), the general form of equation of convenience kinetics for
reversible reaction now can be written as:

\[ v(a, b) = h_A(d, k^A)h_1(d, k^l) \frac{E \prod_i \tilde{a}_i^{\alpha_i} - \frac{E}{k^{eq}} \prod_j \tilde{b}_j^{\beta_j}}{\prod_i (1 + \tilde{a}_i + \cdots + \tilde{a}_i^{\alpha_i}) + \prod_j (1 + \tilde{b}_j + \cdots + \tilde{b}_j^{\beta_j}) - 1} \] (2.9)

### 2.4 Approximated approaches

**S-system and GMA model**

Canonical models captures the non linear dynamics while keeping the mathematics relatively simple. The most promising canonical models in metabolic modeling are Generalized Mass Action (GMA) and S-system structures within Biochemical Systems Theory (BST). In BST framework, each flux is approximated by a power law, that corresponds to Taylor series expansion in the logarithmic space. In the S-system formalism, each equation has a particularly simple format: the change in system variables is given as one set of influxes minus one set of effluxes (cf. \( v_i^+ \) and \( v_i^- \) in equation (2.3)) and each set is collectively written as one product of power law functions. Thus, the generic S-system formulation reads [38], [35], [44]:

\[ \dot{X}_i = \alpha_i \prod_{j=1}^n X_j^{g_{ij}} - \beta_i \prod_{j=1}^n X_j^{h_{ij}}, \quad i = 1, 2, \ldots, n \] (2.10)

where \( \dot{X}_i \) represents a time-dependent variable (metabolite) and \( n \) denotes the number of variables in the system. The non-negative multipliers \( \alpha_i \) and \( \beta_i \) are rate constants which quantify the turnover rate of the production or degradation, respectively. The real numbers \( g_{ij} \) and \( h_{ij} \) are kinetic orders that reflect the strengths of the effects that the corresponding variables \( X_j \) have on a given flux term. A positive value signifies an activating or augmenting effect exerted by \( X_j \), a negative value signifies an inhibitory effect. A kinetic order of zero implies that the corresponding variable \( X_j \) does not have any effect on a given flux. In some instances, \( m \) independent variables, which are typically constant during each mathematical experiment, may be included. They do not have their own equations but enter the power-law terms just like dependent variables, so that the product runs from 1 to \( n + m \).

In the GMA formalism [38], instead of aggregating all influxes and all effluxes into one term each, all influxes and effluxes are approximated individually with power-law terms such that

\[ \dot{X}_i = \sum_{k=1}^{k_i} \left( \pm \gamma_{ik} \prod_{j=1}^n X_j^{f_{ijk}} \right), \quad i = 1, 2, \ldots, n. \] (2.11)
where the rate constants \( \gamma_{ik} \) are non-negative and the kinetic orders \( f_{ijk} \) may have any real values as in the S-system form. Also as before, independent variables may be included. It should be noted that differences between these two formulations only exist at branch points, whereas all other steps are identical [38].

2.5 Dynamic Systems and Simulations

Once a model is constructed with the approaches given above, the kinetic parameters are put together from literature or experimental measurement. With experimental data, which are the time course concentration profile of metabolites in the pathway, the model is extended and improved by estimating the parameters and the initial concentration of metabolites missing in the data. The model is then checked for its consistency with different concentration profile of input, also known as model validation. The model now could be simulated to predict experimental results and analyze the reactions in the model, governing the outputs.

2.6 Parameter estimation

Parameter estimation in systems biology is an iterative process to develop data-driven models for biological systems that should have predictive value [45]. Parameter estimation begins with a guess about the values and are changed to minimize the difference of these values in the model and data using a measurement. A priori identifiability takes into account the model structure (correct structure assumed; unknown parameter values specified as part of model structure), a usually known input and model output (error free data related to measurement variables). We are given a specified measurement equation linking the model to output data. priori identifiability question is “does the model contains enough information to estimate model parameters”? posteriori identifiability shows how well the parameter vector has been determined given dataset that is noisy and sparse [45], [38].

Dynamic models from kinetic equations are typically given in the form of ODEs or DAEs (Differential Algebraic Equation). However, the main theme is such that, within a network of pathways, each metabolite has a rate of change with respect to time. DAEs are of the form:

\[
\begin{cases}
A \frac{dx(t,p)}{dt} = f(t,x(t,p),p,u(t)), & t_0 < t \leq t_c \\
x(t_0, p) = x_o(p)
\end{cases}
\tag{2.12}
\]

ODEs are of the form:

\[
\frac{dx(t,p)}{dt} = \sum_i S_i v_i
\tag{2.13}
\]

\[
= f(t,x(t,p),p,u(t)), & t_0 < t \leq t_c
\]
Which tells us that the rate of change of species $x$ depends upon a function of vector of time $t$, states $x(t,p)$, parameter vector $p$ and input $u(t)$. When component of initial states vector $x_o$ is not known, they are considered as unknown parameters such that $x_o$ may depend on $p$. $A$ is a $n \times n$ diagonal matrix with 0s or 1s in diagonal, more specifically, 1 for ODE and 0 for algebraic equation. $S_i$ is the stoichiometric coefficient and $v_i$ is the velocity of reaction $i$.

We are also given vector of observables:

$$g(t, x(t, p), p, u(t))$$  \hspace{1cm} (2.14)

which are quantities that are state variables or combination of state variable in the model, measured experimentally and possibly vector of non-linear constraint, given as:

$$c(t, x(t, p), p, u(t)) \geq 0$$  \hspace{1cm} (2.15)

If we assume that measurements are taken in N time points, such that: $(y_i, \ldots, y_N)$ are measurements carried out at $t_i, \ldots, t_N$ time points, the model value for parameter vector $\hat{p}$, computed by integrating equation (2.12 - 2.13) and computing observable function for equation (2.14), such that

$$\hat{g}_i = g_i(t_i, x, \hat{p}, u)$$  \hspace{1cm} (2.16)

The vector of discrepancies between model and experiment is given by:

$$|g(t, x(t, p), p, u(t)) - y|$$  \hspace{1cm} (2.17)

The only uncertainty involved in equation (2.12 - 2.13) are the vector of parameter $p$. The optimization problem is to minimize a metrics for discrepancy which are usually euclidean norm or sum of the squares weighted with the error in the measurement:

$$V_{MLE}(p) = \sum_{i=1}^{n} \frac{(g_i(t_i, x, \hat{p}, u) - y)^2}{\sigma_i^2}$$  \hspace{1cm} (2.18)

$$= e^T(p)W(e(p))$$

This measure results form Maximum Likelihood Estimation theory and our aim is to select $p$ such that $V_{MLE}$ is minimized, which is the least square estimate and known as cost function, objective function and goal function.
3 Methods

Contents

3.1 Computational theory, tools and algorithms .................. 18
3.2 Akaike Information Criterion .............................. 24
3.3 Sensitivity Analysis ........................................... 24
3.4 Glycolysis in *L. lactis* within BST framework ........... 26
3.5 Model comparison and extension .............................. 27
3.1 Computational theory, tools and algorithms

3.1.1 Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) is an optimization method, widely used for studying metabolic networks, genome-scale metabolic network and stoichiometric models in particular. FBA poses a linear programming approach that calculates the flow of metabolites (flux) through the metabolic network, which helps to predict the growth rate of an organism or rate of production of biologically important or relevant metabolite (desired metabolite) in the system. FBA has proven useful in the study of metabolic systems. If assumed that the main objective of an organism is to produce the desired metabolite of interest, FBA addresses the problem that, given a set of reactions, what is the combination of fluxes that maximizes the production of the desired metabolite [46], [47], [48].

Metabolic reactions are represented as a stoichiometric matrix $S$ of size $m \times n$. Each row in the matrix represents a metabolite and each column in the matrix represents a reaction ($m$ metabolites and $n$ reactions) as discussed in chapter 2, section 2.2.2 in the example given in equation (2.2). The flux through the network is represented by $v$ vector, of length $n$, concentration of all metabolites are represented by the vector $x$, with length $m$. At steady state, $dx/dt = 0$, and

$$S.v = 0$$  \hspace{1cm} (3.1)

Any $v$ that satisfies this equation is said to be in the null space of $S$. In large networks where the number of reactions are more than the number of compounds, i.e., there are more number of unknown variables than equation, there is no unique solution to this system of equations.

The objective function that FBA maximizes or minimizes is

$$Z = c^T.v$$  \hspace{1cm} (3.2)

where $c$ is a vector of weights which tells how much each reaction contributes to the objective function. When only one reaction is desired for maximization or minimization, vector $c$ contains zeros with 1 at the position of reaction of interest. Such kind of optimization can be done using linear programming. It basically solves the equation $S.v = 0$, given a set of upper and lower bounds on $v$, maximizing or minimizing $Z$. The output is a $v$, that maximizes or minimizes the objective function. The general form of FBA is:

$$\text{maximize} \quad Z = C^T.v$$  \hspace{1cm} (3.3)

subject to $S.v = 0$

$\text{lowerbound} \leq v \leq \text{upperbound}$

$v$ is the vector of fluxes (combination of fluxes) to be determined.
3.1.2 Implementation

FBA is performed in OptFlux on a toy network, downloaded from [http://darwin.di.uminho.pt/optfluxwiki/index.php/OptFlux3:OPK](http://darwin.di.uminho.pt/optfluxwiki/index.php/OptFlux3:OPK) and analyzed. The network structure is given in figure 4.1. The lower bounds and upper bounds for every reaction, except the substrate uptake reaction is kept -10000 and 10000 respectively allowing the uptake of each metabolite in the network. The substrate uptake rate is constrained to -36.5 (realistic uptake rate of glucose in *L. lactis* [24]). Different analysis performed in OptFlux are given in results chapter, section 4.1.

3.1.2.A OptFlux

OptFlux is an open-source software to support *in silico* metabolic engineering tasks. It allows the use of stoichiometric models for phenotypic simulations for both wild-type and mutant organisms, using the method of FBA and analyze pathway through the calculation of Elementary Flux Modes. In a metabolic network, there are sub-networks that allow a metabolic reconstruction network to function in a steady state. Elementary modes can be used to understand cellular objectives for the overall metabolic network, given any environmental conditions allowing to define the fluxes for uptake reactions, or genetic conditions to define genetic modifications over the original strains [30], [49]. OptFlux can be downloaded from [http://www.optflux.org/](http://www.optflux.org/).

A plugin additive for OptFlux, known as Optknock answers how many reactions need to be knocked out or eliminated from a system for the maximization or minimization of a given reaction. OptKnock can be downloaded from [http://darwin.di.uminho.pt/optfluxwiki/index.php/OptFlux3:OPK](http://darwin.di.uminho.pt/optfluxwiki/index.php/OptFlux3:OPK). OptFlux in this work is used to perform FBA on a toy network, given in figure 4.1 and each output is maximized.

**Kinetic modeling**

Kinetic modeling of metabolic process begins with an infrastructure of reaction pathways and their rate laws. When the reactants and their rate laws for conversion to their respective product(s) are known, they are fed into cellDesigner. Once a complete model is set up, the SBML file is exported to COPASI for further analysis such as parameter estimation, dynamic time course simulation, sensitivity analysis etc.

*Systems Biology Markup Language* (SBML) describes the models and are inbuilt in many tools available. The main goal is to exchange models within different environment of simulations and analysis.

3.1.3 cellDesigner

cellDesigner is a user interface based software, that takes in the metabolites and the rate laws of the inter conversion of metabolites and produces a SBML file for further use [28].
Main features of cellDesigner are listed as:

- Representation of biochemical semantics;
- SBML compliant;
- Integration with Systems Biology Workbench - enabled simulation/analysis modules;
- Integration with native simulation library (SBML ODE Solver);
- Capability of database connections;

cellDesigner supplies a process diagram editor with the standardized technology for every computing platform, so that it could confer benefits to as many users as possible. The main standardized features that cellDesigner supports could be summarized as “graphical notation”, “model description” and “application integration environment” [50]. The software is available at [http://www.celldesigner.org/](http://www.celldesigner.org/).

A plugin additive for cellDesigner, known as SBML squeezer generates the kinetic equations for a given reaction. Several systems of rate equations like GMA, Michaelis-Menten kinetics, convenience kinetics, ordered mechanisms can be generated using this plugin. However, the rates should be manually checked as it may differ in the context of regulators for instance cellDesigner does not differentiate between catalysis by a metabolite and an activator. cellDesigner is used in this work to construct kinetic models. A snapshot of cellDesigner, with a model from [51] is given above.
3.1. Computational theory, tools and algorithms

3.1.4 COPASI

COPASI, abbreviated as Complex Pathway Simulator, is a tool that provides a full Graphic User Interface, including functions for creating and editing models and plotting results. COPASI’s graphical interface is similar to windows explorer in operation, where on left, there is a set of functions organized in a hierarchical way; on the right there is a larger window that contains all of the controls to operate the function selected on the left.

Figure 3.2: Snapshot of COPASI with model from [51]

The major group of functions in the program are as follows [29]:

- Model, where the model can be edited and viewed according to a biochemical or mathematical perspective.
- Tasks, consisting of the major numerical operations on the model: steady state, time course, stoichiometry, metabolic control analysis and Lyapunov exponents. Below each task an entry with results sill appear after the task has been run.
- Multiple tasks, which are operations repeating elementary tasks: parameter scanning, optimization and parameter estimation.
- Output is where plots and reports are defined and listed.
- Functions containing the mathematical functions available, such as the rate laws.

COPASI is equipped with a number of diverse optimization algorithms that can be used to minimize or maximize any variable of the model. The algorithms that are used in this
work to minimize the objective function during parameter estimation are particle swarm optimization, evolutionary programming and Hooke & Jeeves algorithm.

**Experimental data**

The experimental data are obtained from *in vivo* NMR time series measurements. *In vivo* Nuclear Magnetic Resonance is a powerful analytical technique to monitor the dynamics of intracellular metabolite and co-factor pools following a glucose pulse, and is also used for characterization of chemical mixtures, the measurement of reaction rates in steady state and the determination of isotopic distribution within molecules. Nuclear magnetic resonance is based on the response of nuclides that possess an intrinsic magnetic moment to an external magnetic field. The experimental data used to estimate the parameters in this work are time series data of 40 mM glucose utilization during anaerobic growth conditions at time zero in *L. lactis*. 40 mM and 80 mM glucose utilization time course data for metabolites ATP, P, Glucose, Lactate, NAD, NADH, PEP and FBP were available, obtained from Neves et al. 2005, [52]. 80 mM glucose impulse data were used to validate the model after estimating the parameters.

The algorithms and their settings used during the development and reconstruction of model to estimate the parameters are briefly discussed below.

### 3.1.5 Algorithms for parameter estimation used

Algorithms used for parameter estimation methods in COPASI determines the minimum of the objective function for a set of parameters obtained by the algorithm. Two classes of methods are widely used which are global optimization and local optimization. Local optimization methods typically converge fast to a minimum and usually have a theoretical proof of convergence to the minimum if the initial guess is sufficiently close to the minimum. Global optimization searches all over the parameter space to find smaller and smaller values for the objective function. However, there is no proof of convergence as in case of global optimization method [45].

#### 3.1.5.A Global Optimization

Global optimization methods usually are of stochastic nature to prevent the search procedure being trapped in a local minimum. The optimization proceeds searching the parameters that maximizes or minimizes the objective function from all the parameter space [45]. Particle swarm algorithm and evolutionary programming are two global optimization algorithms that are used in this work for parameter estimation purpose.

**Particle Swarm Optimization**

It is developed from swarm intelligence and is based on the research of bird and fish flock movement behavior. While the birds are searching for food from one place to another,
there is always a bird that can smell the food very well, i.e., the bird is perceptible of the place where the food can be found, having the better food resource information. Because they are transmitting the information, especially the good information at any time while searching the food from one place to another, conduced by the good information, the birds will eventually flock to the place where food can be found. As far as particle swarm optimization algorithm is concerned, solution swarm is compared to the bird swarm, the birds’ moving from one place to another is equal to the development of the solution swarm, good information is equal to the most optimist solution, and the food resource is equal to the most optimist solution during the whole course. Due to its many advantages including its simplicity and easy implementation, the algorithm can be used widely in the fields such as function optimization, the model classification, machine study, neural network training, signal processing, vague system control, automatic adaptation control etc. [53], [54].

Particle swarm is used in this work for comparison between two models when the parameters were unknown and guessed. While performing the parameter estimation task in COPASI, the swarm size is kept 100 with an iteration limit of 2000, standard deviation of $10^{-6}$, random number generator 1 and seed 0, as given by COPASI.

**Evolutionary Algorithms**

Evolutionary algorithms are inspired by biological evolution. Potential solutions are the individuals of a population. To get new solutions, the individuals are replaced using reproduction, natural selection, mutation, recombination and survival of the fittest.

Initially, a population of random individuals (parameter vectors) is created. Next, the corresponding objective function is evaluated which defines the fitness of the individual. The fitness is assigned a probability and selected for next generation (the higher the probability, greater the fitness is). New individuals are created by two operators: recombination or cross over and mutation. Recombination creates one or more children from two parents while mutation results in one child from one parent. These new population compete with old population for their place in the next generation (survival of the fittest). This process is repeated until a candidate with sufficient quality of solution is obtained. Genetic algorithm, evolutionary computation are examples of evolutionary algorithms [45].

Evolutionary programming in COPASI is used while extending the model in this work. A population size of 100 individuals with 1000 number of generations, random number generator 1 and seed 0 is used to estimate the parameters.
3.1.5.B  Local Optimization

Local optimization maximizes or minimizes the objective function by searching the parameters in a constrained parameter space, this optimization is done specially after global optimization, when the distribution of parameter is known more or less.

Hooke & Jeeves method

This method consists of two steps. First, a series of exploratory changes of the current parameter vector are made, typically a positive and negative perturbation of one parameter at a time. This step returns a direction in which the objective function decreases. In next step, the pattern moves and the information obtained is used to find the best direction of the minimization process. The perturbation is halved and the same process is repeated until a minimum objective function is found [45].

Hooke & Jeeves method for parameter estimation is used in this work after estimating the parameters of a kinetic model by one of the above global optimization algorithms. A tolerance of $10^{-5}$, with tolerance limit of 50 and rho of 0.2 is provided in COPASI for parameter estimation.

3.2  Akaike Information Criterion

Akaike Information Criterion ($AICc$) of a model is given by the following equation

$$AICc = 2k + n \left( \ln \left( \frac{2\pi SSR}{n} \right) + 1 \right) + \frac{2k(k+1)}{n-k-1}$$  \hspace{1cm} (3.4)

Where, $SSR$ is the objective function value, $k$ is the number of parameters and $n$ is the number of data points. The $AICc$ is an information-theory based measure of parsimonious data representation that incorporates the goodness of the fit $SSR$ as well as the complexity of the model $k$ and is used to rank the candidate models, thereby giving an objective measure for model selection and discrimination. The lowest the outcome of the equation or rank, the better the model performance is [55]. $AICc$, in this work is used to rank two different models while comparing different modeling approaches.

3.3  Sensitivity Analysis

Biochemical models are featured by employing a number of parameters, such as enzyme levels, regulators, binding constants, hill coefficients and equilibrium constants. These parameters are treated as constant and their value do not change in the time-scale of interest. These values are usually poorly known and are dependent on external factors such as experimental and cellular conditions. Even a set of experimentally determined parameters are uncertain to approximate a biological system, because some of the parameters are usually
3.3. Sensitivity Analysis

taken from measurements reported by different laboratories with different experimental conditions. Sensitivity analysis provides an insight into which model behavior depends upon which parameter values \cite{55}, \cite{57}, \cite{58}. Saltelli et al. \cite{59} explained sensitivity analysis as “The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.” In biochemical systems modeling, sensitivity analysis tells us how much the output such as concentration of species and reaction fluxes depend upon the parameters.

3.3.1 Local parameter sensitivity analysis

Local parameter sensitivity analysis or forward sensitivity calculates the local sensitivity coefficients of a model \cite{60} \cite{61}. If the model to be analyzed contains a set of ODEs, with model output $y \in \mathbb{R}^N$ and parameter set $P \in \mathbb{R}^{N_p}$, then:

$$\dot{y} = g(t, y, P)$$
$$y(t_0) = y_0(P)$$

The vector $s_i$ represents the sensitivity of the solution $y$ with respect to parameter $P_i$

$$s_i(t) = \frac{\delta y(t)}{\delta P_i}$$

(3.6)

It is often customary to account for different magnitude of the parameters. From equation (3.6),

$$s_i(t) = \frac{\delta y(t)}{y(t)} \times \frac{P_i}{P_i}$$

(3.7)

Accounting the model complexity, we cannot calculate the model sensitivity as described above. Let us consider a model with a single parameter $p$ and model output $y = f(t, p)$. The sensitivity is given as:

$$S = \frac{\delta y}{\delta p}$$
$$= \lim_{h \to 0} \frac{f(t, p + h) - f(t, p)}{h}$$

(3.8)

For a sufficiently small discrete $h$, $S$ can be approximated as:

$$S \approx \frac{f(t, p + h) - f(t, p)}{h}$$

(3.9)

3.3.2 Computational implementation

To implement sensitivity analysis, equation (3.9), is extended by changing the parameter, solving the system and calculating the changed area under the curve given by time-course
of a metabolite with reference to its wild type state. \( RD \) values, which are coefficients that describes the effect of parameter perturbation to an output in a metabolic network are calculated. Here, single parameter is perturbed and the time course of the metabolite with the perturbation is observed. These perturbations are repeated for all the enzyme levels. Each time a parameter is perturbed, the \( RD \) values are calculated. In the end of the experiments, a set of these coefficients for each parameter to a given metabolite is returned.

Perturbation is also carried out with two parameters at once. To begin with, all the possible combination of parameters for double perturbation were analyzed.

The \( RD \) values were calculated as [62],

\[
RD = \frac{\int_{t_0}^{t_f} y_p(t) dt - \int_{t_0}^{t_f} y_c(t) dt}{\int_{t_0}^{t_f} y_c(t) dt}
\]  

(3.10)

Here, \( \int_{t_0}^{t_f} y_p(t) dt \) is the integral of perturbed state and \( \int_{t_0}^{t_f} y_c(t) dt \) is the integral of wild-type. Graphical illustration of calculating the sensitivities or \( RD \) value is given in following figure:

![Graphical illustration of dynamic sensitivity analysis](image)

Figure 3.3: Graphical illustration of dynamic sensitivity analysis. The solid lines is the metabolite time course without any parameter perturbation and the dashed line is the time course of the same metabolite with a parameter perturbation over a range. The shaded area gives the \( RD \) values. Figure adapted from [63].

A MATLAB script was developed for the computational implementation, of which the source code is given in appendix F.

### 3.4 Glycolysis in L. lactis within BST framework

The glycolytic pathway in \( L. \) lactis is shown in figure B.1 adapted from [25], [64], and is simulated and the phenotype are studied. Here, glucose, ATP, \( P_i \) and NAD are given as offline concentration in the form of raw data that were smoothed and splined. Also, the time dependency of glucose consumption (input) is described by a time dependent function as given in equation B.1, which is used to get sigmoid decay of glucose utilization. These
variables, that are splined, are involved in many different reactions in a complete metabolic network and are problematic to include in a small network like the one presented in this section, given in figure B.1 [25], [64].

The parameters were obtained from [64]. Known that the reaction between PEP and P3GA is extremely fast [64], [25], [52], both of the variables are merged into PGAPEP pool, such that PGA = \( k_{45} \times \) PEP and PGAPEP = PGA + PEP. With the rate equations and values of parameters given in table B.1 and table B.2 in appendix B, the system of ODEs was simulated.

**Model mimicking glycolysis**

With a simplified model as given in figure B.1 [25], the validity and efficacy of the proposed mechanism of starting and stopping glycolysis was assessed by mimicking the activation of FBP in glycolysis to a toy network as given in figure C.1 in appendix C [25]. X1 corresponds to G6P, X2 to FBP, X3 to P3GA, X4 to PEP and X5 to pyruvate. Another metabolite X6 is added to assess the regulation of FBP in lactate production later. The simulation results are discussed on results chapter, section 4.2.

In figure C.1 an early metabolite X2 activates the degradation of X4, similar to FBP activating PK reaction. The ATP and PTS based input are given by Input1 and Input2, the values of which are given in table C.1 The parameter \( h_{42} \) signifies FBP activation in X4 → X5 reaction. A metabolite X6 and a regulator \( h_{52} \) is added to the system to explain the regulation of FBP in production of X6 (lactate). \( h_{52} \) is an activator for reaction X5 → X6. In equation (C.1), a sixth equation is added given as:

\[
\dot{X}_6 = X_5^{0.3} X_2^{h_{52}} - X_6^{0.3}
\]  

(3.11)

that accounts for the production of X6, or lactate with X2 (FBP) regulating its expression.

**3.5 Model comparison and extension**

**3.5.1 Model comparison**

The methods presented here in section 3.5.1 have been presented as a poster in Bioinformatics Open Days, 2nd edition, University of Minho, Braga, Portugal; March 2013. Results are discussed in chapter 4 section 4.3.

Mechanistic approach and approximated approaches for model construction are compared. Within approximated approaches, S-systems modeling and GMA modeling as discussed in chapter 2 section 2.4 only differs in terms of influx and out flux of a metabolite in a reaction. Since GMA equations are readily implemented in cellDesigner by SBML
squeezer plugin, comparison of two different models with two different system of equations namely GMA and convenience kinetics were assessed for the network topology from [25], shown in figure 5.1 to get a better model in terms of objective function and its rank by $AIC_c$. In order to get insights into the glycolytic pathway in $L. lactis$ and to elucidate the enzyme levels that affects acetoin and butanediol production, the model is later extended.

Here, the time dependency of glucose decay is omitted in either of modeling approach and PGAPEP are considered two different stat variables. Few species are kept fixed with values for $NAD = 4.21$, $NADH = 3$, $ATP = 1$, $ADP = 5$ and $P = 1$. With known available data and initial concentration for glucose, PEP, P3GA and FBP, the model parameters were estimated. The fits were analyzed and both the models were validated with $80 mM$ glucose impulse at time zero. The reaction rates and initial conditions used in both approaches are given in appendix D. The equations refereed as $v_1 - v_6$ are for reactions catalyzing $\text{glucose} \rightarrow \text{G6P}$, $\text{G6P} \rightarrow \text{FBP}$, $\text{FBP} \rightarrow \text{3PGA}$, $\text{3PGA} \rightarrow \text{PEP}$, $\text{PEP} \rightarrow \text{pyruvate}$ and $\text{pyruvate} \rightarrow \text{lactate}$ conversion respectively in equations (D.1) and equation (D.2). Once both the systems are constructed and parameters estimated, with $(AIC_c)$ from equation (3.4), the two models are ranked. The values of the initial and estimated parameters are given in table 4.6 in results chapter, section 4.3.

3.5.2 Model extension

The network topology is extended to incorporate the glycolytic pathway in increased level of details i.e. including our metabolites of interest. An anaerobic glycolytic model in $L. lactis$, taking account into phosphate transport reaction and ATP degradation reaction is reconstructed. The network topology, rate equations and initial parameters are obtained from [51], [26], [65]. With few parameters unknown, these values were guessed such that the dynamics of metabolites (acetoin, butanediol, ethanol, formate) would stay low or resemble the production in wild type $L. lactis$. The model was then fitted with experimental data which although showed good fittings and validations, did not accounted for transient dynamics of acetoin and butanediol (not shown here). The parameters that were still not available for model to be reconstructed, which were guessed previously, were now obtained from [66]. The experimental data for the analysis of the model fittings and validation were available from Neves et al., 2005 [52]. The parameters are estimated using COPASI and a model with least objective function is selected for further analysis. The model topology obtained is given in figure 4.11 results chapter, section 4.4. The reaction rate laws, parameters, initial conditions used are given in appendix E.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Comparision of stoichiometric and kinetic modeling</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation of Glycolysis within BST framework</td>
<td>35</td>
</tr>
<tr>
<td>4.3</td>
<td>Approximated vs. semi-mechanistic kinetics</td>
<td>40</td>
</tr>
<tr>
<td>4.4</td>
<td>Model extension</td>
<td>42</td>
</tr>
<tr>
<td>4.5</td>
<td>Sensitivity analysis in extended model</td>
<td>46</td>
</tr>
</tbody>
</table>
4.1 Comparision of stoichiometric and kinetic modeling

The goal of this section is to use an optimization strategy in a stoichiometric model, use the same network to translate into kinetic model and observe using sensitivity analysis the enzyme levels responsible for each output metabolite and its coherency with FBA.

![Toy network used to compare FBA and sensitivity analysis. The coefficient of reaction R2 (M1 → M3) are such that two molecules of M1 create one molecule of M3, while the stoichiometric coefficients of all other reactions are 1.](image)

With the model in figure 4.1, the analysis performed are given as follows:

4.1.1 Critical Reaction determination

OptFlux helps us to find the critical reactions/genes in a metabolic network. The critical reactions are the reactions, without which the steady state principle (no accumulation of metabolites inside the cell) will not hold true. The critical reactions in case of the toy network given in figure 4.1 are ‘substrate consumption’ and ‘biomass production’.

4.1.2 Maximize ‘biomass’, ‘desired’ and ‘R_ext’ formation

On the network shown in figure 4.1, FBA is performed to maximize output metabolites formation or flux of different reactions, given as:

![Table 4.1: Flux distribution in figure 4.1, maximizing ‘biomass’, ‘desired’ and ‘R_ext’.](image)

While maximizing for biomass formation, it is observed that the maximization proceeds deleting reactions ‘R2, R3, R_ext, R4’ and ‘Desired’, allowing the flow of flux through ‘R1’.
Net conversion of metabolite M1 to M2 proceeds with a consumption value of -36.5 for M1 and production value of 36.5 for M2.

Two molecules of M1 produces one molecule of M3 comes to play; which can be seen from the flux distribution while maximizing for ‘desired’ reaction, where the substrate consumption rate is double than that of substrate conversion rate to metabolite M3. Reaction ‘R1’ and ‘R_ext’ are deleted to maximize the reaction named ‘Desired’ or metabolite M5_ext production. The net conversion of metabolite M1 to M2 and M5 proceeds with a consumption value of 36.5 for M1 and production value of 18.25 for M2 and M5.

Intuitively, in a small reaction network as mentioned here, we can predict just by looking at the network that to maximize ‘R_ext’ reaction, ‘R1, R4’ and ‘desired’ needs to be deleted, which is true, as given by FBA in table 4.1. Biomass formation is due to the fact that metabolite M2 is being produced by reaction ‘R3’ although reaction ‘R1’ is deleted and in a stoichiometric model, no accumulation of metabolites are allowed since the system is assumed to be in steady state.

4.1.3 Knock-out simulations

Knock-out simulations in OptFlux helps us to delete a reaction from the network and maximize a reaction of interest. Let us backtrack the results in table 4.1, while maximizing for ‘desired’ production by knocking out ‘R_ext’ and ‘R1’ reactions. A knock out simulation for the maximization of ‘desired’ reaction gives a flux distribution as seen in table 4.1 while maximizing for M5_ext.

**Knock-out ‘R4’ and ‘R1’ maximizing Biomass**

The reactions ‘R4’ and ‘R1’ are deleted from the network maximizing the biomass formation. Following results are obtained after performing FBA in the network.

**Table 4.2:** Flux distribution in figure 4.1, maximizing biomass and knocking out R4 and R1.

<table>
<thead>
<tr>
<th>Flux distribution after knock out of R4 and R1</th>
<th>Maximization</th>
<th>Substrate</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R_ext</th>
<th>R4</th>
<th>Biomass</th>
<th>Desired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass</td>
<td>-36.5</td>
<td>0</td>
<td>18.25</td>
<td>18.25</td>
<td>18.25</td>
<td>0</td>
<td>18.25</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Metabolite M1 is consumed with a value of 36.5 and metabolites M4 and M2 are being produced with values of 18.25 each.

4.1.4 OptKnock

While maximizing ‘desired’ reaction formation, optKnock predicts the knock out of two reactions namely, ‘R_ext’ and ‘R1’, which is the same case as in table 4.1 (under column
Say, in the next case, we want to know reactions to be removed for maximum production of metabolite ‘M4_ext’, constraining the substrate uptake to 36.5. OptKnock predicted the knockout of reactions ‘R1, R4’ and ‘Desired’ to maximize ‘R_ext’. The flux distribution after deletion of ‘R1, R4’ and ‘Desired’ reactions are given as:

Table 4.3: Flux distribution in figure 4.1 predicted by OptKnock while maximizing for R_ext.

<table>
<thead>
<tr>
<th>Flux distribution predicted by OptKnock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximization</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>R_ext</td>
</tr>
</tbody>
</table>

With FBA, flux distribution given the constraints on the fluxes are determined. FBA is restrictive to stoichiometric models making itself different in the context of optimization principle and the modeling formalism. When the network is quite large and the information about the fluxes (constraints) is known or the time course of the intermediate species are not very well known, then a stoichiometric model is favored. Here, we have a smaller pathway with significant time course data of the transient species, which makes us to perform kinetic modeling over stoichiometric modeling. This model is translated to a kinetic model using Michaelis Menten equation and sensitivity analysis is performed to cross refer the results with FBA in the same network which is presented in section 4.1.5. With a kinetic model, time course as well as flux distribution in either steady state or during the time course can be calculated.

### 4.1.5 Sensitivity Analysis

A dynamic model of the network structure in figure 4.1 is constructed with Michaelis-Menten kinetics. The rate equations involved, concentration and parameters used are given in appendix A. With a kinetic model, sensitivity analysis is performed on the toy network, of which the results are supportive to FBA. The sensitivity analysis of the toy network, is performed as explained in section 3.3 and calculated as given by equation (3.10). The results of which are shown as follows:
4.1. Comparison of stoichiometric and kinetic modeling

4.1.5.A Sensitivities of M2_ext

Figure 4.2: Sensitivity analysis of toy model for species ‘M2_ext’ after translating into a dynamic model. (Parameter perturbation of +3% : left, −3% : right)

Sensitivity analysis for M2_ext production infers that perturbing two enzyme levels, for ‘R1’ and ‘R2’ affects the production of M2_ext. Decreasing $V_{\text{max}} R_1$ accounts for increasing M2_ext and vice versa. Increasing $V_{\text{max}} R_2$ and $V_{\text{max}} \text{Substrate}$ accounts for increasing M2_ext and vice versa as seen from +3% perturbation of enzyme levels.

4.1.5.B Sensitivities of M4_ext

Figure 4.3: Sensitivity analysis of toy model for species ‘M4_ext’ after translating into a dynamic model. (Parameter perturbation of +3% : left, −3% : right)

Sensitivity analysis for M4_ext production infers that perturbing all enzyme levels affects the production of M2_ext. Increasing $V_{\text{max}} R_{\text{ext}}$, $V_{\text{max}} R_2$, $V_{\text{max}} R_3$, $V_{\text{max}} \text{Desired}$ and $V_{\text{max}} \text{Substrate}$ accounts for increasing M4_ext and vice versa. Decreasing $V_{\text{max}} R_1$ and $V_{\text{max}} R_4$ accounts for increasing M4_ext and vice versa as seen from −3% perturbation of enzyme levels.
4.1.5. C Sensitivities of M5_ext

Figure 4.4: Sensitivity analysis of toy model for species ‘M5_ext’ after translating into a dynamic model. (Parameter perturbation of +3% : left, −3% : right)

Sensitivity analysis for M5_ext production infers that perturbing all enzyme levels except the enzyme levels for ‘biomass’ and ‘desired’ reactions affects the production of M5_ext. Increasing \( V_{\text{max}} R_2, V_{\text{max}} R_3, V_{\text{max}} R_4 \) and \( V_{\text{max}} \text{Substrate} \) accounts for increasing M5_ext and vice versa. Decreasing \( V_{\text{max}} R_{\text{ext}} \) and \( V_{\text{max}} R_1 \) accounts for increasing M5_ext and vice versa as seen from −3% perturbation of enzyme levels.

4.1.6 Sensitivity analysis and Flux Balance Analysis

Comparing sensitivity analysis with FBA from section 4.1, if we observe closely, it is seen that the results inferred from sensitivity analysis supports the results obtained from FBA. The sensitivity and FBA are compared in table 4.4.

Table 4.4: FBA and sensitivity analysis compared in figure 4.4 the sensitivity analysis with +3% parameter perturbation presented below reads: + for increased sensitivity of an enzyme level to a metabolite, when perturbed. +↑ for high levels of sensitivity of an enzyme level to a metabolite and vice versa (sensitivity when enzyme levels perturbed positively), “.” : insignificant or no effect of enzyme levels to desired metabolite.

<table>
<thead>
<tr>
<th>Maximization of</th>
<th>Substrate</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R_{\text{ext}}</th>
<th>R4</th>
<th>Biomass</th>
<th>Desired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass</td>
<td>-36.5</td>
<td>36.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36.5</td>
<td>0</td>
</tr>
<tr>
<td>Desired</td>
<td>-36.5</td>
<td>0</td>
<td>18.25</td>
<td>18.25</td>
<td>0</td>
<td>18.25</td>
<td>18.25</td>
<td>18.25</td>
</tr>
<tr>
<td>R_{\text{ext}}</td>
<td>-36.5</td>
<td>0</td>
<td>18.25</td>
<td>18.25</td>
<td>18.25</td>
<td>0</td>
<td>18.25</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolite</td>
</tr>
<tr>
<td>Biomass</td>
</tr>
<tr>
<td>Desired</td>
</tr>
<tr>
<td>R_{\text{ext}}</td>
</tr>
</tbody>
</table>
In case of ‘desired’ or M5_ext production, R_ext and R1 are the reactions where FBA calculates the flux distribution as zero or these two reactions are needed to be eliminated. Sensitivity analysis for M5_ext in figure 4.4 shows that decreasing the enzyme levels of ‘R_ext’ and ‘R1’ accounts increasing M5_ext. Also for maximization of ‘R_ext’ or M4_ext, FBA predicts that R1 and R4 needs to be deleted, which is supported by sensitivity analysis for ‘R_ext’ or M4_ext, which gives that the enzyme levels for ‘R1’ and ‘R4’ if decreased accounts increasing M4_ext, or conversely enzyme levels for ‘R1’ and ‘R4’ if increased accounts decreasing M4_ext.

Sensitivity analysis is performed in a kinetic model while FBA is limited to stoichiometric models. A key benefit of FBA is that it requires minimal amount of biological knowledge and data required to make quantitative predictions. However, FBA concentrates only on flux distribution and not on cellular metabolite concentration. When the time course data of the network intermediates are not known and the network is large, then stoichiometric models are selected to get flux distribution at steady state condition using FBA. However, in this work, considerable amount of data from the network intermediates are known thus, kinetic modeling is favored over stoichiometric modeling. The flux distribution, which FBA returns can also be calculated using a kinetic model where it calculates the flux distribution during a time course.

Glycolysis in L. lactis was also modeled as a stoichiometric model. However, the tools used, like OptFlux and COBRA (Constrained Based Analysis and Reconstruction) toolbox in MATLAB, required more information such as boundary reactions, exchange reactions, metabolite formula in charged and neutral state, total charge etc. and also the tools used showed that the model had reaction gaps in its pathway. Since FBA is applied at steady state, any compound entering the system should always exit. When this does not happen, the tools showed reaction gaps in the pathway which are to be filled to validate the steady state assumption. Since, a considerable amount of data, network structure and kinetic parameters were already available, stoichiometric modeling is now left out and only kinetic modeling is focused.

4.2 Simulation of Glycolysis within BST framework

With data available for network intermediates, kinetic models are then chosen over stoichiometric models to study. Glycolysis in L. lactis was simulated using S-systems of kinetics given by the rate equations in chapter B equation (B.1) and using parameters from table B.1 for aerobic conditions and table B.2 for anaerobic conditions. The characteristic phenotype of L. lactis is observed via this modeling approach [64]. The role of regulators FBP and P are studied on PYK reaction that governs the changes in PEP concentration which serves as
Chapter 4. Results and Discussion

driving force for glucose uptake in \( L.\text{ lactis} \). The response of the metabolites are given as:

![Image of time course dynamics of glycolytic model in L. lactis](image)

**Figure 4.5:** Time course dynamics of glycolytic model in \( L.\text{ lactis} \). (Aerobic conditions left, anaerobic conditions right). The legends of each figure is given as: • : glucose measured, — : glucose simulated, * : FBP measured, * : FBP simulated, • : P3GA measured, — : P3GA simulated, • : PEP measured, — : PEP simulated, • : lactate measured, — : lactate simulated.

Any available glucose is taken up by the cell causing a short accumulation of all the metabolites. Also, unneeded intermediates are undesirable in a system and are to be minimized. However, the accumulation of PEP allows the reaction ‘\( \text{PTS:glucose: phospho-tranferase} \)’, that phosphorylates glucose for G6P production, initiating glycolysis, which is also a characteristic of starved cells. This phenomena also rises a question that how could cells maintain high PEP concentration during glucose starvation. This can be explained by regulation as hypothesized in [25]. FBP is a strong activator of ‘\( \text{PYK: Pyruvate Kinase} \)’. When glycolysis proceeds and FBP declines while \( P_i \) is increasing, \( \text{PYK} \) gets deactivated and PEP cannot be converted to Pyruvate. Glucose transport system is also slowed and PEP along with P3GA during this phase switch within themselves such that high noise is observed. To assess the mechanism of start and stop of glycolysis, the toy model in figure. [C1] from [25] is used.

**Model mimicking Glycolysis**

Initially, the system starts at steady state. After \( t = 10 \), \( X_4 \) the main input is removed and the system behaves such that only \( X_4 \) is produced and not degraded. Since \( X_2 \) is close to zero, the activation by \( X_2 \) stops. Because of the model simplicity, it deviates from the \( L.\text{ lactis} \) model in terms of PEP depletion at \( t = 60 \). When \( t = 60 \), the input is restored and the system resumes its activity.

The control of FBP or \( X_2 \) upon PEP degradation is observed by removing the activation of FBP, setting \( h_{42} = 0 \). Here, it is observed that \( X_4 \) is decreased along the other metabolites initially. After time 60, the low amount of PEP takes up substrate very slowly. Following
figures illustrates the explanation.

![Diagram 1](image1)
![Diagram 2](image2)

**Figure 4.6:** Response of simplified glycolytic pathway in *L. lactis* from figure C.1

When $X_6$ and $h_{52}$ are added to the system, following are the behavior seen:

![Diagram 3](image3)
![Diagram 4](image4)

**Figure 4.7:** Response of mimicked glycolytic pathway in figure C.1, with one extra species added. The three figures are referred as A, B and C from left and are explained in text with the conditions explained below.

In figure A, initially, $h_{52}$ is assigned a value of 0.5. In the beginning, the effect of $h_{52}$ on production of $X_6$ is observed. When the system starts, the time course profile of each species starts rising. At time 10 to 60 units, the regulation of FBP for ‘LDH: lactate dehydrogenase’ reaction ($h_{52}$) is removed. At 10 units of time, it is observed that the production of $X_6$ declines quite abruptly and rises very slowly. After 60 time units, $h_{52}$ is restored to 0.5 again. During this period, it can be seen that $X_6$ production rises sharply till the production reaches to a steady state.

In figure B, second figure from left, $h_{52}$ is assigned a value of 0.5. The system starts production of the intermediates unless at time 10 units, $h_{42}$ and $h_{52}$ are removed, and get restored to their original values at time 60 units. The production elevates steeply till time 60. After 60 units of time, $X_6$ production is increased vertically because of the introduction of activators $h_{42}$ and $h_{52}$. When the activators are restored to their original numbers, then a steep rise in the time course profile is seen for all species, especially for $X_5$ and $X_6$, which later declines with a same pattern and goes to a steady state.
Figure C, third figure from left, is similar to the first/left figure of figure 4.6. Here, initially there is no activation for $X_6$ production. At 10 units of time, the second input is shut off which is restored after 60 units of time. The activator for $X_5$ production is then removed after 60 units of time i.e. $h_{42} = 0$. A steady state is observed in the beginning, which then changes such that $X_4$ production reaches to a maximum level and starts decreasing very slowly. After time 60, although the second input is restored the production ceases since there are no activators in the system.

For the exact inference of regulation, the glycolytic system from [25], [64] is studied using a graphical user interface for glycolysis in $L. lactis$ in aerobic conditions, called GUI-SIMGLY [67]. Also, in the simplified model, a inhibitor $p_i$ for PK is left out which comes in play in the real network. Four key parameters in PK reaction: $\beta_{51}, h_{513}, h_{515}, h_{51p_i}$ that accounts for rate constant, FBP activation, enzyme level and $p_i$ inhibition are changed as follows:

| Table 4.5: Parameters changed for FBP and P regulations |
|-----------------|-----------------|-----------------|-----------------|
| $\beta_{51}$ : rate | $h_{513} : FBP+$ | $h_{515} : PK$ | $h_{51p_i} : P_i$ |
| No Regulation_a | 15.3 | 0 | 0.0382342 | 0 |
| No Regulation_b | 15.3 | 0 | 0 | 0 |
| Excl. inhibition_a | 8 | 0 | 0.0382342 | 0.211149 |
| Excl. inhibition_b | 8 | 0 | 0 | 0.211149 |
| Excl. activation_a | 1.1 | 0.7663 | 0.0382342 | 0 |
| Excl. activation_b | 1.1 | 0.7663 | 0 | 0 |

It is seen in figure 4.8 that in first two cases, i.e. no regulation on PYK and exclusive inhibition by $p_i$ on PYK, the P3GA PEP pool at first are steeply consumed, then gets produced extremely fast, reaching to a maximum level and as the glucose starts depleting, the PGA PEP again decreases steeply, which could be interpreted as PEP is being used for glucose transport system in the beginning. This produces FBP, which does not now activate PYK and due to inhibition by $p_i$ or without any activation, starts ceasing the PYK reaction. PEP declines and shuts off after a while thereby not allowing the cell to uptake glucose. On the other hand, when there is exclusive activation by FBP, then PGA PEP pool first depletes steeply and gets accumulated and then decreases very slowly, confirming sufficient PEP pool for future glucose utilization.
Figure 4.8: Effect of different regulations on PYK. First row: no regulation on PYK a(left), b(right), second row: exclusive inhibition by P<sub>i</sub> on PYK a(left), b(right), third row: exclusive activation by FBP on PYK a(left), b(right). Table 4.5 refers a and b. The legends of each figure is given as: •: glucose measured, ---: glucose simulated, •: FBP measured, ---: FBP simulated, •: P3GA measured, ---: P3GA simulated, •: PEP measured, ---: PEP simulated, •: lactate measured, ---: lactate simulated.

The regulation by FBP takes place accompanied by phosphate inhibition. Without FBP, this regulation is said to be sensitive to P<sub>i</sub> fluctuations in the cell, affecting glycolysis via glucose transport reaction by ATP, which is activated by ADP. The large transient pool of FBP may be interpreted as even if there are large number of P<sub>i</sub> in the cell, then FBP would activate formation of pyruvate.
Chapter 4. Results and Discussion

The smoothing and interpolation of the data while simulating glycolysis, explained in section 3.4 of which the characteristics are shown in figure 4.5 and 4.8 was not possible in the tools used such as cellDesigner or COPASI. The parameters that takes place the powers of the rate equations in S-system modeling could not be estimated using COPASI despite of the possibility to replicate time dependency of a metabolite in these tools. Other tools for parameter estimation purpose were not used in this work. All analysis in this section are performed in MATLAB.

4.3 Approximated vs. semi-mechanistic kinetics

Two different systems of modeling approaches with different rate equations, GMA and convenience kinetics are compared neck to neck for the model in [25] given in figure 8.1. The motivation here is to get supportive evidence for a type of modeling approach that overcasts the other in terms of fitting and predictions in order to extend the model to incorporate acetoin and butanediol during glycolysis in L. lactis. After the model is validated, control points in the network are identified using sensitivity analysis that are involved in acetoin and butanediol production.

4.3.1 Model construction and parameter estimation

The models were constructed in cellDesigner with the rate equations given in appendix D. The parameters used are given in table 4.6 and table 4.7. With the available experimental data [52], of anaerobic glucose consumption, the parameters are estimated in COPASI using particle swarm algorithm. The following are the dynamics of the systems after the parameters are estimated with 40 mM glucose pulse utilization data in anaerobic conditions:

Figure 4.9: Fitting of the model with 40 mM glucose impulse at time zero. The legends of each figure is given as: ○: data, ---: convenience kinetics, —-: GMA system
4.3. approximated vs. semi-mechanistic kinetics

Table 4.6 and table 4.7 gives the initial and estimated parameters for convenience kinetics and GMA system, respectively.

**Table 4.6:** Initial and estimated parameters used in the simulation of ODE’s in equation (D.2) for anaerobic conditions, using convenience kinetics equations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial</th>
<th>Estimated</th>
<th>Parameter</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{\text{max}}^{r_e} ) GLUC</td>
<td>1000</td>
<td>999.974</td>
<td>( k_{\text{M}}^{p5 \text{GA}} )</td>
<td>0.4</td>
<td>0.00155238</td>
</tr>
<tr>
<td>( k_{\text{re}}^{1} ) G6P</td>
<td>0.5</td>
<td>27.0065</td>
<td>( V_{\text{max}}^{r_e} )</td>
<td>0.5</td>
<td>25.8279</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{PEP}} )</td>
<td>0.5</td>
<td>0.28782</td>
<td>( k_{\text{re}}^{4} ) PEP</td>
<td>0.5</td>
<td>42.791</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{PYR}} )</td>
<td>10</td>
<td>933.699</td>
<td>( k_{\text{M}}^{p3 \text{GA}} )</td>
<td>0.8</td>
<td>0.0212757</td>
</tr>
<tr>
<td>( V_{\text{max}}^{r_e} ) G6P</td>
<td>100</td>
<td>997.025</td>
<td>( k_{\text{M}}^{\text{PYR}} )</td>
<td>10</td>
<td>933.699</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{ATP}} )</td>
<td>0.5</td>
<td>0.176978</td>
<td>( V_{\text{max}}^{r_e} )</td>
<td>0.8</td>
<td>89.5565</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{FBP}} )</td>
<td>4</td>
<td>912.073</td>
<td>( k_{\text{M}}^{\text{GLUC}} )</td>
<td>0.5</td>
<td>27.0065</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{ADP}} )</td>
<td>1</td>
<td>2.55299</td>
<td>( k_{\text{M}}^{\text{PEP}} )</td>
<td>0.3</td>
<td>96.4009</td>
</tr>
<tr>
<td>( V_{\text{max}}^{r_e} )</td>
<td>0.3</td>
<td>13.129</td>
<td>( k_{\text{M}}^{\text{FBP}} )</td>
<td>2</td>
<td>729.239</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{FBP}} )</td>
<td>0.2</td>
<td>49.7098</td>
<td>( V_{\text{max}}^{r_e} )</td>
<td>10</td>
<td>14.2325</td>
</tr>
<tr>
<td>( k_{\text{M}}^{\text{ADP}} )</td>
<td>0.2</td>
<td>35.7922</td>
<td>( k_{\text{M}}^{\text{PYR}} )</td>
<td>0.8</td>
<td>2.617 \times 10^{-7}</td>
</tr>
<tr>
<td>( k_{\text{M}}^{NAD} )</td>
<td>0.3</td>
<td>1.75246</td>
<td>( k_{\text{M}}^{\text{NADH}} )</td>
<td>0.2</td>
<td>10.2408</td>
</tr>
</tbody>
</table>

**Table 4.7:** Parameters used in the simulation of ODE’s in equation (D.1) for anaerobic conditions, using GMA system of equations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial</th>
<th>Estimated</th>
<th>Parameter</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_{\text{ass}}^{1} )</td>
<td>1</td>
<td>434.751</td>
<td>( k_{\text{ass}}^{r_e} )</td>
<td>0.1</td>
<td>29.4374</td>
</tr>
<tr>
<td>( k_{\text{ass}}^{2} )</td>
<td>0.75</td>
<td>72.2652</td>
<td>( k_{\text{ass}}^{r_e} )</td>
<td>2</td>
<td>151.426</td>
</tr>
<tr>
<td>( k_{\text{ass}}^{3} )</td>
<td>0.6</td>
<td>0.060175</td>
<td>( k_{\text{ass}}^{r_e} )</td>
<td>0.1</td>
<td>203.026</td>
</tr>
<tr>
<td>( k_{\text{ass}}^{4} )</td>
<td>0.7</td>
<td>39.1418</td>
<td>( k_{\text{ass}}^{r_e} )</td>
<td>2</td>
<td>10^{-9}</td>
</tr>
<tr>
<td>( k_{\text{ass}}^{5} )</td>
<td>0.5</td>
<td>93.6445</td>
<td>( k_{\text{ass}}^{r_e} )</td>
<td>0.5</td>
<td>99.921</td>
</tr>
</tbody>
</table>

The model is now validated to assess its prediction capability.

4.3.2 Model validation

Validation is a process where a model is assessed how predictive it is by changing the initial condition of the model input and checking the simulation against experimental data. Here, the glucose impulse is changed to 80 mM and following predictions were made:
Chapter 4. Results and Discussion

Figure 4.10: Validation of the model with 80 mM glucose impulse at time zero. The legends of each figure is given as: •: data, ---: convenience kinetics, ---: GMA system.

It is seen that while fixing the concentration of $NAD = 4.21$, $NADH = 3$, $ATP = 1$, $ADP = 5$ and $P = 1$, the dynamics of the metabolites in the pathway is defined properly, which was not the case when these metabolites were considered as state variables (results not shown). Also, it is observed that the validation of the model follows the experimental data closely. The AIC$_c$ (eqn (3.4)) gives: $\text{Convenience kinetics} = 339.337$; $\text{GMA} = 905.928$, which concludes that while taking care of the network topology and accounting for each variables that contributes to the involvement in other pathways, convenience kinetics equations describes a model better than GMA system of equations in terms of validation and AIC$_c$. In cases where the GMA out rules the convenience kinetics, we still can argue that the dynamics of the metabolites are very poorly described and GMA systems have very few number of parameters involved.

4.4 Model extension

It is observed from section 4.3 that convenience kinetics describes better the dynamics of a system with respect to time. An anaerobic glycolysis model as shown in figure 4.11 is reconstructed as explained in methods chapter, section 3.5.2. The parameters that were not known in literature reviews were obtained from [66], which then gave good behavior of the organism’s phenotype. The model topology obtained is given as:
Figure 4.11: Network structure extended and reconstructed in this work [51], [26], [65]. The blue lines are the inhibitions and the red lines are the activation. The abbreviations corresponds to the reactions, which are given in appendix E.

Parameter estimation and model validation

After the model has been constructed, it is further improved by estimating parameters using $^{13}$C NMR in vivo time course data. While modeling different organism or collecting
the parameters, most of the times, the experimental kinetics are determined in isolated enzyme kinetic experiments, or from different experiment and even sometimes from different organism, which leads to inaccuracies in the model and incorrect predictions. To make an accurate kinetic model, the kinetic parameters should be obtained in conditions that simulate in vivo environment.

The dynamics of the system after estimating the parameters are given in figure 4.12. With a glucose impulse of 40 mM at time zero, the model is trained with available experimental data available from Neves et. al. 2005, [52]. Ten different runs of parameter estimation were performed to get significant results. It can be seen that the best fitted model (blue curve) in figure 4.12 follows the experimental data well except in the case of FBP, where it does not produces a bell shaped peak. Since the model is trained using 40 mM data, inferring any conclusions without validating will give uncertain results. Thus, the model is now further validated changing the glucose impulse (input).

Figure 4.12: Fittings of the reconstructed model with 40 mM glucose impulse at time zero with evolutionary programming in COPASI. The legends of each figure is given as: ○: data, - - : simulation, —— : best fit.
The trained model is now validated using 80 mM glucose impulse at time zero. When comparing with the predictions made by the model against the experimental data, as given in figure 4.13. It can be observed that the response of the model for the changed glucose impulse (input) follows the experimental data closely. When the predictions for FBP, ATP and P are observed, it can be said that some aspect in the models are lacking which leads to the divergence of model predictions from experimental data. Considering that ATP and P are involved in entire metabolic network and not only limited to the glycolytic model studied here, it can be argued that this model provides a sufficient base for analyzing the control points or reactions responsible for our desired metabolites, i.e. acetoin and butanediol.

Figure 4.13: Validation of the reconstructed model with 80 mM glucose impulse at time zero. The legends of each figure is given as: ○: data, −: validation, −−: best validation.

After a good predictive model in terms of validation, fitting and its overall characteristics
is obtained, it is later used for metabolic engineering purposes. The rate equations, initial concentrations, estimated concentrations, initial parameters and estimated parameters used while extending the model are given in appendix E.

4.5 Sensitivity analysis in extended model

For the extended model, the sensitivity analysis was performed as explained in the methods chapter, section 3.3. Here, only the sensitivity of the enzyme levels ($V_{max}$) in the reactions for acetoin and butanediol production are assessed. The $RD$ values were calculated using equation (3.10) and plotted as bar graphs. The analysis is performed in MATLAB by perturbing the parameter set and solving the system each time when parameter(s) are perturbed. The implementation and source code of which are given in methods chapter, section 3.3 and appendix F respectively.

4.5.1 Single parameter perturbation

Each enzyme level was perturbed at a time and the system was solved. Every time the system is solved, the shift or change in the area between the curve in reference to wild type strain given by a metabolite’s time course is calculated.

4.5.1.A Analysis for Acetoin

The enzyme levels are perturbed by $\pm 3\%$ and $\pm 1\%$ (results not shown) and its effect on the desired metabolites is assessed using sensitivity analysis. A positive sensitivity reads perturbing the parameter by some $x$ percentage will give a rise in the desired metabolite’s production and vice versa.

Figure 4.14: Sensitivity analysis of the reconstructed model for acetoin. (Parameter perturbation of $+3\%$ : left, $-3\%$ : right)

Here it is seen that two enzymes, phosphotranferase: glucose ($PTS:glucose$) and lactate dehydrogenase ($LDH$) has a significant effect on the system. Increasing $PTS:glucose$ level 1%
or 3% gives a rise in acetoin production while increasing LDH decreases the acetoin production. The enzyme levels of reactions pyruvate kinase (PYK), acetolactate synthase; acetolactate decarboxylase (PA), acetoin transportase (AT), mannitol phosphatase (MP), FBP\textsubscript{ase}, ATP\textsubscript{ase} influences positively to the system when the parameters are increased. The enzyme levels of phosphofructokinase (PFK), LDH, pyruvate dehydrogenase (PDH) and acetate kinase (AC) has a negative influence on the system when the parameters are increased. It is also observed that phosphoglucoisomerase (PGI) has a negative influence when it is increased more since when it was perturbed by 1%, the influence was not seen and when perturbed by 3%, the influence although little, is seen. Similar but opposite inference is drawn when the parameters are perturbed negatively.

### 4.5.1. B Analysis for Butanediol

![Figure 4.15: Sensitivity analysis of the reconstructed model for butanediol. (Parameter perturbation of +3% : left, −3% : right)](image)

When observing the above plots, it is seen that more enzyme levels have a negative influence in the production of butanediol than acetoin. Enzyme levels of \textit{PTS: glucose, PYK, PA, butanediol dehydrogenase (AB), AC, mannitol phosphate dehydrogenase (MPD), MP, phosphotransferase: mannitol (pts:Mannitol), ATP\textsubscript{ase}, mannitol transportase (MT)} has a positive influence on production of butanediol. \textit{PGI} influence is similar as in acetoin’s sensitivity analysis where it influence the system only when large perturbation is done in the system’s parameter. \textit{PFK, LDH and PDH} are the common enzymes whose levels has negative influence on the system. Besides these, glyceraldehyde 3-phosphate dehydrogenase (\textit{GAPDH}), \textit{AT, acetaldehyde dehydrogenase; alcohol dehydrogenase (AE)} and \textit{FBP\textsubscript{ase}} have a negative influence in the production of butanediol.

Table 4.8 and 4.9 summarizes main results of single parameter perturbation.
Table 4.8: RD values for Acetoin with +3% of single perturbation. Several other enzyme levels’ sensitivity are not shown. Refer to figure 4.14, +3% parameter perturbation. The reactions are arranged accordingly to decreasing sensitivity effect to the system.

<table>
<thead>
<tr>
<th>Reactions</th>
<th>Sensitivity</th>
<th>Reactions</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDH</td>
<td>-0.0798</td>
<td>FBP&lt;sub&gt;ase&lt;/sub&gt;</td>
<td>0.0061</td>
</tr>
<tr>
<td>PTS: Glucose</td>
<td>0.0676</td>
<td>ATP&lt;sub&gt;ase&lt;/sub&gt;</td>
<td>0.0045</td>
</tr>
<tr>
<td>PA</td>
<td>0.0197</td>
<td>AB</td>
<td>0.0033</td>
</tr>
<tr>
<td>PDH</td>
<td>-0.0174</td>
<td>AT</td>
<td>0.0033</td>
</tr>
<tr>
<td>MP</td>
<td>0.0118</td>
<td>PYK</td>
<td>0.0023</td>
</tr>
<tr>
<td>PFK</td>
<td>-0.0119</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: RD values for Butanediol with +3% of single perturbation. Several other enzyme levels’ sensitivity are not shown. Refer to figure 4.15, +3% parameter perturbation. The reactions are arranged accordingly to decreasing sensitivity effect to the system.

<table>
<thead>
<tr>
<th>Reactions</th>
<th>Sensitivity</th>
<th>Reactions</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDH</td>
<td>-0.0894</td>
<td>MP</td>
<td>0.009</td>
</tr>
<tr>
<td>PTS: Glucose</td>
<td>0.0819</td>
<td>ATP&lt;sub&gt;ase&lt;/sub&gt;</td>
<td>0.0087</td>
</tr>
<tr>
<td>PA</td>
<td>0.0247</td>
<td>FBP&lt;sub&gt;ase&lt;/sub&gt;</td>
<td>-0.0128</td>
</tr>
<tr>
<td>PDH</td>
<td>-0.0266</td>
<td>PK</td>
<td>0.0047</td>
</tr>
<tr>
<td>AT</td>
<td>-0.0175</td>
<td>AE</td>
<td>-0.0032</td>
</tr>
<tr>
<td>PFK</td>
<td>-0.0092</td>
<td>AB</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AC</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

4.5.2 Double parameter perturbation

Two parameters, from a set of enzyme levels were perturbed and its effect on the production of desired metabolites: acetoin and butanediol is assessed. All possible combination of two parameters from 21 set of enzyme levels were seen. The idea is similar to the single parameter perturbation, the step where it differs is such that at each iteration of solving the system of ODE’s, two parameters were perturbed. The computational implementation is given in appendix F.
4.5. Sensitivity analysis in extended model

4.5.2.A Analysis for acetoin

Figure 4.16: Sensitivity analysis of the reconstructed model as color bars for acetoin with parameter perturbation of +3% (left), −3% (right)

When the enzyme levels are perturbed positively, the combination of enzyme levels phosphate transportase \((PT)\) along with acetalactate synthase; acetalactate decarboxylase \((PA)\) and mannitol phosphatase \((MP)\); \(PA\) along with \(MP\), \(FBPase\), \(ATPase\) and Pyruvate kinase \((PYK)\); and \(PTS: glucose\) along with all enzyme levels except lactate dehydrogenase \((LDH)\) exerts a significant effect on production of acetoin. Increasing these enzyme levels in \(L. lactis\) will account for enhanced production of acetoin. Table 4.10 presents the set of enzymes with significant effect on production of acetoin.

When the enzyme levels are perturbed negatively, the combinations of enzyme levels pyruvate dehydrogenase \((PDH)\) along with phosphofructokinase \((PFK)\), and \(LDH\) along with all enzyme levels except \(PTS:gluc\) have a significant effect on production of acetoin. Decreasing these enzyme levels will result in enhanced production of acetoin.
4.5.2.B Analysis for butanediol

Figure 4.17: Sensitivity analysis of the reconstructed model as color bars for butanediol with parameter perturbation of $+3\%$ (right), $-3\%$ (left)

When the enzyme levels are perturbed positively, the combination of enzymes levels acetolactate synthase; acetolactate decarboxylase (PA) along with mannitol phosphatase (MP) and $ATP_{asc}$; and phosphotransferase: glucose ($PTS:gluc$) along with all enzyme levels except $LDH$ has a significant effect on butanediol production. Increasing these enzyme levels in $L.\ lactis$ will account for enhanced production of butanediol. Table 4.10 presents the set of enzymes with significant effect on production of butanediol.

When the enzyme levels are perturbed negatively, the combination of enzyme levels pyruvate dehydrogenase (PDH) along with Acetaldehyde dehydrogenase; alcohol dehydrogenase (AE) and $FBP_{asc}$; and $LDH$ along with all enzyme levels except $PTS:gluc$ exerts an increased effect on production of butanediol.

Table 4.10 summarizes the main results of double parameter perturbation.

**Table 4.10:** Enzyme levels with significant effect for acetoin and butanediol production with $+3\%$ of double perturbation. Several other enzyme levels’ sensitivity are not shown. Refer to figure 4.16 and figure 4.17 for the plots. The reactions are arranged accordingly to decreasing sensitivity effect to the system.

<table>
<thead>
<tr>
<th>Sensitivity for Acetoin</th>
<th>Sensitivity for Butanediol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme levels</td>
<td>Enzyme levels</td>
</tr>
<tr>
<td>$PTS$: Glucose; all except LDH</td>
<td>$PTS$: Glucose; all except LDH</td>
</tr>
<tr>
<td>PT; MP</td>
<td>PA; PYK</td>
</tr>
<tr>
<td>PT; $FBP_{asc}$</td>
<td>PA; AB</td>
</tr>
<tr>
<td>PT; $ATP_{asc}$</td>
<td>PA; AC</td>
</tr>
<tr>
<td>PA; PYK</td>
<td>PA; MP</td>
</tr>
<tr>
<td>PA; AT</td>
<td>PA; $ATP_{asc}$</td>
</tr>
<tr>
<td>PA; AB</td>
<td></td>
</tr>
<tr>
<td>PA; MP</td>
<td></td>
</tr>
<tr>
<td>PA; $FBP_{asc}$</td>
<td></td>
</tr>
<tr>
<td>PA; $ATP_{asc}$</td>
<td></td>
</tr>
</tbody>
</table>
4.5. Sensitivity analysis in extended model

With a kinetic model available for the pathway of interest and performing sensitivity analysis in these models elucidates the significance of each enzyme levels responsible for acetoin and butanediol production. This model can now be used for metabolic engineering tool to design wet lab experiments. The reactions or the enzyme levels that showed limelight for acetoin and butanediol production corresponds with the previous experimental evidence where deleting the LDH reaction increased the acetoin production \cite{27} and decreasing the enzyme level responsible for ethanol production: alcohol dehydrogenase designated as AE accounted for decreasing butanediol levels in *Klebsiella oxytoca* \cite{68}, \cite{69}. Here, in case of *L. lactis*, increasing $V_{max}$ of AE accounted for decreasing butanediol levels and vice versa but with lesser extent.
Conclusions and Future work
The behavior of a metabolic system can be replicated in silico, where the behavior governed by enzyme catalyzed reactions are brought up to a mathematical model using the kinetic laws governing each reaction and formulating ODEs with these rate laws or using the information from stoichiometry of metabolites and reaction flux. The work here began with analysis of a toy stoichiometric model using an optimization strategy known as FBA. FBA predicted the flux of a reaction given an objective to maximize and fluxes constraining the reactions. Constraining the substrate uptake rate allowed over expression or under expression of reaction fluxes. The same network was then translated to a kinetic model using Michaelis-Menten kinetics and sensitivity analysis was performed. The results from sensitivity analysis and FBA were compared. Both analysis gave similar results when interpreted. When only the information about the stoichiometry and network topology of central metabolism in L. lactis were used to construct a stoichiometric model, the construction of stoichiometric model was yet not feasible to use FBA with the tools on hand because of the reaction gaps in the pathway structure. With kinetic model, the flux of each reaction as well as time course concentration profiles can be known while simulating the model.

After it was concluded that kinetic models are to be used ahead and not stoichiometric models, glycolysis in L. lactis was simulated and the characteristics were studied from Voit et al., 2006 [25] and Vinga et al., 2010 [64] using S-systems modeling. After concluding that S-system approach was not feasible to use with the tools available at hand, a simple modeling approach (GMA) in the context of number of parameters and a mechanism based model (convenience kinetics) which is complicated than GMA model were used to compare between the modeling approach. In a reversible reaction, with one species producing one product, the minimum number of parameters GMA requires is 2 while convenience kinetics takes 4 parameters. The model was constructed using known network topology from previous published work [25]. With some adjustments such as keeping the metabolites that are involved in other metabolic network fixed in the model (NAD, NADH, ATP and P), and omitting the time dependency of glucose decay, it is inferred that the GMA models are suppressed by convenience kinetics model in terms of model predictions and AICc.

Once it is inferred that convenience kinetics models are better than GMA models, an anaerobic glycolytic model in L. lactis was reconstructed. This model is now used to gain insights into the glycolysis pathway. Using Sensitivity analysis, the enzyme levels that are responsible for the production of desired compounds or metabolites; acetoin and butanediol in our case were seen. The analysis could also be used for the production of mannitol or ATP, depending upon the reaction or metabolite of interest.

These information are not available to biologists readily and is problematic when the knowledge regarding the dynamics of the system is not known for metabolic engineering.
To relate metabolomics to genome study, altering the $V_{\text{max}}$ of a reaction in a pathway causes the activity of the enzyme to be upregulated, downregulated or removed. This when cross-referred in genetic perturbations, could be achieved in bacteria by knocking out the gene responsible for the enzyme or by increasing or decreasing the corresponding transcript levels for upregulation and downregulation respectively.

On summarizing, the work presented here discusses the background of LAB, their applications in biotechnology and how their kinetics of metabolic pathway can be utilized to understand the pathway. Different types of modeling approaches towards constructing a kinetic model, exchanging it with different analysis environment and solving the system with different parameter perturbations can be learned from this work explicitly. The model devised here, can be later used for metabolic engineering purposes via engineering the enzyme levels. Also, further analysis such as a combination of different numbers of parameter perturbation can be done to get more insights in the model. However it is also to be noted that the model should be stable in all cases. Multi-enzyme modulation strategies can be devised in the presented model to get a combination of parameters to perturb. This model can also be extended to include other pathways, citric acid cycle in $L. \text{lactis}$ for instance and analyze it for the outputs that it is designated to produce. The significance of the binding constants ($k_M$) values could also be studied using sensitivity analysis.

A complete kinetic model of a cell would give rise to very deep insights into the reaction network and model predictions. In addition if data is available for all or most of the intermediate metabolites in a model, a kinetic model would replicate the strain or organism (considering level of details in model) in detail. These tools and methods that are discussed here to construct and analyze metabolic models can be used to create $\text{in silico}$ mutant strains, and then to create a real mutant strain designing the experiments on the basis of model. These strategies can be used extensively in biotechnological applications like food industries, fermentation industries, pharmaceutical industries, flavanoids production etc. The modeling approaches presented can be used on a new organism and even the model reconstructed can serve as a starting point to model other LAB such as $\text{Streptococcus pneumoniae}$.
Bibliography


Rate equations and initial conditions used while translating stoichiometric model to kinetic model
Appendix A. Rate equations and initial conditions used while translating stoichiometric model to kinetic model

Rate equations

The rate equations used to translate the stoichiometric model from results chapter, section 4.1, figure 4.1 are given as:

\[ v_{\text{substrate}} = \frac{V_{\text{sub}} \left[ M1_{\text{ext}} \right]}{k_{\text{sub}} M1_{\text{ext}} + \left[ M1_{\text{ext}} \right]} \]  
\[ v_{R1} = \frac{V_{R1} \left[ M1 \right]}{k_{R1} M1 + \left[ M1 \right]} \]  
\[ v_{R2} = \frac{V_{R2} \left( \left[ M1 \right] \frac{k_{R2}}{k_{M1}} \right)^2}{1 + \frac{\left[ M1 \right]}{k_{M1}} + \left( \frac{\left[ M1 \right]}{k_{M1}} \right)^2} \]  
\[ v_{R3} = \frac{V_{R3} \left[ M3 \right]}{k_{R3} M3 + \left[ M3 \right]} \]  
\[ v_{R4} = \frac{V_{R4} \left[ M4 \right]}{k_{R4} M4 + \left[ M4 \right]} \]  
\[ v_{\text{desired}} = \frac{V_{\text{des}} \left[ M5 \right]}{k_{\text{des}} M5 + \left[ M5 \right]} \]  
\[ v_{\text{biomass}} = \frac{V_{\text{biom}} \left[ M2 \right]}{k_{\text{biom}} M2 + \left[ M2 \right]} \]  
\[ v_{R_{ext}} = \frac{V_{R_{ext}} \left[ M4 \right]}{k_{R_{ext}} M4 + \left[ M4 \right]} \]

Parameters

The parameters used to simulate the model are given as:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{sub}}_{\text{max}}$</td>
<td>100</td>
<td>$k_{\text{sub}} M1_{\text{ext}}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$V_{R1}_{\text{max}}$</td>
<td>25</td>
<td>$k_{R1} M1$</td>
<td>1</td>
</tr>
<tr>
<td>$V_{R2}_{\text{max}}$</td>
<td>25</td>
<td>$k_{R2} M1$</td>
<td>1</td>
</tr>
<tr>
<td>$V_{R3}_{\text{max}}$</td>
<td>25</td>
<td>$k_{R3} M3$</td>
<td>1</td>
</tr>
<tr>
<td>$V_{R4}_{\text{max}}$</td>
<td>2</td>
<td>$k_{R4} M4$</td>
<td>1</td>
</tr>
<tr>
<td>$V_{\text{des}}_{\text{max}}$</td>
<td>60.8</td>
<td>$k_{\text{des}} M5$</td>
<td>7.88</td>
</tr>
<tr>
<td>$V_{\text{biom}}_{\text{max}}$</td>
<td>72.2</td>
<td>$k_{\text{biom}} M2$</td>
<td>1</td>
</tr>
</tbody>
</table>

The initial concentration of $M1_{\text{ext}}$ is kept 40, while others are set to zero.
Network structure, rate equations and parameters for simulation of glycolysis within BST
Appendix B. Network structure, rate equations and parameters for simulation of glycolysis within BST

The network structure for simulation procedure in chapter 3, section 3.4 is given as:

![Glycolysis Network Structure](image)

Figure B.1: Glycolysis in *L. lactis* figure from [25]

The kinetic equations of each state variables here are given as:

\[
\begin{align*}
\frac{d}{dt} G &= -k(1 + \alpha t^\beta) G, \\
\frac{d}{dt} G_6P &= \beta_1 G_6P^{h_{11}} G_6P^{h_{12}} PEP^{h_{25}} - \beta_2 G_6P^{h_{12}} ATP^{h_{2ATP}}, \\
\frac{d}{dt} FBP &= \beta_2 G_6P^{h_{2}} ATP^{h_{2ATP}} - \beta_3 FBP^{h_{33}} P_i^{h_{33}} NAD^{h_{SNAD}}, \\
\frac{d}{dt} PGAPEP &= 2\beta_3 FBP^{h_{33}} P_i^{h_{33P}} NAD^{h_{SNAD}} - \beta_1 G_6P^{h_{11}} G_6P^{h_{12}} PEP^{h_{25}} \\
&\quad - \beta_51 FBP^{h_{513}} P_i^{h_{513P}} P_i^{h_{513P}} - \beta_52 PEP^{h_{525}}, \\
\frac{d}{dt} P &= \beta_1 G_6P^{h_{11}} G_6P^{h_{12}} PEP^{h_{25}} + \beta_51 FBP^{h_{513}} P_i^{h_{513P}} P_i^{h_{513P}} \\
&\quad - \beta_61 P^{h_{613}} FBP^{h_{613}} NAD^{h_{SNAD}} - \beta_62 P^{h_{626}}, \\
\frac{d}{dt} L &= \beta_61 P^{h_{613}} FBP^{h_{613}} NAD^{h_{SNAD}}
\end{align*}
\]
The parameters of each reaction are given as:

1. PTS (Phosphotransferase):
   Glucose + PEP $\rightarrow$ G6P + PYR
   parameters:
   $k, \alpha, \beta, \beta_{11}, h_{11}, h_{12}$.

2. PFK (Phosphofructokinase): G6P + ATP $\rightarrow$ ADP + FBP
   parameters:
   "$h_{22} : PFK", h_{2ATP}$

3. Enzyme (PGAPEP are two metabolites which are plumped here):
   FBP + 2 NAD + 2 ADP +2 $P_i$ $\rightarrow$ 2(3-PGAPEP) + 2 NADH + 2 ATP
   parameters:
   $\beta_3, "h_{33} : enzyme", h_{3P}, h_{3NAD}$

4. 2 PEP $\Rightarrow$
   parameters:
   $\beta_{52}, h_{525}$

5. PYK (Pyruvate Kinase):
   2 PEP + 2 ADP $\rightarrow$ 2 ATP + 2 Pyr
   parameters:
   $\beta_{51}, "h_{513} : FBP^+", "h_{515} : PYK", "h_{51P} : P_i^+"$

6. LDH (Lactate dehydrogenase): 2 Pyr + 2 NADH $\rightarrow$ 2 Lactate + 2 NAD
   parameters:
   $\beta_{61}, "h_{616} : LDH", "h_{613} : FBP^+", h_{61NAD}$
Appendix B. Network structure, rate equations and parameters for simulation of glycolysis within BST

Table B.1: Parameters used in the simulation of ODEs in equation (B.1) for aerobic conditions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>0.053025</td>
<td>$h_{2ATP}$</td>
<td>1.51599</td>
<td>$h_{515}$</td>
<td>0.038234</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.041996</td>
<td>$\beta_3$</td>
<td>0.338423</td>
<td>$h_{516}$</td>
<td>0.211149</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2.68092</td>
<td>$h_{33}$</td>
<td>1.09298</td>
<td>$h_{51}$</td>
<td>0.032474</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>7.20321</td>
<td>$h_{3P_i}$</td>
<td>0.258372</td>
<td>$h_{61}$</td>
<td>0.675486</td>
</tr>
<tr>
<td>$h_{11}$</td>
<td>0.997546</td>
<td>$h_{3NAD}$</td>
<td>-0.09666</td>
<td>$h_{613}$</td>
<td>1.03221</td>
</tr>
<tr>
<td>$h_{12}$</td>
<td>-1.48643</td>
<td>$\beta_{52}$</td>
<td>0.134164</td>
<td>$h_{61NAD}$</td>
<td>-0.05194</td>
</tr>
<tr>
<td>$h_{25}$</td>
<td>0.38576</td>
<td>$h_{525}$</td>
<td>0.094045</td>
<td>$\beta_{62}$</td>
<td>1.74742</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.348889</td>
<td>$\beta_{51}$</td>
<td>0.862421</td>
<td>$h_{626}$</td>
<td>1.40312</td>
</tr>
<tr>
<td>$h_{22}$</td>
<td>1.54399</td>
<td>$h_{513}$</td>
<td>0.7663</td>
<td>$\beta_7$</td>
<td>0.0386</td>
</tr>
<tr>
<td>$k_{45}$</td>
<td></td>
<td></td>
<td></td>
<td>$h_{45}$</td>
<td>2.04035</td>
</tr>
</tbody>
</table>

Table B.2: Parameters used in the simulation of ODEs in equation (B.1) for anaerobic conditions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>0.124738</td>
<td>$h_{2ATP}$</td>
<td>0.26203</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.134194</td>
<td>$\beta_3$</td>
<td>0.181082</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2.6674</td>
<td>$h_{33}$</td>
<td>1.02783</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>5.86735</td>
<td>$h_{3P_i}$</td>
<td>0.137778</td>
</tr>
<tr>
<td>$h_{11}$</td>
<td>1.25193</td>
<td>$h_{3NAD}$</td>
<td>0.174484</td>
</tr>
<tr>
<td>$h_{12}$</td>
<td>-1.06961</td>
<td>$\beta_{52}$</td>
<td>0.000447956</td>
</tr>
<tr>
<td>$h_{25}$</td>
<td>0.288687</td>
<td>$h_{525}$</td>
<td>0.103001</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.379794</td>
<td>$\beta_{51}$</td>
<td>0.683548</td>
</tr>
<tr>
<td>$h_{22}$</td>
<td>2.83465</td>
<td>$h_{513}$</td>
<td>0.854603</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{45}$</td>
<td>1.54695</td>
</tr>
</tbody>
</table>

B-4
Network structure and rate equations for model mimicking glycolysis
Appendix C. Network structure and rate equations for model mimicking glycolysis

The toy network, used to mimic glycolysis in methods chapter, section 3.4 is given as:

![Network Diagram]

Figure C.1: Generic linear feedforward activated pathway in which a downstream metabolite ($X_4$) is needed as a second substrate for the first step \(^{25}\).

The model equations are given as:

\[
\begin{align*}
\dot{X}_1 &= \text{Input}_1 + \text{Input}_2 X_4^{0.5} - X_1^{0.5} \\
\dot{X}_2 &= X_1^{0.5} - X_2^{0.75} \\
\dot{X}_3 &= 2X_2^{0.75} - 2X_3^{0.4} \\
\dot{X}_4 &= 2X_3^{0.4} - \text{Input}_2 X_4^{0.5} - X_2^{h_{42}} X_4^{0.5} \\
\dot{X}_5 &= X_2^{h_{42}} X_4^{0.5} - X_5^{0.5}
\end{align*}
\]

(C.1)

Table C.1: Parameter and species used in the simulation of ODEs in equation (C.1).

<table>
<thead>
<tr>
<th>Species</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input1</td>
<td>0.01</td>
<td>h_{42}</td>
<td>0.75</td>
</tr>
<tr>
<td>Input2</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_4</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rate equations and initial conditions used in Approximated vs. Mechanism based modeling approach
Modeling via GMA system

The rate equations used in GMA model are given as:

\[ v_1 = \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{1 + \left( \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{K_{\text{M}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]} \right) + \left( \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{K_{\text{M}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]} \right), \quad (D.1) \]

\[ v_2 = k_{\text{rel}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}], \]

\[ v_3 = k_{\text{rel}}^{\text{FBP}} \cdot [\text{FBP}] \cdot [\text{NAD}]^2 \cdot [\text{ADP}]^2 \cdot [\text{P}]^2, \]

\[ v_4 = k_{\text{rel}}^{\text{P3GA}} \cdot [\text{P3GA}]^2 - k_{\text{rel}}^{\text{PEP}} \cdot [\text{PEP}]^2, \]

\[ v_5 = \frac{k_{\text{rel}}^{\text{PYR}} \cdot [\text{FBP}]}{k_{\text{rel}}^{\text{PYR}} \cdot [\text{FBP}] + [\text{P}]} \cdot \frac{k_{\text{rel}}^{\text{PYR}} \cdot [\text{FBP}] \cdot [\text{P}] \cdot [\text{PEP}] \cdot [\text{ADP}]^2}{k_{\text{rel}}^{\text{PYR}} \cdot [\text{FBP}] \cdot [\text{P}] \cdot [\text{PEP}] \cdot [\text{ADP}]^2}, \]

\[ v_6 = \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{1 + \left( \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{K_{\text{M}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2} \right) \cdot \left( \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{K_{\text{M}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2} \right).} \]

The values of the initial and estimated parameters are given in table 4.7 in results chapter, section 4.3.

Modeling via Convenience kinetics

The rate equations used in convenience kinetics model are given as:

\[ v_1 = \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{1 + \left( \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{K_{\text{M}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]} \right) + \left( \frac{v_{\text{max}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]}{K_{\text{M}}^{\text{GLUC}} \cdot [\text{GLUC}] \cdot [\text{PEP}]} \right), \quad (D.2) \]

\[ v_2 = \frac{v_{\text{max}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}]}{1 + \left( \frac{v_{\text{max}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}]}{K_{\text{M}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}]} \right) + \left( \frac{v_{\text{max}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}]}{K_{\text{M}}^{\text{G6P}} \cdot [\text{G6P}] \cdot [\text{ATP}]} \right), \]

\[ v_3 = \frac{v_{\text{max}}^{\text{FBP}} \cdot [\text{ADP}]}{1 + \left( \frac{v_{\text{max}}^{\text{FBP}} \cdot [\text{ADP}]}{K_{\text{M}}^{\text{FBP}} \cdot [\text{ADP}]} \right) + \left( \frac{v_{\text{max}}^{\text{FBP}} \cdot [\text{ADP}]}{K_{\text{M}}^{\text{FBP}} \cdot [\text{ADP}]} \right), \]

\[ v_4 = \frac{v_{\text{max}}^{\text{P3GA}} \cdot [\text{P3GA}]}{1 + \left( \frac{v_{\text{max}}^{\text{P3GA}} \cdot [\text{P3GA}]}{K_{\text{M}}^{\text{P3GA}} \cdot [\text{P3GA}]} \right) + \left( \frac{v_{\text{max}}^{\text{P3GA}} \cdot [\text{P3GA}]}{K_{\text{M}}^{\text{P3GA}} \cdot [\text{P3GA}]} \right), \]

\[ v_5 = \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{1 + \left( \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{K_{\text{M}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2} \right) \cdot \left( \frac{v_{\text{max}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2}{K_{\text{M}}^{\text{PYR}} \cdot [\text{PYR}]^2 \cdot [\text{NADH}]^2} \right).} \]
The ordinary differential equations are now formulated following mass balance as following:

\[
\begin{align*}
\frac{d[GLUC]}{dt} &= -v1 \\
\frac{d[G6P]}{dt} &= v1 - v2 \\
\frac{d[FBP]}{dt} &= v1 - v3 \\
\frac{d[P3GA]}{dt} &= 2v3 - 2v4 \\
\frac{d[PEP]}{dt} &= 2v4 - 2v5 - v1 \\
\frac{d[PYR]}{dt} &= 2v5 - 2v6 + v1 \\
\frac{d[LAC]}{dt} &= 2v6
\end{align*}
\]  

(D.3)

The initial concentrations and mass balance equation used for both systems are given in table [D.1] and equation (D.3).

Table D.1: Initial concentrations used in the simulation of ODE’s in equation (D.1) and equation (D.2) for anaerobic conditions.

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Concentration</th>
<th>Metabolites</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>40</td>
<td>PEP</td>
<td>14.7628</td>
</tr>
<tr>
<td>G6P</td>
<td>0</td>
<td>PYR</td>
<td>0.1</td>
</tr>
<tr>
<td>FBP</td>
<td>15.3</td>
<td>LAC</td>
<td>0</td>
</tr>
<tr>
<td>P3GA</td>
<td>23.8372</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reactions, rate laws, parameters and species concentration used in extended model
following are the enzyme catalyzed reactions that are used in the extended model.

1. Phosphotransferase: glucose (PTS:GLUC)
   \[ \text{PEP} + \text{GLUC} \rightarrow \text{G6P} + \text{PYR} \]
   \[ \nu_{\text{PTS:gluc}} = \frac{V_{\text{PTS:gluc}}^{\text{max}} \frac{[\text{GLUC}]}{K_M^{\text{GLUC}}} \frac{[\text{PEP}]}{K_M^{\text{PEP}}}}{1 + \frac{[\text{GLUC}]}{K_M^{\text{GLUC}}} + \left(1 + \frac{[\text{PEP}]}{K_M^{\text{PEP}}} + \left(1 + \frac{[\text{G6P}]}{K_M^{\text{G6P}}} \right) \frac{[\text{PYR}]}{K_M^{\text{PYR}}} \right)} - 1', \tag{E.1} \]

2. Phosphoglucoisomerase (PGI)
   \[ \text{G6P} \rightleftharpoons \text{F6P} \]
   \[ \nu_{\text{PGI}} = \frac{V_{\text{PGI}}^{\text{max}} \frac{[\text{G6P}]}{K_M^{\text{G6P}}} \frac{[\text{F6P}]}{K_M^{\text{F6P}}}}{1 + \frac{[\text{G6P}]}{K_M^{\text{G6P}}} + \left(1 + \frac{[\text{F6P}]}{K_M^{\text{F6P}}} \right)} - 1' \]

3. Phosphofructokinase (PFK)
   \[ \text{F6P} + \text{ATP} \rightarrow \text{FBP} + \text{ADP} \]
   \[ \nu_{\text{PFK}} = \frac{V_{\text{PFK}}^{\text{max}} \frac{[\text{F6P}]}{K_M^{\text{F6P}}} \frac{[\text{ATP}]}{K_M^{\text{ATP}}}}{1 + \frac{[\text{F6P}]}{K_M^{\text{F6P}}} + \left(1 + \frac{[\text{F6P}]}{K_M^{\text{F6P}}} \right) \frac{[\text{ADP}]}{K_M^{\text{ADP}}} \right)} - 1' \]

4. Fructose-bisphosphate aldolase (FBA)
   \[ \text{FBP} \rightleftharpoons 2\text{G3P} \]
   \[ \nu_{\text{FBA}} = \frac{V_{\text{FBA}}^{\text{max}} \frac{[\text{FBP}]}{K_M^{\text{FBP}}} \frac{[\text{G3P}]}{K_M^{\text{G3P}}}}{1 + \frac{[\text{FBP}]}{K_M^{\text{FBP}}} + \frac{[\text{G3P}]}{K_M^{\text{G3P}}} + \left(1 + \frac{[\text{G3P}]}{K_M^{\text{G3P}}} \right)^2} \]

5. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
   \[ \text{G3P} + \text{P} + \text{NAD} \rightleftharpoons \text{BPG} + \text{NADH} \]
   \[ \nu_{\text{GAPDH}} = \frac{V_{\text{GAPDH}}^{\text{max}} \frac{[\text{G3P}]}{K_M^{\text{G3P}}} \frac{[\text{NAD}]}{K_M^{\text{NAD}}} \frac{[\text{P}]}{K_M^{\text{P}}}}{1 + \frac{[\text{G3P}]}{K_M^{\text{G3P}}} \cdot \left(1 + \frac{[\text{NAD}]}{K_M^{\text{NAD}}} \right) \cdot \left(1 + \frac{[\text{P}]}{K_M^{\text{P}}} \right) \cdot \left(1 + \frac{[\text{NADH}]}{K_M^{\text{NADH}}} \right) - 1} \]

6. Enolase (ENO)
   \[ \text{BPG} + \text{ADP} \rightleftharpoons \text{PEP} + \text{ATP} \]
   \[ \nu_{\text{ENO}} = \frac{V_{\text{ENO}}^{\text{max}} \frac{[\text{BPG}]}{K_M^{\text{BPG}}} \frac{[\text{ADP}]}{K_M^{\text{ADP}}} \frac{[\text{PEP}]}{K_M^{\text{PEP}}} \frac{[\text{ATP}]}{K_M^{\text{ATP}}}}{1 + \frac{[\text{BPG}]}{K_M^{\text{BPG}}} + \left(1 + \frac{[\text{ADP}]}{K_M^{\text{ADP}}} \right) + \left(1 + \frac{[\text{PEP}]}{K_M^{\text{PEP}}} \right) + \left(1 + \frac{[\text{ATP}]}{K_M^{\text{ATP}}} \right) - 1} \]
7. Pyruvate kinase (PYK)

\[
V_{\text{PYK}} = \frac{[\text{FBP}]}{[\text{FBP}] + K_M^{\text{FBP}}} \cdot (1 + \frac{[\text{ADP}]}{K_M^{\text{ADP}}}) \cdot (1 + \frac{[\text{PEP}]}{K_M^{\text{PEP}}}) \cdot V_{\text{max}}^{\text{PYK}} \cdot \frac{[\text{ADP}]}{k_M^{\text{ADP}}} \cdot \frac{[\text{PEP}]}{k_M^{\text{PEP}}} \cdot \left(1 + \frac{[\text{PYR}]}{K_M^{\text{PYR}}} - 1\right)
\]

8. Lactate dehydrogenase (LDH)

\[
V_{\text{LDH}} = \frac{[\text{FBP}]}{[\text{FBP}] + K_M^{\text{FBP}}} \cdot (1 + \frac{[\text{NAD}]}{K_M^{\text{NAD}}}) \cdot V_{\text{max}}^{\text{LDH}} \cdot \frac{[\text{PYR}]}{k_M^{\text{PYR}}} \cdot \frac{[\text{NAD}]}{k_M^{\text{NAD}}} \cdot \left(1 + \frac{[\text{LAC}]}{K_M^{\text{LAC}}} - 1\right)
\]

9. Acetolactate synthase; acetolactate decarboxylase (PA)

\[
V_{\text{PA}} = \left(1 + \frac{[\text{PYR}]}{K_M^{\text{PYR}}} + \frac{[\text{NAD}]}{K_M^{\text{NAD}}} + 1 + \frac{[\text{Acetoin}]}{K_M^{\text{Acetoin}}} - 1\right)
\]

10. Acetoin Transportase (AT)

\[
V_{\text{Acetoin}_{\text{transp}}} = \frac{V_{\text{max}}^{\text{AT}} \cdot [\text{Acetoin}]}{k_M^{\text{Acetoin}} \cdot [\text{Acetoin}]} 
\]

11. Butanediol dehydrogenase (AB)

\[
V_{\text{AB}} = \frac{[\text{Acetoin}]}{k_M^{\text{Acetoin}}} \cdot [\text{NAD}] \cdot \frac{[\text{Butanediol}]}{k_M^{\text{Butanediol}}} \cdot [\text{NAD}] \cdot \left(1 + \frac{[\text{Butanediol}]}{k_M^{\text{Butanediol}}} + \frac{[\text{NAD}]}{k_M^{\text{NAD}}} - 1\right)
\]

12. Pyruvate dehydrogenase (PDH)

\[
V_{\text{PDH}} = \frac{K_M^{\text{PDH}}}{[\text{GAP}] + K_M^{\text{PDH}}} \cdot V_{\text{max}}^{\text{PDH}} \cdot \frac{[\text{PYR}]}{k_M^{\text{PYR}}} \cdot \frac{[\text{CoA}]}{k_M^{\text{CoA}}} \cdot \left(1 + \frac{[\text{AcCoA}]}{k_M^{\text{AcCoA}}} \cdot \frac{[\text{Formate}]}{k_M^{\text{Formate}}} - 1\right)
\]

13. Acetaldehyde dehydrogenase; alcohol dehydrogenase (AE)

\[
V_{\text{AE}} = \frac{K_M^{\text{AE}}}{[\text{ATP}] + K_M^{\text{AE}}} \cdot V_{\text{max}}^{\text{AE}} \cdot \frac{[\text{AcCoA}]}{k_M^{\text{AcCoA}}} \cdot \frac{[\text{NAD}]}{k_M^{\text{NAD}}} \cdot \left(1 + \frac{[\text{AcCoA}]}{k_M^{\text{AcCoA}}} + \left(\frac{[\text{NAD}]}{k_M^{\text{NAD}}}\right)^2\right) - 1
\]
14. Acetate kinase (AC)
\[ \text{Ac.Co.A} + \text{ADP} \rightarrow \text{Acetate} + \text{ATP} + \text{Co} \cdot \text{A} \]

\[ v_{AC} = \frac{V_{\text{max}}^{AC} \cdot [\text{Ac.Co.A}] \cdot [\text{ADP}]}{k_M^{AC} \cdot \text{Ac.Co.A} \cdot k_M^{AC} \cdot \text{ADP}} \]

\[ = \left( 1 + \frac{[\text{Ac.Co.A}]}{k_M^{AC} \cdot \text{Ac.Co.A}} \right) \cdot \left( 1 + \frac{[\text{ADP}]}{k_M^{AC} \cdot \text{ADP}} \right) \cdot \left( 1 + \frac{[\text{Co.A}]}{k_M^{AC} \cdot \text{Co.A}} \right) - 1 \]

15. Mannitol phosphate dehydrogenase (MPD)
\[ \text{F6P} + \text{NADH} \rightleftharpoons \text{M1P} + \text{NAD} \]

\[ v_{\text{MPD}} = \frac{V_{\text{max}}^{\text{MPD}} \cdot \text{F6P} \cdot \text{NADH}}{k_M^{\text{MPD}} \cdot \text{F6P} \cdot k_M^{\text{MPD}} \cdot \text{NADH}} - \frac{V_{\text{max}}^{\text{MPD}} \cdot \text{M1P} \cdot \text{NAD}}{k_M^{\text{MPD}} \cdot \text{M1P} \cdot k_M^{\text{MPD}} \cdot \text{NAD}} \]

\[ = \left( 1 + \frac{[\text{F6P}]}{k_M^{\text{F6P}} \cdot \text{F6P}} \right) \cdot \left( 1 + \frac{[\text{NADH}]}{k_M^{\text{NADH}} \cdot \text{NADH}} \right) + \left( 1 + \frac{[\text{M1P}]}{k_M^{\text{M1P}} \cdot \text{M1P}} \right) \cdot \left( 1 + \frac{[\text{NAD}]}{k_M^{\text{NAD}} \cdot \text{NAD}} \right) - 1 \]

16. Mannitol phosphatase (MP)
\[ \text{M1P} \rightarrow \text{Mannitol} \]

\[ v_{\text{MP}} = \frac{V_{\text{max}}^{\text{MP}} \cdot [\text{M1P}]}{1 + \frac{[\text{M1P}]}{k_M^{\text{M1P}}} + \frac{\text{Mannitol}}{k_M^{\text{Mannitol}}} - 1} \]

17. Mannitol Transportase (MT)
\[ \text{Mannitol} \rightarrow \text{Mannitol}_{\text{ext}} \]

\[ v_{\text{Mannitol}_{\text{ext}}} = \frac{V_{\text{max}}^{\text{MT}} \cdot [\text{Mannitol}] \cdot [\text{Mannitol}_{\text{ext}}]}{1 + \frac{[\text{Mannitol}]}{k_M^{\text{Mannitol}}} + \frac{\text{Mannitol}_{\text{ext}}}{k_M^{\text{Mannitol}_{\text{ext}}}} \]

18. Phosphotransferase: mannitol (pts:Mannitol)
\[ \text{Mannitol}_{\text{ext}} + \text{PEP} \rightarrow \text{M1P} + \text{PYR} \]

\[ v_{\text{pts}:\text{man}} = \frac{V_{\text{max}}^{\text{pts}} \cdot [\text{Mannitol}_{\text{ext}}] \cdot [\text{PEP}]}{1 + \frac{[\text{Mannitol}_{\text{ext}}]}{k_M^{\text{Mannitol}_{\text{ext}}}} + \frac{[\text{PEP}]}{k_M^{\text{PEP}}}} \cdot \left( 1 + \frac{[\text{M1P}]}{k_M^{\text{M1P}} \cdot \text{M1P}} \right) \cdot \left( 1 + \frac{[\text{PYR}]}{k_M^{\text{PYR}} \cdot \text{PYR}} \right) - 1 \]

19. ATPase
\[ \text{ATP} \rightarrow \text{ADP} + \text{P} \]

\[ v_{\text{ATPase}} = \frac{V_{\text{max}}^{\text{ATPase}} \cdot [\text{ATP}]}{1 + \left( \frac{[\text{ATP}]}{k_M^{\text{ATPase}} \cdot \text{ATP}} \right)^{n_{\text{ATPase}}}} + 1 \]
20. Phosphate transportase (PT)

\[ \text{P}_{\text{ext}} \rightarrow \text{P} \]

\[ \text{V}_{\text{PT}} = \frac{K_{\text{PT}} \text{P}_{\text{ext}}}{k_{\text{PT}}^{\text{i}} \text{P} + k_{\text{PT}}^{\text{i}} \text{P}_{\text{ext}}} \cdot \frac{[\text{ATP}]}{k_{\text{M}}^{\text{ATP}}} \cdot \frac{[\text{P}_{\text{ext}}]}{k_{\text{M}}^{\text{P}_{\text{ext}}}} \]

\[ \cdot \left( 1 + \frac{[\text{P}]}{k_{\text{M}}^{\text{P}}} + \left( \frac{[\text{P}]}{k_{\text{M}}^{\text{P}}} \right)^2 \right) \cdot \left( 1 + \frac{[\text{ADP}]}{k_{\text{M}}^{\text{ADP}}} \right) - 1 \]

21. FBPase

\[ \text{FBP} \rightarrow \text{F6P} + \text{P} \]

\[ \text{V}_{\text{FBP}} = \frac{V_{\text{FBP}}^{\text{max}} \cdot [\text{FBP}]}{k_{\text{M}}^{\text{FBP}} + \left( 1 + \frac{[\text{F6P}]}{k_{\text{M}}^{\text{F6P}}} \right) \cdot \left( \frac{[\text{P}]}{k_{\text{M}}^{\text{P}}} \right)} \]

The mass balance equations with their reaction rate abbreviated (table E.1), are given as:

\[ \frac{d[G6P]}{dt} = v1 - v3 \] (E.2)

\[ \frac{d[F6P]}{dt} = v3 - v4 - v18 + v21 \]

\[ \frac{d[FBP]}{dt} = v4 - v5 - v21 \]

\[ \frac{d[G3P]}{dt} = 2 \times v5 - v6 \]

\[ \frac{d[BPG]}{dt} = v6 - v7 \]

\[ \frac{d[PEP]}{dt} = -v1 + v7 + v8 - v20 \]

\[ \frac{d[PYR]}{dt} = v1 + v8 - v9 - 2 \times v10 - v11 + v20 \]

\[ \frac{d[Acetoin]}{dt} = v10 - v12 - v14 \]

\[ \frac{d[Ac.Coa]}{dt} = v11 - v15 - v16 \]

\[ \frac{d[COA]}{dt} = -v11 + v15 + v16 \]

\[ \frac{d[ATP]}{dt} = -v2 - v4 + v7 + v8 + v16 - v22 \]

\[ \frac{d[ADP]}{dt} = v2 + v4 - v7 - v8 - v16 + v22 \]
Appendix E. Reactions, rate laws, parameters and species concentration used in extended model

\[
\begin{align*}
\frac{d[NAD]}{dt} &= -v6 + v9 + v14 + 2 \times v15 + v18 \\
\frac{d[NADH]}{dt} &= v6 - v9 - v14 - 2 \times v15 - v18 \\
\frac{d[P]}{dt} &= 2 \times v2 - v6 + v21 + v22 \\
\frac{d[Mannitol]}{dt} &= v19 - v23 \\
\frac{d[M1P]}{dt} &= v18 - v19 + v20 \\
\frac{d[Formate]}{dt} &= v11 \\
\frac{d[GLUC]}{dt} &= -v1 \\
\frac{d[LAC]}{dt} &= v9 \\
\frac{d[Acetoin_{ext}]}{dt} &= v12 \\
\frac{d[Butanediol]}{dt} &= v14 \\
\frac{d[Etoh]}{dt} &= v15 \\
\frac{d[Mannitol_{ext}]}{dt} &= -v20 + v23 \\
\frac{d[ACETATE]}{dt} &= v16 \\
\frac{d[P_{ext}]}{dt} &= -v2
\end{align*}
\]

The reaction numbers are abbreviated as given in table E.1

<table>
<thead>
<tr>
<th>Reaction abbreviated</th>
<th>Name</th>
<th>Reaction abbreviated</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>PTS: GLUC</td>
<td>v12</td>
<td>AC</td>
</tr>
<tr>
<td>v2</td>
<td>PT</td>
<td>v14</td>
<td>AB</td>
</tr>
<tr>
<td>v3</td>
<td>PGI</td>
<td>v15</td>
<td>AE</td>
</tr>
<tr>
<td>v4</td>
<td>PFK</td>
<td>v16</td>
<td>AC</td>
</tr>
<tr>
<td>v5</td>
<td>GAPDH</td>
<td>v18</td>
<td>MPD</td>
</tr>
<tr>
<td>v6</td>
<td>FBA</td>
<td>v19</td>
<td>MP</td>
</tr>
<tr>
<td>v7</td>
<td>ENO</td>
<td>v20</td>
<td>pts: man</td>
</tr>
<tr>
<td>v8</td>
<td>PYK</td>
<td>v21</td>
<td>FBP_{anc}</td>
</tr>
<tr>
<td>v9</td>
<td>LDH</td>
<td>v22</td>
<td>ATP_{anc}</td>
</tr>
<tr>
<td>v10</td>
<td>PA</td>
<td>v23</td>
<td>MT</td>
</tr>
<tr>
<td>v11</td>
<td>PDH</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table E.1: Reactions name and abbreviations of equation (E.2)
Initial and estimated values for enzyme levels:
The enzyme levels or $V_{\text{max}}$ values used initially and estimated are given in table E.2. The initial values were obtained from [26], [66].

<table>
<thead>
<tr>
<th>Enzyme levels</th>
<th>Initial</th>
<th>Estimated</th>
<th>Enzyme levels</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{PTS}}$</td>
<td>3.62109</td>
<td>6.06723</td>
<td>$V_{\text{AT}}$</td>
<td>0.144479</td>
<td>0.189505</td>
</tr>
<tr>
<td>$V_{\text{PGI}}$</td>
<td>6.63613</td>
<td>12.4705</td>
<td>$V_{\text{AB}}$</td>
<td>5.23063</td>
<td>10.9469</td>
</tr>
<tr>
<td>$V_{\text{PFK}}$</td>
<td>6.05634</td>
<td>10.625</td>
<td>$V_{\text{PDH}}$</td>
<td>1.64741</td>
<td>2.42745</td>
</tr>
<tr>
<td>$V_{\text{FBA}}$</td>
<td>85.5311</td>
<td>69.4044</td>
<td>$V_{\text{AE}}$</td>
<td>6.86628</td>
<td>33.0617</td>
</tr>
<tr>
<td>$V_{\text{GDH}}$</td>
<td>125.313</td>
<td>167.112</td>
<td>$V_{\text{IGE}}$</td>
<td>1.71644</td>
<td>0.098253</td>
</tr>
<tr>
<td>$V_{\text{ENO}}$</td>
<td>148.458</td>
<td>465.946</td>
<td>$V_{\text{MPD}}$</td>
<td>4.89428</td>
<td>41.6883</td>
</tr>
<tr>
<td>$V_{\text{PYK}}$</td>
<td>15.1824</td>
<td>29.3514</td>
<td>$V_{\text{MP}}$</td>
<td>0.618081</td>
<td>1.37966</td>
</tr>
<tr>
<td>$V_{\text{LDH}}$</td>
<td>964.084</td>
<td>152.179</td>
<td>$V_{\text{MT}}$</td>
<td>4.26383</td>
<td>3.7254</td>
</tr>
<tr>
<td>$V_{\text{PA}}$</td>
<td>0.514555</td>
<td>0.47069</td>
<td>$V_{\text{ATP}}$</td>
<td>1.64597</td>
<td>4.30884</td>
</tr>
<tr>
<td>$V_{\text{PT} \text{ max}}$</td>
<td>0.314882</td>
<td>0.0059386</td>
<td>$V_{\text{ATP max}}$</td>
<td>4.04847</td>
<td>2.50595</td>
</tr>
<tr>
<td>$V_{\text{PTS man}}$</td>
<td>0.405875</td>
<td>4.30884</td>
<td>$V_{\text{FDP max}}$</td>
<td>4.04847</td>
<td>2.50595</td>
</tr>
<tr>
<td>$V_{\text{FDP max}}$</td>
<td>0.405875</td>
<td>4.30884</td>
<td>$V_{\text{ATP max}}$</td>
<td>4.04847</td>
<td>2.50595</td>
</tr>
<tr>
<td>$V_{\text{FDP max}}$</td>
<td>0.405875</td>
<td>4.30884</td>
<td>$V_{\text{FDP max}}$</td>
<td>4.04847</td>
<td>2.50595</td>
</tr>
</tbody>
</table>

Initial and estimated values for regulators:
The initial and estimated values of the regulators are given in table E.3.

Table E.3: Initial and estimated regulator levels for system of ODE’s for extended model.

<table>
<thead>
<tr>
<th>Regulators</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{PTS P}}$</td>
<td>0.679653</td>
<td>1.04566</td>
</tr>
<tr>
<td>$K_{\text{PTS FBP}}$</td>
<td>2.21297</td>
<td>6.01892</td>
</tr>
<tr>
<td>$K_{\text{PYK P}}$</td>
<td>0.990707</td>
<td>1.37804</td>
</tr>
<tr>
<td>$K_{\text{PYK FBP}}$</td>
<td>0.130987</td>
<td>0.331419</td>
</tr>
<tr>
<td>$K_{\text{PYK P}}$</td>
<td>2.27146</td>
<td>2.03591</td>
</tr>
<tr>
<td>$K_{\text{DPR FBP}}$</td>
<td>0.02328</td>
<td>0.040485</td>
</tr>
<tr>
<td>$K_{\text{DPR FBP}}$</td>
<td>0.02328</td>
<td>0.040485</td>
</tr>
<tr>
<td>$K_{\text{DPR FBP}}$</td>
<td>0.02328</td>
<td>0.040485</td>
</tr>
<tr>
<td>$K_{\text{DPR FBP}}$</td>
<td>0.02328</td>
<td>0.040485</td>
</tr>
<tr>
<td>$K_{\text{DPR FBP}}$</td>
<td>0.02328</td>
<td>0.040485</td>
</tr>
</tbody>
</table>
Appendix E. Reactions, rate laws, parameters and species concentration used in extended model

Equilibrium constants and hill coefficients:
The values of the hill coefficient and equilibrium constants are given in table E.4 (these values are considered to be known thus, are not estimated). These values are obtained from [51], [26], [66].

Table E.4: Equilibrium constants and hill coefficients used for system of ODE’s for extended model.

<table>
<thead>
<tr>
<th>Equilibrium Constants</th>
<th>Values</th>
<th>Hill coefficients</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_{PGI} )</td>
<td>0.43</td>
<td>( n_{ATP_{ext}} )</td>
<td>3</td>
</tr>
<tr>
<td>( K_{FBA} )</td>
<td>0.056</td>
<td>( n_{PYK} )</td>
<td>3</td>
</tr>
<tr>
<td>( K_{GAPDH} )</td>
<td>0.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_{ENO} )</td>
<td>27.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_{P} )</td>
<td>900000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_{PDH} )</td>
<td>650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_{AB} )</td>
<td>1400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_{MPD} )</td>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initial concentration used:
The initial and estimated concentration used are given as:

Table E.5: Initial concentrations used to simulate extended model. The concentration that was not known were estimated.

<table>
<thead>
<tr>
<th>Species</th>
<th>Initial</th>
<th>Estimated</th>
<th>Species</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6P</td>
<td>0</td>
<td>-</td>
<td>Ac. Co A</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>ATP</td>
<td>1.9</td>
<td>-</td>
<td>Acetoin</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>ADP</td>
<td>5.96</td>
<td>4.6</td>
<td>Mannitol</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>P</td>
<td>38.26</td>
<td>14.5</td>
<td>MIP</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>F6P</td>
<td>0</td>
<td>-</td>
<td>Co. A</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>FBP</td>
<td>15.3</td>
<td>33.173</td>
<td>( P_{ext} )</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>G3P</td>
<td>0</td>
<td>-</td>
<td>LAC</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>BPG</td>
<td>1.04</td>
<td>-</td>
<td>GLUC</td>
<td>40</td>
<td>-</td>
</tr>
<tr>
<td>PEP</td>
<td>5.24</td>
<td>1.94</td>
<td>Acetoin_\text{ext}</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>NAD</td>
<td>4.67</td>
<td>-</td>
<td>Mannitol_\text{ext}</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>NADH</td>
<td>3.1 \times 10^{-8}</td>
<td>3.06 \times 10^{-7}</td>
<td>Formate</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Initial and estimated $k_M$ values:
The initial $k_M$ values were obtained from [26], [66] and are given as:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_M^{GLUC}$</td>
<td>0.038457</td>
<td>0.0288522</td>
</tr>
<tr>
<td>$k_M^{PEP}$</td>
<td>0.398442</td>
<td>0.892559</td>
</tr>
<tr>
<td>$k_M^{G6P}$</td>
<td>0.25773</td>
<td>0.238339</td>
</tr>
<tr>
<td>$k_M^{PYR}$</td>
<td>0.138748</td>
<td>0.228952</td>
</tr>
<tr>
<td>$k_M^{F6P}$</td>
<td>0.660528</td>
<td>1.98118</td>
</tr>
<tr>
<td>$k_M^{ATP}$</td>
<td>0.010726</td>
<td>0.0198023</td>
</tr>
<tr>
<td>$k_M^{ADP}$</td>
<td>1.51727</td>
<td>5.68754</td>
</tr>
<tr>
<td>$k_M^{FGP}$</td>
<td>10.1035</td>
<td>11.5643</td>
</tr>
<tr>
<td>$k_M^{GDP}$</td>
<td>0.183764</td>
<td>0.132406</td>
</tr>
<tr>
<td>$k_M^{NAD}$</td>
<td>0.021761</td>
<td>0.014478</td>
</tr>
<tr>
<td>$k_M^{NADH}$</td>
<td>2.02738</td>
<td>2.67903</td>
</tr>
<tr>
<td>$k_M^{BP</td>
<td>$</td>
<td>0.184883</td>
</tr>
<tr>
<td>$k_M^{BPG}$</td>
<td>0.055127</td>
<td>0.0655325</td>
</tr>
<tr>
<td>$k_M^{ADP}$</td>
<td>0.029594</td>
<td>0.014377</td>
</tr>
<tr>
<td>$k_M^{PEP}$</td>
<td>0.933378</td>
<td>0.43378</td>
</tr>
<tr>
<td>$k_M^{PYR}$</td>
<td>1.10746</td>
<td>0.901918</td>
</tr>
<tr>
<td>$k_M^{ATP}$</td>
<td>0.359807</td>
<td>0.436221</td>
</tr>
<tr>
<td>$k_M^{ADP}$</td>
<td>3.02456</td>
<td>4.93837</td>
</tr>
<tr>
<td>$k_M^{PEP}$</td>
<td>0.127951</td>
<td>0.210271</td>
</tr>
<tr>
<td>$k_M^{ATP}$</td>
<td>100</td>
<td>136.805</td>
</tr>
<tr>
<td>$k_M^{PYR}$</td>
<td>47.1936</td>
<td>26.1164</td>
</tr>
<tr>
<td>$k_M^{NAD}$</td>
<td>0.229881</td>
<td>0.643823</td>
</tr>
<tr>
<td>$k_M^{LAC}$</td>
<td>13.8881</td>
<td>6.97436</td>
</tr>
<tr>
<td>$k_M^{NAD}$</td>
<td>4.61517</td>
<td>5.58543</td>
</tr>
<tr>
<td>$k_M^{Mannitol}$</td>
<td>0.333555</td>
<td>0.024406</td>
</tr>
<tr>
<td>$k_M^{Mannitolext}$</td>
<td>1.70598</td>
<td>0.39178</td>
</tr>
<tr>
<td>$k_M^{PEP}$</td>
<td>0.01</td>
<td>0.0078102</td>
</tr>
<tr>
<td>$k_M^{MIP}$</td>
<td>1.78052</td>
<td>1.53154</td>
</tr>
<tr>
<td>$k_M^{ATP}$</td>
<td>0.841895</td>
<td>0.376914</td>
</tr>
<tr>
<td>$k_M^{P_{ext}}$</td>
<td>3.638</td>
<td>1.495</td>
</tr>
<tr>
<td>$k_M^{P}$</td>
<td>0.028421</td>
<td>0.00214667</td>
</tr>
</tbody>
</table>
Calculation of $RD$ values using MATLAB
Appendix F. Calculation of \( RD \) values using MATLAB

**Single parameter perturbation:**

Single parameter was perturbed at a time and its effect was seen in desired metabolite.

```matlab
% Initialize all the parameters and initial conditions
% simulation
[tROS, XROS] = ode15s(@returnVelocity, time, species, [],vmax, k);
sim_butanediol = XROS(:,21);
area_stndrd_butanediol = trapz(tROS, sim_butanediol); % area under the curve
plot(tROS,sim_butanediol);

% Single Parameter perturbation
lvmax = (-3/100)*vmax; % perturbation of -3%
rd_butanediol = zeros(21,1); % max value at each time of dynsens
figure;
for r = 1:length(vmax)
    newvmax_1 = vmax; % newvmax_1 is the matrix we want to parse to ODE ...
    newvmax_1(r,1) = lvmax(r,1)+ vmax(r,1); % every time, one value ...
    [tROS2, XROS2] = ode15s(@returnVelocity_2, time, species, ...
                            [],newvmax_1, k);
sim2_butanediol = XROS2(:,21);
area_butanediol = trapz(tROS2, sim2_butanediol); % area under the curve
rd_butanediol(r,1) = (area_butanediol - ...
                         area_stndrd_butanediol)/area_stndrd_butanediol; % RD value
end

% Solve the ODE with original parameters (function: returnVelocity)
% Solve the ODE with perturbed parameters (function: returnVelocity_2)
```

F-2
Double parameter perturbation:
Double parameters were perturbed at a time and its effect was seen in desired metabolite. All possible combination of 2 enzymes from 21 set of enzyme levels were assessed.

```matlab
1 % Initialization and simulation as above
2 %---------------TWO parameters perturbed at once---------------
3 lvmax = (-3/100)*vmax;
4 figure;
5 combos = combnns(1:21,2); % All combination of 2 enzyme levels from 21 enzyme levels
6 findex = combos(:,1);
7 bindex = combos(:,2);
8 rd_butanediol = zeros(length(combos),1); % max value at each time of dynsens
9 for r = 1:length(combos)
10    newvmax_1 = vmax; % newvmax_1 is the matrix we want to parse to ODE ...
11       solver
12       newvmax_1(findex(r)) = lvmax(findex(r))+vmax(findex(r)); % every ...
13       time, one value gets replaced from the ... original parameter, leaving rest unchanged
14       newvmax_1(bindex(r)) = lvmax(bindex(r))+vmax(bindex(r));
15       [tROS2, XROS2] = ode15s(@returnVelocity_2, time, species, ... [],newvmax_1, k);
16       sim2_butanediol = XROS2(:,21);
17       area_butanediol = trapz(sim2_butanediol);
18       rd_butanediol(r,1) = (area_butanediol ... area_stndrd_butanediol)/area_stndrd_butanediol;
19 end
20 % Solve the ODE with original parameters (function: returnVelocity)
21 % Solve the ODE with perturbed parameters (function: returnVelocity_2)
22 % End
```
Appendix F. Calculation of $RD$ values using MATLAB