Mika Toivanen

# SIMULATION OF *LACTOCOCCUS LACTIS* METABOLIC NETWORKS

Thesis for the degree of Master of Science in Technology.

Espoo, 11th August 2005

| | |
|---|---|
| Supervisor: | Academy Prof. Kimmo Kaski |
| Instructors: | Antti Nyyssölä, D. Sc. (Tech.) |
| | Juha-Pekka Pitkänen, M. Sc. (Tech.) |

# Foreword

The work was carried out in the Laboratory of Computational Engineering between June 2004 and March 2005. Preliminary studies were done earlier in 2003 and 2004. The work was part of a joint research project to study the systems biology of the lactic acid bacteria with D.Sc. Antti Nyyssölä at the Laboratory of Bioprocess Engineering at Helsinki University of Technology (TKK) and Prof. Airi Palva at the Faculty of Veterinary Medicine at University of Helsinki. I am grateful for Academy Prof. Kimmo Kaski and Prof. Matti Leisola who made it possible for me to participate in such a multidisciplinary project.

I also want to thank my instructors Antti Nyyssölä and Juha-Pekka Pitkänen for guidance and suggestions on the manuscript. They have helped me staying in the subject and focusing on the relevant issues. I also wish to thank Kristiina Kiviharju and Ulla Moilanen at the Bioprocess Engineering laboratory at TKK for running the wet lab experiments. In addition, I want to acknowledge the help from Michael Patra and Jari Saramäki relating to Fortran language and compiling problems I have had. Generally I wish to thank my room mates and other colleagues at the laboratory for interesting and enlightening discussions on research subjects and personal issues.

Finally, I want to thank my wife Saara for her patience and support during the writing process.

Hanko, 11th August 2005

Mika Toivanen

| Author | Date |
|---|---|
| Mika Toivanen | 11th August 2005 |
| | Pages |
| | 71 (+18) |

| Title of Thesis | |
|---|---|
| SIMULATION OF *LACTOCOCCUS LACTIS* METABOLIC NETWORKS | |

| Chair | Chair Code |
|---|---|
| Computational Systems Biology | S-114 |

| Supervisor | |
|---|---|
| Academy Professor Kimmo Kaski | |

| Instructors | |
|---|---|
| Antti Nyyssölä, D. Sc. (Tech.) | |
| Juha-Pekka Pitkänen, M. Sc. (Tech.) | |

This thesis investigates if a model based on enzyme kinetics can be used to simulate bacterial metabolism. The target pathway is the *Lactococcus lactis* glycolytic pathway including 26 reactions and 34 metabolites. A metabolic engineering aspect is included in the study by addition of an reconbinant pathway that enables production of xylitol from xylose.

Two kinetic models of *L. lactis* were composed together with a parameter estimation method. The models were based on mechanistic and power-law rate equations. A stoichiometric model and the rate equations together formed an ordinary differential equation system which was solved numerically. A combination of a random search and a Monte-Carlo minimization methods was used to estimate the 164 and 146 parameters of the mechanistic and power-law models, respectively. A series of six cultivation experiments was carried out and data was collected for parameter estimation.

The literature parameters were not usable with the mechanistic model. Additionally, the parameter estimation was not satisfactory because the squared difference of model predictions and measurements remained above 414000. The mechanistic and power-law models were halted in an equilibrium state after 30 and 0.5 minutes, respectively. This was caused by depletion of phosphate in the simulations. As a result, an intermediate glycolytic reaction ceased completely. The mechanistic model was found to be five times faster than the power-law model in terms of calculation time. This was most probably due to the slow evaluation of the power function by the computer.

The models did not perform as expected but with suggested changes these models could be used to simulate bacterial metabolism. Both models were excessively complex and sensitive to their parameter values. The number of parameters should be brought down to a third of the current value due to the lack of detail seen in the data. Additionally, the initial concentrations should be subjects to parameter estimation to make the model less static.

TEKNILLINEN KORKEAKOULU  DIPLOMITYÖN TIIVISTELMÄ

Kemian tekniikan osasto

| Tekijä | Päiväys |
|---|---|
| Mika Toivanen | 11.8.2005 |
| | Sivumäärä |
| | 71 (+18) |

| Työn nimi | |
|---|---|
| *LACTOCOCCUS LACTIS* -MAITOHAPPOBAKTEERIN METABOLIAREITTIEN SIMULOINTI | |

| Professuuri | Koodi |
|---|---|
| Laskennallinen systeemibiologia | S-114 |

| Valvoja | |
|---|---|
| Akatemian Professori Kimmo Kaski | |

| Ohjaajat | |
|---|---|
| TkT Antti Nyyssölä | |
| DI Juha-Pekka Pitkänen | |

Työssä tutkittiin bakteeriaineenvaihdunnan simulointia entsyymikinetiikkaan perustuvan mallin avulla. Mallinnuksen kohteena oli *Lactococcus lactis* -maitohappobakteerin glykolyysireitti, joka sisältää 26 reaktiota ja 34 metaboliittia. Työssä tutkittiin myös mallin sovelluksia metaboliamuokkauksessa lisäämällä malliin reitti, jonka avulla kanta tuotti ksyloosista ksylitolia.

Työssä tehtiin kaksi eri aineenvaihduntamallia. Ne perustuivat reaktiomekanismeista johdettuihin nopeusyhtälöihin sekä potenssifunktioihin. Yhdessä stoikiometriamallin kanssa mallit muodostivat differentiaaliyhtälösysteemin, joka ratkaistiin numeerisesti. Mekanistisen mallin 164 ja potenssifunktiomallin 146 parametria estimoitiin satunnaishaun ja Monte-Carlo minimoinnin yhdistelmällä. Parametrien estimoimiseksi tehtiin kuusi kasvatuskoetta, joiden tulokset sovitettiin malliin.

Parametrien estimointi ei tuottanut tyydyttäviä tuloksia. Neliövirhe oli molemmilla malleilla yli 414000. Mekanistinen malli ja potenssifunktiomalli ajautuivat tasapainotilaan 30 minuutin ja 0,5 minuutin simuloinnin jälkeen. Aineenvaihdunnan pysähtymisen aiheutti fosfaatin loppuminen, minkä seurauksena eräs glykolyysin välireaktio pysähtyi kokonaan. Mekanistinen malli oli laskuissa viisi kertaa nopeampi kuin potenssifunktiomalli, mikä johtui todennäköisesti potenssifunktion laskemisen hitaudesta tietokoneella.

Mallit eivät toimineet täysin odotetulla tavalla, mutta muutamilla korjauksilla niistä voidaan kehittää toimivia. Molemmat mallit olivat liian monimutkaisia ja herkkiä parametriensä suhteen. Parametrien määrä pitäisi laskea kolmannekseen nykytasosta, koska datan perusteella ei saada luotettavia estimaatteja näin monelle parametrille. Lisäksi simuloinnin alkukonsentraatiot pitäisi estimoida kuten muutkin parametrit, jotta vältettäisiin simuloinnin liian aikainen pysähtyminen tasapainotilaan.

# Table of Contents

# Nomenclature

## Abbreviations

**CDW**  Cell dry weight

*E. coli*  *Escherichia coli*

*L. lactis*  *Lactococcus lactis*

**LAB**  Lactic acid bacterium

**MCA**  Metabolic control analysis

**ODE**  Ordinary differential equation

**PMF**  Proton motive force.

*S. cerevisiae*  *Saccharomyces cerevisiae*

## Symbols

$\mathbb{C}$  Relative complexity of a rate equation

$C_i$  Concentration of metabolite $i$ in $mM = mmol\ dm^{-3}$

$k_j$  Chemical rate constant of reaction step $j$ in $min^{-1}$

$K$  Collective symbol for (enzyme) kinetic parameters

$K_{eq}$  Equilibrium constant (dimension depends on the reaction)

$K_{il}$  Inhibition constant of metabolite $I$ in $mM$

$K_{ml}$  Michaelis constant for metabolite $I$ in $mM$

$S$  Stoichiometric matrix

$v$  Reaction rate in $mmol\ min^{-1}\ mg(CDW)^{-1}$

$V$  Limiting rate in $mmol\ min^{-1}\ mg(CDW)^{-1}$

# Chapter 1

# Introduction

The interest in computational methods in biological applications has recently been increasing. This is not only due to increased complexity and computational demand in data analysis, but also due to a whole new systems approach to biological sciences. There is a clear trend of learning to model biological phenomena in a more quantitative manner.

The scientific ground for metabolic modeling dates back to the 18th century, when Antoine Lavoisier discovered the law of mass conservation and demonstrated it by a yeast fermentation experiment (Hudson, 1994). This experiment proved that living organisms do obey the laws of chemistry and physics (although he was not aware of the living nature of yeast). Scientists of the 20th century have further studied cells and described their structure and functions. A proper theory of enzyme kinetics was established in 1960's and 70's (Cleland, 1963a,b,c) and no major changes to it have been made since. By the end of the millennium, databases were filled with gene and amino acid sequences, reaction pathways and crystallographic data. However, not many attempts have so far been made to integrate this knowledge into a comprehensive model. It should be emphasized that data in itself is not scientific knowledge, but the models constructed upon them are.

Chemical synthesis has been the production method for several compounds since the industrial revolution. Since the 1970's biotechnology has been offering an increasing amount of alternative paths to produce many of these compounds. Biotechnological processes are usually cleaner and less energy consuming than their chemical counterparts. The drawback of biotechnological processes is that the yields and product concentrations are commonly low which decreases profitability and makes the chemical synthesis a more attractive choice. However, the toolbox of biotechnology is ever-increasing since new reaction pathways are discovered constantly. The reaction pathways inside a cell have been used for centuries (e.g. in brewing) but modern biotechnology allows scientists to maximize the activity of a chosen pathway by the methods of metabolic engineering. By genetic or biochemical modifications the functions of a metabolic network can be changed to favor a particular product and increase yield and productivity of the whole process. The eventual goal is to make the process more competitive.

A model is a measure of knowledge and a very useful tool for metabolic engineering. With the model scientists can effectively design proper changes to the network with fewer wet lab experiments. The metabolic model is basically composed of the network structure (topology) and the kinetic descriptions of the reactions. The time course of the concentrations of compounds involved can then be calculated with the model. However, the kinetic parameters usually hold a great deal of uncertainty. Thus, the parameters should be fitted to *in vivo* measurements in order to get a reliable model. Measurement of the external metabolites is everyday routine in many laboratories, but analyzing the internal metabolite concentrations reliably is more difficult. Several rapid sampling devices have been introduced that can take up to five samples a second from the fermentation broth of which the extra- and intracellular metabolites are analyzed (Buchholz *et al.*, 2002; Visser *et al.*, 2002).

A lactic acid bacterium was chosen for the study because it has many advantages over the more extensively studied microorganism such as *Escherichia coli* and *Saccharomyces cerevisiae*. First of all, compared to *S. cerevisiae*, LAB are single-cell prokaryotes, which simplifies the biological picture. Only intracellular and extracellular compartments are needed, and these are assumed to be completely mixed. *S. cerevisiae* is a eukaryote and the cell structure is more complicated including nuclear and mitochondrial compartments. Secondly, LAB have more simple energy production mechanisms than the other two microbes have. It generates energy by product level phosphorylation without respiration. Third, it is by far the simplest of the three in terms of protein coding genes; According to GenomeNet (Anonymous, 2004) *L. lactis*, *E. coli* and *S. cerevisiae* have 2266, 4289, 5855 protein coding genes, respectively.

In this master's thesis one of the main goals was to compose a model that could be used in metabolic engineering. This model should be able to predict the behavior of the organism in different environments. However, the first target was to see whether a kinetic model could be fitted to experimental measurements at all. The tested models were composed of enzyme kinetics and power-law rate equations. In addition, suitability of literature parameters and parameter estimation routines were analyzed. The study also evaluates the current methods to measure intra- and extracellular metabolite concentrations.

# Part I

# Literature survey

# Chapter 2

# Metabolism of lactic acid bacteria

This chapter describes the general properties of lactic acid bacteria (LAB) and *Lactococcus lactis* in particular. It also presents the significant properties that have to be taken into account when modeling the metabolism and physiology of LAB *in silico*.

LAB have a long history of metabolic engineering (de Vos and Hugenholtz, 2004; Hugenholtz *et al.*, 2002). They are popular engineering targets because they are widely used in sour milk products, they are a well known member of human biota and they are classified as "generally recognized as safe" (GRAS) -organisms. They also have an interesting property of having rather separate catabolic and biosynthetic pathways as will be discussed later. This property gives the researcher greater freedom for engineering, since the mutations only effect a specific part of the whole metabolic network. Therefore, they are ideal production hosts of various bioproducts.

## 2.1 Physiology

The group of lactic acid bacteria has no exact definition. It is based on historical classification and contains several different genera of modern bacterial taxonomy. Common characteristics of lactic acid bacteria is that they are Gram-positive, aerotolerant, acid-tolerant, strictly fermentative rods or cocci that produce lactic acid as their main product (Axelsson, 1998). The species of interest, *L. lactis*, is a member of the homofermentative LAB and it produces under ideal conditions almost exclusively lactic acid, while the members of the heterofermentative LAB exhibit a mixed acid fermentation mode and produce ethanol, acetate, formate and carbon dioxide in addition to lactic acid. The core reactions of these pathways are presented in figure 2.1. The biggest difference of the two fermentation pathways is in the reactions between glucose-6-phosphate (G6P) and glyceraldehyde-3-phosphate (GAP). It should be noted that also homolactic LAB can exhibit mixed acid fermentation, but this depends on the redistribution of carbon at the level of pyruvate, not on the enzymes specific for heterofermentative LAB.

## Homolactic pathway

Glucose
**GLK** — ATP → ADP
G6P
**PGI**
F6P
**PFK** — ATP → ADP
FBP
**FBA**
GAP ⇌ DHAP
**GAPDH** — NAD$^+$, Pho → NADH
1,3PG
**PGK** — ADP → ATP
3PG
**PMG**
2PG
**ENO**
PEP
**PK** — ADP → ATP
Pyruvate
**LDH** — NADH → NAD$^+$
Lactate

## Heterolactic pathway

Glucose
**GLK** — ATP → ADP
G6P
**G6PDH** — NAD$^+$ → NADH
6–phospho–gluconate
**6PGDH** — NAD$^+$ → NADH, $CO_2$
Ribulose–5–phosphate
**PHE**
Xylulose–5–phosphate
**PHK** — Pho
GAP / Acetyl–phosphate

**GAPDH** — NAD$^+$, Pho → NADH
1,3PG
**PGK** — ADP → ATP
3PG
**PMG**
2PG
**ENO**
PEP
**PK** — ADP → ATP
Pyruvate
**LDH** — NADH → NAD$^+$
Lactate

Acetyl–phosphate
**PTA** — CoA → Pho
Acetyl–CoA
**ACDH** — NADH → NAD$^+$
Acetaldehyde
**ADHE** — NADH → NAD$^+$
Ethanol

**Figure 2.1:** The major homo- and heterolactic fermentation pathways based on Axelsson (1998). Metabolites are written in normal face and enzymes in bold face. Abbreviations for metabolites are *Glc*: glucose, *G6P*: glucose-6-phosphate, *F6P*: fructose-6-phosphate, *FBP*: fructose-1,6-bisphosphate, *DHAP*: dihydroxyacetonephosphate, *GAP*: glyceraldehyde-3-phosphate, 1,3*PG*: 1,3-diphosphoglycerate, 3*PG*: 3-phosphoglycerate, 2*PG*: 2-phosphoglycerate, *PEP*: phosphoenolpyruvate, *CoA*: coenzyme-A, *Pho*: inorganic phosphate, *NAD$^+$* and *NADH*: nicotinamide adenine dinucleotide oxidized and reduced forms, respectively, *ATP*: adenosine-triphosphate, *ADP*: adenosine-diphosphate and *$CO_2$*: carbon dioxide. Enzyme abbreviations not listed in table 6.1 are **G6PDH**: Glucose-6-phosphate dehydrogenase, **6PGDH**: 6-phospho-gluconate dehydrogenase, **PHE**: Phosphoribulose epimerase and **PHK**: Phosphoketolase.

5

LAB tolerate relatively low pH values. Lactococci particularly tolerate external pH of 3 – 4 while maintaining internal pH above a threshold value of 5.0 (Axelsson, 1998). This gives LAB a great advantage in competition for energy sources. They acidify the habitat which severely inhibits the growth of other organisms.

LAB are very demanding in terms of the nutrient content of the growth medium. This is because their metabolism generally cannot produce the macromolecules needed for growth. Therefore LAB need most of the amino acids, nucleic acids, vitamins and growth factors from the medium. For this reason it is important that LAB have an efficient method for taking over nutrient rich habitats.

Generally the catabolic pathways[1] of LAB are distinct from the anabolic[2] ones. This simplifies the network since most of the glucose is used for energy production rather than as biosynthetic precursor. Novák and Loubiere (2000) have studied the distinction between catabolic and anabolic pathways of *L. lactis* strain NCDO 2118 by ^{14}C-labeled substrates. They came to the conclusion that only 5 % of glucose is utilized in anabolic pathways in synthetic medium at exponential growth phase. However, this low percentage should not be underestimated since it corresponds to 66 % of total biomass carbon making glucose a major precursor of biomass (Novák and Loubiere, 2000). The percentage of biomass originating from glucose diminishes in cultivations on complex medium. On the other hand, in this thesis the presumed production strategy is based on use of cultures of resting cells which results in close-to-zero activity of anabolic pathways. This makes the LAB, with distinct catabolic and anabolic pathways, excellent production hosts for resting cell processes.

Resting cells have some important benefits compared to growing cell cultures. Resting cells do not require nutrients for growth under production conditions. Once an appropriate amount of biomass is grown the growth medium is changed to a buffer solution that contains the raw materials for the desired product. Continually growing cells, however, require a constant feed of rich growth medium. Complex growth media are generally expensive and their excessive use decreases the profitability of the process. In addition, resting cells are often reusable. The same batch of cells can be used multiple times with little decrease in productivity. In a continuous culture, the cells are lost once they are flushed out of the reactor unless a cell recycling unit is implemented.

## 2.2 Membrane transport

The cell membrane is unarguably the most important structure of living organisms. It has a vital role in controlling the flow of molecules in and out of the cell. This role poses a problem for metabolic engineering. In order to correctly model the behavior of the cell, the functions of

---

[1]Catabolism is degradation of carbon source, e.g. glucose, to produce energy and metabolic end products such as lactic acid.

[2]Anabolism is formation of biomass constituents, that is cells, from simple precursor molecules such as amino acids and carbohydrates.

the cell membrane and cell wall have to be properly characterized. These functions have been reviewed by Saier *et al.* (1996) and Poolman (2002). Membrane transport of solutes can be divided into primary and secondary transport and group translocation systems. These transport mechanisms differ in the source of the driving force for the transport.

The energy driving primary transport originates from direct hydrolysis of ATP to ADP and phosphate. In this case there are two types of mediators; ATPase-type transporters and so called ATP binding cassettes (ABC). Both of these mediators have a motif called nucleotide binding domain (NBD) where ATP hydrolysis occurs (Nicholls and Ferguson, 2001). These membrane-bound protein complexes are common for Gram-positive and -negative bacteria with similarity to mammalian transport proteins. This similarity is an indication of the universality of these systems. The primary transport reactions are in effect irreversible due to the hydrolysis of the phosphate bond (Axelsson, 1998).

Secondary transport is based on simultaneous transport of two different metabolites. In this case, the uphill transport of the other metabolite is made possible by the Gibbs free energy difference associated to the downhill transport of the other. This mode of transport is controlled by a membrane associated carrier or a permease protein. Many solutes are transported this way, particularly amino acids and sugars (Axelsson, 1998). In many cases the co-transported molecule is proton. Proton gradient is generated by a $H^+$ ATPase membrane protein that pumps protons out of the cell while simultaneously hydrolyzing ATP. The proton gradient generates a proton motive force (PMF). PMF consists of two trans-membrane components; pH gradient and electrical potential gradient. Thus a symporter permease can transport a proton and e.g. a D-xylose molecule into the cell by utilizing energy stored in PMF. On the contrary to primary transport, secondary transport is reversible.

The third method of solute transport is the group translocation system. It consists of several enzyme complexes which, in effect, take a phosphoryl group from phosphoenolpyruvate (PEP) and transfer it to a glucose molecule that is transported from extracellular space to the cytosol as glucose-6-phosphate (G6P) in the process. Thus, PEP is converted to pyruvate and at the same time extracellular glucose is converted to intracellular G6P. The enzymes involved in the translocation process are enzyme complex I (EI), a heat-stable protein (HPr) and sugar specific enzyme complexes IIA (EIIA) and IIBC (EIIBC), latter of which is membrane bound (Axelsson, 1998; Poolman, 2002). However, the enzyme complex II can adopt multiple arrangements and in *L. lactis* in particular there is no free glucose specific EIIA, but an EIIAB complex. Only for mannitol and cellobiose there is a free EIIA in *L. lactis* (Bolotin *et al.*, 2001). In addition, even thought EIIA and EIIBC are mostly sugar specific, it is known that mannose and glucose share these enzymes in LAB. The group translocation is tightly connected to the metabolism of the transported molecule. Thus, group translocation is usually used for metabolizable sugars (Axelsson, 1998; Saier *et al.*, 1996) and alditols (Saier *et al.*, 1996).

## 2.3   Control of membrane transport

The tightly connected uptake and metabolism of a carbohydrate results in *hierarchical control* where the most preferable substrate available is used first. The name reflects the hierarchy of the substrates. The other method of regulation by PTS is *autoregulation* where the rate of metabolism is adjusted to the needs of the microbe. The difference between the two methods is that autoregulation controls the activity of a single pathway while hierarchical control affects multiple pathways (Poolman, 2002).

The most important control factor of LAB metabolite transport are HPr and its different phosphorylated variates, namely the one phosphorylated at histidine-15, HPr(His-P), and the other phosphorylated at serine-46, HPr(Ser-P). The HPr(His-P) is directly involved in phosphoryl transfer from PEP to glucose, while HPr(Ser-P) has only controlling nature (Figure 2.2). Basically, serine phosphorylation controls the rate of PTS mediated transport by adjusting the amount of HPr taking part in phosphoryl transfer to glucose. On the other hand, HPr(Ser-P) also has direct allosteric control over the secondary (PMF-dependent) transport and transcriptional control over the catabolite responsive element (CRE). The balance between HPr and HPr(Ser-P) depends on two enzymes; the ATP-dependent fructose-1,6-bisphosphate (FBP) -activated protein kinase and the phosphate-activated protein phosphatase (Poolman, 2002). The hierarchical control (also known as *catabolite repression*) is explained to operate in three different ways (Figure 2.2):

1. Transcriptional control of the genes involved in the uptake of carbohydrates. HPr(Ser-P) forms a ternary complex with catabolite control protein (CcpA) and catabolite responsive element (CRE) and prevents transcription of many metabolically important enzymes. Free CcpA, on the other hand, activates the transcription of *las* operon which encodes for three important glycolytic enzymes; phosphofructokinase, pyruvate kinase and lactate dehydrogenase (Poolman, 2002).

2. Preventing further uptake of the less preferable carbohydrates by inhibiting corresponding enzymes (*inducer exclusion*). HPr(Ser-P) inhibits secondary transport (carrier proteins) while EII complexes that are specific for the more preferable carbohydrates have also higher affinity for HPr(His-P), thus inhibiting uptake of less preferable carbohydrates via the PTS pathway.

3. Dephosphorylation and consequent pumping out of the less preferable carbohydrate when a more preferable one is available (*inducer expulsion*). HPr(Ser-P) unlocks the carrier proteins and leaves the channel fully open discharging the sugar concentration gradient. This means that the intracellular sugar concentrations settles down to the significantly lower sugar concentration of the extracellular medium. The shift towards equilibrium results in pumping out (*expulsion*) of the corresponding sugars. To enable this mechanism also for sugar phosphates (PTS transported sugars), A HPr(Ser-P) activated sugar phosphatase II (PaseII) dephosphorylates sugar phosphates and simultaneously transports the sugar out of the cell via the EIIC (Poolman, 2002).
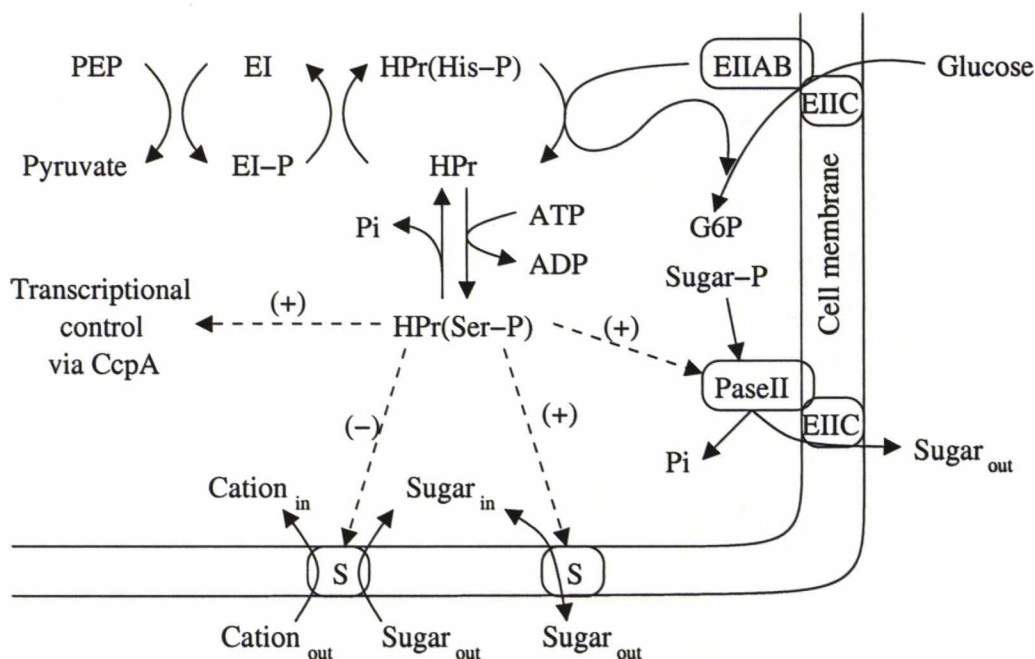
**Figure 2.2:** Phosphoenolpyruvate phosphotransferase system (Poolman, 2002). Solid lines are reactions and dashed lines are regulatory connections where sign indicates the nature of regulation. S denotes secondary transport protein. See text for other abbreviations.

## 2.4 Glycolytic pathway

The possible pathways in *L. lactis* are quite well known because it is one of the best studied micro-organisms besides *Escherichia coli* and *Saccharomyces cerevisiae* and because its genome was recently sequenced by Bolotin *et al.* (2001). The genome sequence indicates the presence of potential enzymes which can mostly be identified through homology comparisons without the need for wet lab characterization of their properties. In the following a brief overview of the glycolytic pathway of *L. lactis* is given. Many of the enzymes on the glycolytic pathway have been characterized decades ago and nowadays the research focuses on regulatory matters.

In *L. lactis* the homofermentative glycolytic pathway greatly resembles that presented in many biochemistry textbooks (see e.g. Stryer, 1995, p. 490). In addition to the homofermentative pathway, *L. lactis* also has many enzymes of the heterofermentative pathway and of the electron transfer chain present, but these pathways are not complete, and thus, mostly inoperable. Under certain conditions, however, the homofermentative pathway is capable of adopting a mixed-acid fermentation mode by changing the fluxes at the level of pyruvate. The central metabolism of *L. lactis* is reviewed by Cocaign-Bousquet *et al.* (2002) and the network has glucose, lactose and galactose as carbon sources, and lactate, formate, acetate, ethanol and carbon dioxide as end products. The relevant part of the network for this thesis is shown in figure 6.1 on page 37. The overall reaction for the homofermentative pathway with fully homolactic fermentation

mode and balanced NADH/NAD$^+$-ratio is given below (Equation 2.1)

$$Glucose \rightarrow 2\ Lactate + 2\ ATP. \tag{2.1}$$

For the homolactic fermentation in full mixed acid mode the overall reaction is a bit different (Equation 2.2)

$$Glucose \rightarrow 2\ Formate + Ethanol + Acetate + 3\ ATP. \tag{2.2}$$

The equations above indicate that the ATP yield from glucose is higher in the mixed acid fermentation mode. In the real situation the fermentation mode is rarely fully homolactic or mixed acid type. The amounts of produced ATP and other end products depend on the amount of pyruvate channeled to the reactions catalyzed by LDH or PFL. According to this logic, measuring the ratio of lactate to the other end products gives the ATP yield. In the next section, the factors regulating the fermentation mode are discussed.

## 2.5   Control of glycolytic flux and fermentation mode

The glycolytic flux and fermentation mode are not independent. It has been reported (Garrigues *et al.*, 1997, 2001; Cocaign-Bousquet *et al.*, 2002) that in *L. lactis* homolactic fermentation mode is generally related to high glycolytic flux (rapidly metabolizable sugars) while mixed-acid fermentation mode is related to low glycolytic flux (slowly metabolizable sugars). The control over the fermentation mode is based on the concentrations of glycolytic intermediates. The intermediates can control enzyme levels with transcriptional control and enzyme activities with direct inhibition and allosteric interactions. The most significant controllers are NADH/NAD$^+$-ratio, fructose-1,6-bisphosphate (FBP) and triose phosphates (GAP and DHAP).

Lactate dehydrogenase (LDH) activity has great impact on the carbon flux dispersion at the level of pyruvate, modulating the overall glycolytic rate and the fermentation mode. Therefore, activators and inhibitors of LDH are of special interest. *In vitro* enzyme assays have indicated that FBP is a major activator of the LDH, but the FBP concentration *in vivo* is typically high enough to fully activate the enzyme (Garrigues *et al.*, 1997). Thus, another significant regulator must exist.

The regulator of LDH was found by an observation that at high glycolytic rate the metabolite pools of GAP, DHAP and FBP were high indicating a rate limiting step at glyceraldehyde-phosphate dehydrogenase (GAPDH). Interestingly, this enzyme, as well as the LDH, is dependent on the concentrations of NADH and NAD$^+$. This observation indicates that the glycolytic rate is modulated by a double regulation mechanism of GAPDH and LDH and that the fundamental controlling factor is the NADH/NAD$^+$-ratio (Garrigues *et al.*, 1997).

The pyruvate formate lyase (PFL) branch of the pyruvate metabolism is also regulated, though indirectly, by the NADH/NAD$^+$-ratio. Asanuma and Hino (2000) have shown that high concentrations of GAP and DHAP, caused by high NADH/NAD$^+$-ratio, inhibit the activity of PFL.
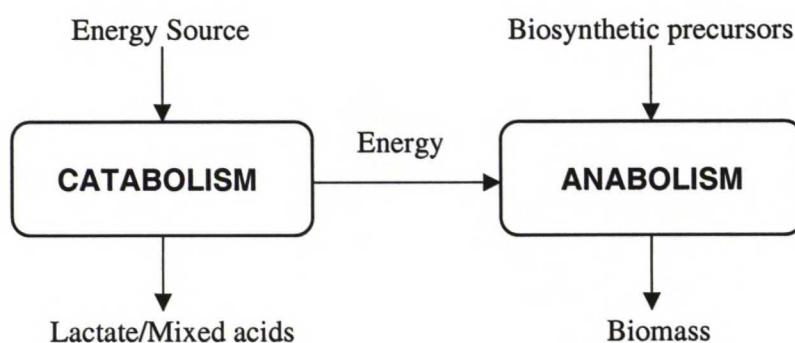
**Figure 2.3:** Schematic diagram of the energy balance between the catabolic and anabolic pathways. Energy has here a wide meaning including ATP and reduction potential (NADH).

However, Melchiorsen *et al.* (2001) argue that allosteric inhibition is not enough to explain the transition between glycolytic modes. They propose that also transcriptional regulation exists. These regulators, in addition to $NADH/NAD^+$-ratio, further direct the carbon flux towards lactate at the level of pyruvate and strengthen the homolactic fermentation mode.

To summarize, the $NADH/NAD^+$-ratio is the most significant control factor of the glycolytic rate although the factors controlling that ratio are more obscure. However, Garrigues *et al.* (2001) have argued that the mode of glycolysis, and thus the $NADH/NAD^+$-ratio, depend on the balance of catabolism and anabolism (Figure 2.3). In other words, if anabolic pathways lack raw material (biosynthetic precursors), cell growth will slow down and homolactic metabolism is likely to occur. Reciprocally, if the catabolic pathway can not provide enough energy for the anabolic pathways, the cell will switch to mixed-acid fermentation to gain the extra ATP from acetate kinase as indicated by equations 2.1 and 2.2.

The previous discussions pose a difficulty concerning the approach of this thesis; they deal with exponentially growing cells. The situation in resting cells is, without a doubt, very different. In the spirit of Garrigues *et al.* (2001) and figure 2.3, one would assume that in theory anabolic energy demand is enormous since nitrogen source is absent and biomass should be synthesized from the carbon source alone. This is obviously impossible and it is more reasonable to assume that the anabolic energy demand is zero since no biomass is synthesized. However, ATP is needed for other tasks than biosynthesis like transport and repair.

# Chapter 3

# Methods of modeling

Models of metabolic networks can be divided in a number of ways. For example, Wiechert (2002) divides them in six categories; Structural models, stoichiometric model, carbon flux models, stationary and non-stationary mechanistic models and models with gene regulation. For the scope of this thesis a clearer choice would be to divide the models in only two categories; steady-state and kinetic. Both of these categories assume that the structure (topology) of the network is known.

Steady-state models approximate the flux distribution in the network given the consumption and production rates of the boundary metabolites. This category includes the four first categories of Wiechert (2002). Kinetic models (also known as transient or dynamic models) simulate the time dependent changes in the network. They are sometimes based on enzyme kinetics (like in this thesis) but their basis can also be completely arbitrary. In addition to the metabolic reaction networks of steady-state models, kinetic models also require input of the regulatory connections in the network. Kinetic models include the last two models of Wiechert (2002).

The networks themselves are also models. Metabolic network models of novel organisms are based on gene homology comparisons between the novel organism and a database of known sequence-function records. The objective is to propose a set of possible pathways and intermediates (edges and nodes, respectively) based on prior information. Network models differ from the two models of the previous paragraph in that their topology is initially unknown. Network models could be considered as part of the structural models group of Wiechert (2002) but structural models also include knowledge of cofactors and effectors.

In any case, the division of models to categories is subjective and other well justified divisions probably exist. The following sections present some steady-state methods in less detail and kinetic methods in more detail, as they are more important for this thesis.

## 3.1 Steady-state models

Flux balance analysis (FBA) approximates all steady-state fluxes (reaction rates) of the network. This is carried out by setting the rate of change of metabolites to zero (pseudo steady-state assumption) and finding the corresponding reaction rates that satisfy this requirement and the boundary fluxes for the given network topology (Equation 3.1).

$$\frac{dC}{dt} = 0 = S \times v, \tag{3.1}$$

where $S$ stands for stoichiometric matrix and $v$ is the rate vector (see equation 6.1 for details). Initially, the resulting equation system is underdetermined because there are usually more reactions than metabolites in metabolic networks (Stephanopoulos *et al.*, 1998). Thus, fixing some fluxes by experiments reduces the degrees of freedom in the system. However, a solution can be found for the underdetermined system by additional constraints and by an optimization criterion. For example, giving limits to some fluxes and optimizing for example cell growth may produce a realistic solution. Depending on the case, different limits and optimization criteria may have to be tested before a solution is found.

In order to get a single solution, the number of undetermined rates must be less than or equal to the number of internal metabolites. If the numbers are equal the system is said to be determined and a unique solution exists. If more fluxes are measured than necessary the system is overdetermined and has a least square solution (Stephanopoulos *et al.*, 1998). In the case of overdetermined system, validity of the assumption and the integrity of the system can be tested to gain additional confidence to the model.

Recently, several modifications to FBA have been developed (Kauffman *et al.*, 2003). Incorporation of regulatory constraints decreases the number of possible solutions as solutions that require inconsistent regulatory events to occur are removed. Also thermodynamic constraints have been introduced to decrease the solution space. Finally, there have been reports that utilization of straightforward optimization of cell growth does not produce similar flux distributions to what has been observed *in vivo*. Therefore, alternative objective functions have been introduced (Kauffman *et al.*, 2003).

## 3.2 Kinetic models

In kinetic modeling the rate of a reaction or a group of reactions is expressed as a function of environmental variables and kinetic parameters. Environmental variables are usually metabolite concentrations, pH or temperature. Kinetic parameters include rate, affinity or inhibition constants and enzyme activities. The different classes of functions are presented next.

### 3.2.1 Mechanistic models

The most common class of kinetic models is the one based on mechanistic rate equations of enzyme kinetics. The justification is quite clear; they reflect the current physical and chemical knowledge of enzymatic reactions *in vitro*. These equations are reviewed thoroughly in chapter 4. The key question is, do enzymatic reactions behave similarly *in vivo* as *in vitro*? This is not trivial since the intracellular environments and assay conditions are typically very different. Usually *in vitro* assay solutions include a buffer and a controlled amount of activating or inhibiting substances, whereas the *in vivo* environment contains a variety of pH adjusting compounds and mechanisms as well as activating and inhibiting substances. Additionally, in a common *in vitro* assay the enzyme concentration is lower than under typical *in vivo* conditions. This setup is due to the fact that in an enzyme assay the enzyme concentration is fixed sufficiently low allowing the initial rate to be directly proportional to the substrate concentration.

The matter of *in vitro* versus *in vivo* kinetics has been extensively discussed, but no definite conclusions have been made. Teusink *et al.* (2000) have modeled yeast glycolysis with *in vitro* parameters in order to find out whether the results fit *in vivo* measurements. The network model they used originally was direct with feedback regulation by ATP and NAD$^+$ regeneration. This model did not perform well. The concentrations of FBP, DHAP and GAP increased to very high levels while other metabolites reached a steady concentration. This kind of behavior is due to the "turbo design" of the network model, which is described in an earlier paper by the same authors (Teusink *et al.*, 1998). The accumulation of the mentioned metabolites is caused by insufficient inhibition in a rate equation in the upper part of the glycolytic pathway. Thus, it is a property of the model rather than of an enzyme. In the direct network ATP consuming and producing reactions must cancel each others out to maintain constant amount of nucleotide phosphates (ATP, ADP and AMP). Since the ATP/ADP concentrations are the only methods of control for the hexokinase, the reaction rate is fixed at steady nucleotide phosphate concentrations. Thus, additional method of regulation is needed which would allow regulation of hexokinase independently of ATP (Teusink *et al.*, 1998, 2000). The authors have found that a mutant with a deleted trehalose-6-phosphate synthase is incapable of growing at high glucose concentrations due to the accumulation of glycolytic intermediates. They argue that the enzyme itself or the product of the reaction it catalyzes (trehalose-6-phosphate) inhibit hexokinase activity or glucose transport and slow down the turbo mechanism (Teusink *et al.*, 1998).

The other mechanism of preventing intermediate accumulation is the addition of other ATP consuming reactions (Teusink *et al.*, 2000). This approach lets the network find a steady-state condition for ATP that simultaneously allows steady-state for the glycolytic intermediates. Adding glycogen, trehalose, glycerol and succinate pathways to the direct model yields a so-called branched model (Teusink *et al.*, 2000). This branched model performed better than the direct one in the sense that the model did reach a steady-state for all intermediate metabolites. However, it did not give better fit to the *in vivo* measurements.

As the predictions did not fully fit the *in vivo* measurements, Teusink *et al.* (2000) calculated

how much each of the parameters should be adjusted to reach the steady-state flux with measured metabolite concentrations. Only changes to $V_{max}$, equilibrium and Michaelis constants were considered. The objective of this procedure was to see whether the failure in predicting concentrations with *in vitro* parameters was due to a mismatch in one or two enzymes or if it was a universal property of the network. As a result they found values for each parameter that gave the correct flux through the enzyme. These fitted parameters deviated from the original ones as much as 1000-fold, but for half of the enzymes the *in vitro* determined $V_{max}$ values were within a factor of two with the optimized value. They could, however, find an explanation for the discrepancies in most cases. Hexokinase and phosphofructokinase seemed to be the two most influential enzymes of the network.

However, comparing the optimized parameter values is of little significance since for fixed metabolite concentrations and reaction rate there is a continuum of parameter values that solve the equation. Thus, changing only one parameter out of several others that most probably have wrong values does not tell how far that particular *in vitro* parameter value is from the actual *in vivo* value. Obviously, the distance may change with another set of *in vitro* parameters. The problem is that with one steady-state measurement it is not possible to optimize multiple parameter values. For that the number of steady-state measurements would have to be equal to the number of parameters that are to be defined. These measurements should be made for significantly different metabolite concentrations while preventing activation of additional pathways absent from the original steady-state network.

Many other examples of modeling with mechanistic rate equations exist. A model for *L. lactis* by Hoefnagel *et al.* is presented separately in chapter 5. In the following the kinetic model of *S. cerevisiae* sugar metabolism is presented (Rizzi *et al.*, 1997). The organism is different from the model organism of the present thesis, but it is an example of a more complex model than what is pursued here. The model of the catabolic pathway of *S. cerevisiae* is more complex in the sense that is has multiple intracellular compartments; cytoplasm and mitochondria. This type of model is commonly used for eukaryotes but not for prokaryotes which lack intracellular compartments. Of course, both these model types have also the extracellular compartment. The model of Rizzi *et al.* (1997) has 23 reactions and it uses rate equations that are reported in the scientific literature and are based on mechanistic enzyme kinetics. The authors report a sophisticated experimental setup to measure the time course of seven internal metabolites and five co-metabolites during a glucose pulse experiment (Theobald *et al.*, 1997). The kinetic parameters of the model are estimated and the simulations are compared to the data. Interestingly, the model predictions for two metabolites that were measured are not presented (Theobald *et al.*, 1997; Rizzi *et al.*, 1997).

The results were not convincing. Out of 14 plots of predictions and data, in three plots (cytosolic ATP, ADP and NADH) the predicted curves had the same shape as the measurements but there was a constant bias. In the case of pyruvate even the shape of the predicted curve was different from the measurements, that is, the result was qualitatively wrong. The co-metabolite concentrations seem to be most error prone. This is unfortunate since co-metabolites have most

15

control in straightforward reaction networks like glycolysis. Also the pyruvate concentration is central in the control of glycolysis. In addition, the authors have used the same data to predict the parameters and to compare the concentration profiles with data. The work raises the question why proper statistical analysis was not used to estimate parameters and the predictive power of the model from separate data sets. Suitable methods would, for example, be cross validation or bootstrap. These methods give the data probability for the used model structure. The difference is that the method used by Rizzi *et al.* (1997) gives the error of fit while cross validation gives the error in model prediction. In metabolic engineering predictions are naturally more interesting. However, the authors settled for visual inspection of fit, which makes the error analysis incomplete. Despite these drawbacks, the model exhibits good qualitative results and encourages the use of mechanistic rate equations. At the same time it shows the importance of including all regulatory properties and pathways, not forgetting proper parameter estimation schemes.

### 3.2.2 Power-law models

The power-law models are based on rate equations used in simple chemical kinetics, such as that in equation 3.2.

$$v(S_1, S_2) = k \cdot S_1^a \cdot S_2^b \tag{3.2}$$

where $k$ is rate constant, $S_1$ and $S_2$ are substrate concentrations and $a$ and $b$ are kinetic order parameters for the substrates. The order parameters can have both positive and negative values. A negative value indicates that the corresponding substrate inhibits the reaction, while a positive value indicates that the substrate activates the reaction or is a reactant. The total rate of a reversible reaction is written as a sum of the forward and reverse rates (Equation 3.3). The scale of these parameters is quite narrow in most cases. A rule of thumb by Voit (2000) states that quite often the order parameters scale from $-0.5$ to $1.0$. It should be emphasized, however, that sometimes the order parameters may have values several magnitude larger than these.

It is noteworthy that these rate equations do not assume any specific reaction mechanism as Michaelis-Menten kinetics do. They only assume that the reaction rate depends on the rate of encounters of the substrates (or products), the kinetic energy of the molecule(s) and a steric factor representing a fraction of possible angles of encounter that allow the reaction to happen. These assumptions are a commonplace for any analysis of chemical kinetics.

The reaction-based definition of power-law kinetics is called generalized mass action (GMA) model by Voit (2000). Rather than reaction-based, the author favors a metabolite-based model describing the rate at which each metabolite is produced and consumed. These models are called S-systems, where S stands for synergism and saturation (Voit, 2000). S-systems do, however, produce constraints on rate terms of adjacent metabolites since the rate at which the previous metabolite is consumed is the rate at which the next is generated. GMA models can avoid this problem by writing the system in terms of a stoichiometric matrix and a rate vector.

16

The two models are similar in most cases, but differ for example at branch points. For the metabolites next to a branch point the two models are identical at only one metabolic state (for example, at steady-state). Therefore, stoichiometry is not obeyed off the steady-state and conservation of flux does not hold (Voit, 2000). The author argues that the error is small if "the deviation from steady-state is not too large". However, this opposes a problem for kinetic modeling where the model should be valid for a large range of metabolic environments far away from steady-state.

The reversible power-law rate equation for a reaction $A + B \rightleftharpoons P + Q$ is shown in equation 3.3.

$$v = k_+ A^a B^b - k_- P^p Q^q \tag{3.3}$$

where the $k_+$ and $k_-$ are the forward and reverse rate constants, respectively, and $a$, $b$, $p$ and $q$ are the order parameters for the corresponding metabolites. External inhibitor or activator are included to one or both terms, depending on their function.

A "tendency modeling" approach, presented by Visser *et al.* (2000), is a variate of the power-law models. It introduces tendency kinetics, which is essentially equal to the previously presented power-law kinetics with allosteric effectors. Tendency modeling approach has an additional feature that decreases the number of parameters in the model by use of time scale analysis. Reactions are divided into four groups depending on the time scale of the reaction compared with the time scale of the experiment. The groups in order of decreasing rate are pseudo equilibrium, pseudo steady-state, dynamic and frozen reactions. The reactions at the extremes of the time scale are then approximated.

For fast reactions a pseudo equilibrium assumption is made, which means that these reactions are fast enough to constantly operate in the close proximity of the equilibrium. This behavior is achieved by giving these reactions a very large fixed rate constant and taking the equilibrium constant as their sole parameter. A pseudo steady-state assumption is made for somewhat slower or unknown processes. This assumption actually concerns the metabolites rather than the reactions. The time derivative of these metabolites is fixed to zero, which implies a subsystem in a dynamic model that is solved with flux balance analysis (see section 3.1). Thus the reaction rates in the steady-state sub-system are determined by the boundary conditions. The dynamic rate equations follow the power-law kinetics as presented above. Finally, the slowest processes are assumed to be frozen, which means that their reaction rate is zero. One could argue, that the steady-state and frozen reactions have not been modeled at all, but they only support the actual kinetic model composed of pseudo equilibrium and dynamic reactions.

The structure of an example model of tendency kinetics in Visser *et al.* (2000) is the same as the structure of the mechanistic model in Rizzi *et al.* (1997). Particularly, the model organism is *S. cerevisiae* in both of these studies, so it is not surprising that the results of Visser *et al.* (2000) are similar to those of Rizzi *et al.* (1997). The advantage of the tendency model, due to the simplifications, is that the number of parameters is decreased to 35, compared with 84 for the Rizzi *et al.* (1997) model. The fit of data and model predictions are slightly worse for the tendency model. Interestingly, both models fail to model pyruvate concentration accurately.

The study of Visser *et al.* (2000) indicates that metabolic models commonly suffer from excess complexity and that simple models may give adequate results for most applications.

### 3.2.3 Log-linear models

Log-linear kinetics is another type of rate equation used in transient modeling. This type of rate equation was first proposed in 1930's and it was derived from thermodynamics (Nielsen, 1997; Visser and Heijnen, 2003). At first hand, it sounds odd that kinetic properties could be derived from thermodynamics. However, the dependency on thermodynamics can be explained with Michaelis-Menten kinetics by recalling that the rate limiting reactions step is $E \cdot S \rightarrow E + P$. Thus, the reaction rate dependents on the concentration of enzyme substrate complex ($E \cdot S$), concentration of which depends on the affinity of the substrate and the enzyme (a thermodynamic property). The log-linear kinetics is used to expand the applicability of metabolic control analysis (MCA). This method determines the elasticities of fluxes as a function of metabolite concentrations near steady-state conditions (Visser and Heijnen, 2003). Indeed, log-linear kinetics is determined relative to a steady-state, which makes it different from mechanistic and tendency kinetics. However, it is the method of choice for transient methods where a steady-state is disturbed with a metabolite pulse, a usual experiment in MCA. The mathematical form of a log-linear rate equation is shown below (Equation 3.4).

$$v(S, P, A) = e(k_1 + k_2 \ln S + k_3 \ln A + k_4 \ln P) \tag{3.4}$$

where $e$ is enzyme activity, $k_i$ are parameters, $S$, $A$ and $P$ are substrate, allosteric effector and product concentrations, respectively. The sign of the $k$ parameter determines whether a metabolite is a substrate, product, activator or inhibitor. In matrix form, the whole system can be described as

$$v = K \times \begin{bmatrix} 1 \\ \ln C_i \end{bmatrix} \tag{3.5}$$

where $v$ is rate vector, $K$ is the parameter matrix and the last vector has logarithms of the metabolites except in the first element. This system is linear in terms of $\ln C_i$ and relatively easy to solve in comparison to power-law functions of rational functions of mechanistic rate equations. Particularly power functions oppose some demand on computational resources. However, the limited applicability and steady-state nature of the model makes it a poor candidate for kinetic studies.

### 3.2.4 Arbitrary models

Basically any function can be used to model a reaction rate. The most obvious candidate is the linear model. In other words, a reaction rate is modeled with the linear combination of all metabolite concentrations. Naturally, many of the coefficients are zero, but this method could expose unknown effectors.

The arbitrary models have two important benefits. First of all, these models are highly structured and homogeneous such as the power-law model. This means that an algorithm can be developed to evaluate the reaction rates of a reaction system (and the partial derivatives thereof) based only on the stoichiometry of the system and a parameter pool. The mechanistic model is the other extreme; the rate equations are written out by hand which makes the coding exhaustive and prone to errors. The other benefit is that the simple arbitrary equations are swiftly evaluated by a computer. Addition, subtraction and multiplication are relatively fast to evaluate while division, logarithmic and power functions are slow. Thus, a model with short and simple equations evaluates faster than a model with longer and more complex equations. This is a significant property from the parameter estimation point of view.

In any case, extensive screening is required to find a new rate equation candidate. This is facilitated by the rapidly evaluated arbitrary models. However, arbitrary models may have limited range of applicability. Thus they may not be a realistic choice for some modeling applications, but their status as a model class should be recognized. Eventually, the question is which model produces smallest modeling error with the same computational effort.

## 3.3 Software

The kinetic models usually produce a system of ordinary differential equations (ODE). In order to solve these equations initial or boundary conditions have to be defined. This kind of mathematical problem is most common in engineering and there are numerous ways to solve it. Practically any numerical mathematics software offers routines to solve different kinds of ODE-systems. Additionally, to achieve best performance, the model may be built with a programming language of choice. This is made possible by the vast number of freely available ODE-solver libraries written in C or Fortran. Many of these libraries have algorithms originally presented in mathematical journals and they have been used for years to confirm efficiency and stability, thus giving the same or superior reliability as mathematical software. In most cases these libraries outperform the mathematical software, while the mathematical software is much easier to use.

There are also numerous software dedicated to simulation of (bio)chemical reaction networks such as Gepasi (Mendes, 1993) and DBSolve (Goryanin et al., 1999). Although these software are easy to use and look pretty, they are practically nothing more than ODE-solvers with a graphical user interface. Many of these programs have additional features such as parameter estimation (Mendes and Kell, 1998) and graphical representation of the results. The benefit of biochemical simulations software is that they facilitate reporting of findings and comparison of results since a reference to a software version and a method defines the details of the procedures. In any case, these programs are hard to integrate to external software and hard to change for testing. A short feature list usually limits their innovative use and these software are only suitable for day-to-day routine analysis. For the enthusiastic model developer general mathematical software is the best choice.

# Chapter 4

# Rate equations in mechanistic models

In this chapter the research on mechanistic rate equations is reviewed. The emphasis is put on those equations that most probably occur in the model. Most of this chapter is based on Cornish-Bowden (1999) and Kuby (1991). In the latter part of the chapter the complexities of different rate equations are compared.

The derivation of mechanistic rate equations is quite straightforward. First, there has to be a model of the reaction mechanism and the individual steps or elemental reactions it has. Second, each step is given its own rate equation based on chemical kinetics. For example, the rate equation for elemental reaction $A + B \rightleftharpoons C$, where the forward and reverse rate constants are $k_1$ and $k_{-1}$, respectively, is $k_1 AB - k_{-1} C$. Here we have assumed first order kinetics for all reactants which is a common assumption in enzyme kinetics. Then, there are two choices on how to continue:

1. Assume that the intermediates of the whole reaction are at steady-state. This is expressed in mathematics as $dC_i/dt = 0$ for all intermediates $i$. The reaction rate is then written as the time derivative of a product concentration which is equal to a negative time derivative of a substrate with equal stoichiometric coefficient.

2. Assume that an individual reaction is rate limiting. Thus, the other reactions are relatively fast and operate very close to equilibrium. Because of the equilibrium approximation, the concentrations of the intermediates can be calculated from the concentrations of the reactants and the rate constants of the elemental reactions. The rate equation for the limiting reaction is solved to give the overall rate equation.

Additional constraints are needed to solve the equations explained above. In catalytic reactions, and in enzyme kinetics in particular, it is commonly required that the total catalyst (enzyme) concentration is constant ($e_0$). In other words, the sum of the concentrations of the enzyme constituents is constant.

As the reaction mechanisms become more complicated than the Michaelis-Menten kinetics, direct application of this technique is not very productive and more advanced mechanisms

have to be developed. One such mechanism is the method of King and Altman described in Cornish-Bowden (1999) and Kuby (1991). The method itself is not explained here because of its complexity, but it is applied to the reaction mechanism of GAPDH in appendix A.

## 4.1 Michaelis-Menten rate equations

Michaelis-Menten equation is the best known enzyme kinetics equation among the biochemical community. The irreversible Michaelis-Menten rate equation corresponding to the following reaction mechanism

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \bullet A \overset{k_2}{\rightarrow} E + P \tag{4.1}$$

and assuming that the latter reaction is rate limiting, gives

$$v = \frac{k_2 e_0 A}{\frac{k_{-1}+k_2}{k_1} + A} = \frac{VA}{K_{mA} + A} \tag{4.2}$$

where $v$ is the reaction rate relative to the cell dry weight, CDW ($mmol\ mg(CDW)^{-1}\ min^{-1}$), $k_i$ are chemical rate constants ($min^{-1}$), $e_0$ is enzyme concentration ($mmol\ mg(CDW)^{-1}$), $V$ is the limiting reaction rate at substrate saturation ($mmol\ mg(CDW)^{-1}\ min^{-1}$), $A$, $P$ and $E$ correspond to substrate, product and enzyme concentrations ($mM$), respectively and $K_{mA}$ is the Michaelis constant ($mM$) that represents the substrate concentration that generates reaction rate half of $V$. The limiting rate $V$ is sometimes called $V_{max}$, but use of this term is discouraged by the Nomenclature Committee IUBMB because it is not a maximum in mathematical sense but a limit (Cornish-Bowden, 1999). Here the reaction rate and enzyme concentrations are given relative to the cell dry weight rather than volume. That is because molar amount per weight is more convenient for *in vivo* calculations and amount per volume suits better for *in vitro* calculations, where the cells are absent.

In reality, the irreversible rate equation is useful only for initial rate studies where product concentration is very low. The reversible Michaelis-Menten rate equation has wider applicability and a general form of equation 4.3.

$$v = \frac{e_0(k_A A - k_P P)}{1 + \frac{A}{K_{mA}} + \frac{P}{K_{mP}}} \tag{4.3}$$

where the definitions of the kinetic constants depend on the assumed reaction mechanism although the equation itself assumes no particular mechanism (Cornish-Bowden, 1999). For a reaction mechanism where a chemical reaction step between the association and dissociation steps is included (step 2 in equation 4.4)

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \bullet A \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E \bullet P \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P \tag{4.4}$$

the kinetic parameters of equation 4.3 have the following values (Cornish-Bowden, 1999).

$$k_A = \frac{k_1 k_2 k_3}{k_{-1}k_{-2}+k_{-1}k_3+k_2 k_3}; \quad K_{mA} = \frac{k_{-1}k_{-2}+k_{-1}k_3+k_2 k_3}{k_1(k_{-2}+k_2+k_3)}$$
$$k_P = \frac{k_{-1}k_{-2}k_{-3}}{k_{-1}k_{-2}+k_{-1}k_3+k_2 k_3}; \quad K_{mP} = \frac{k_{-1}k_{-2}+k_{-1}k_3+k_2 k_3}{k_{-3}(k_{-1}+k_{-2}+k_2)}$$

It is interesting to note that the mechanism of equation 4.4 has seven parameters altogether ($k_i$'s and $e_0$) while the Michaelis-Menten equation (Equation 4.3) has only five. This is probably due to the equilibrium assumptions. Recall that the rate equation is derived by assuming a rate limiting step, and that the other reactions are close to equilibrium. In this mechanism two reaction steps are assumed to be at equilibrium which eliminates the two parameters of the initial rate equations. This phenomenon is also observed in the other rate equations later on.

The numerator of equation 4.3 can also be written in the following way in order to introduce the equilibrium constant

$$e_0(k_A A - k_P P) = e_0\, k_A (A - \frac{k_P P}{k_A}) = V(A - \frac{P}{K_{eq}}) \tag{4.5}$$

where $K_{eq}$ is the product of all equilibrium constants of the individual reaction steps $k_i/k_{-i}$ which can be verified for the kinetic parameters of the previous paragraph.

The two-substrate two-product Michaelis-Menten rate equation is very similar to the one substrate one product equation. For reaction

$$A + B \rightleftharpoons P + Q \tag{4.6}$$
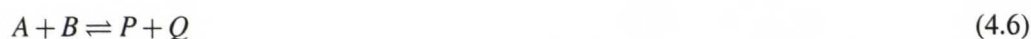
the reversible Michaelis-Menten rate equation according to Teusink *et al.* (2000) is

$$v = \frac{\frac{V}{K_{mA} K_{mB}}(AB - \frac{PQ}{K_{eq}})}{(1 + \frac{A}{K_{mA}} + \frac{P}{K_{mP}})(1 + \frac{B}{K_{mB}} + \frac{Q}{K_{mQ}})} \tag{4.7}$$

These authors did not derive the equation from any assumptions but merely presented it. However, the same equation has also been used by Hoefnagel *et al.* (2002b).

## 4.2  Two-substrate two-product rate equations

The group of two-substrate two-product enzymatic reactions is of special importance since 60 % of the enzyme catalyzed reactions belong to this class (Cornish-Bowden, 1999). There are a number of ways how four metabolites can interact at the active site of an enzyme. These mechanism can be divided in two groups. There are reactions that obey ternary-complex kinetics and reactions that obey substituted enzyme kinetics (Cornish-Bowden, 1999). A ternary-complex mechanism propagates through a complex of the enzyme and both substrates, hence the name. This mechanism can be further divided into mechanisms where the binding of the substrates is ordered or random (Sections 4.2.1 and 4.2.2, respectively). The substituted enzyme mechanism propagates through an enzyme intermediate that is covalently bound to a molecule being transferred from substrate to the product (Section 4.2.3). This mechanism is called the ping-pong mechanism by Kuby (1991) to emphasize the fact that the substrates react in turns.
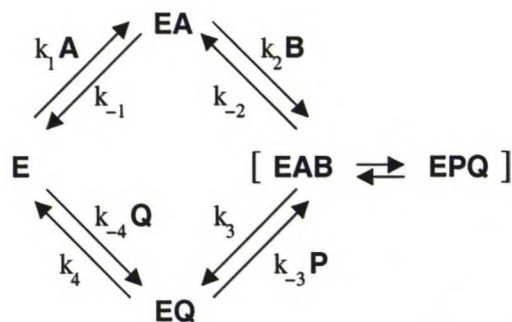
**Figure 4.1:** Diagram of the ordered sequential reaction mechanism. Redrawn following Kuby (1991).

### 4.2.1 Ordered sequential rate equation

In ordered binding mechanism the free enzyme has a binding site only for the first substrate. Binding of the substrate induces a structural transformation that enables the binding of the second substrate. Also release of the products is in most cases ordered so that the second substrate is released first. There are, however, exceptions. For the whole reaction of $A + B \rightleftharpoons P + Q$, where $Q$ is the product of $A$ and $P$ is the product of $B$, the assumed mechanism is shown in figure 4.1. As the figure indicates, the interconversion of the substrates to products is assumed to be at equilibrium. This does not, however, effect the form of the rate equation (Equation 4.8), but it does change the values of the kinetic parameters ($K_i$'s) in terms of chemical rate constants ($k_i$'s) (Kuby, 1991). The rate equation of an ordered sequential reaction is shown below (Equation 4.8). This equation is derived by the method of King and Altman and the details can be found elsewhere (Cornish-Bowden, 1999; Kuby, 1991).

$$
v = \frac{\frac{V}{K_{iA}K_{mB}}\left(AB - \frac{PQ}{K_{eq}}\right)}{\begin{array}{l} 1 + \frac{A}{K_{iA}}\left(1 + \frac{B}{K_{mB}}\left(1 + \frac{P}{K_{iP}}\right)\right) + \frac{B}{K_{mB}}\left(1 + \frac{K_{mA}Q}{K_{iA}K_{iQ}}\right) + \\ \frac{Q}{K_{iQ}}\left(1 + \frac{P}{K_{mP}}\left(1 + \frac{B}{K_{iB}}\right)\right) + \frac{K_{mQ}P}{K_{iQ}K_{mP}}\left(1 + \frac{A}{K_{iA}}\right) \end{array}}
\tag{4.8}
$$

The values of the kinetic parameters can be found with the other details. However, the values of the kinetic parameters in terms of chemical rate constants are very sensitive to the actual mechanism of the reaction (Cornish-Bowden, 1999). Thus, failing to fulfill the assumptions yields different values for the kinetic parameters, although the structure of the rate equation is the same. Therefore, equation 4.8 is significant even though the parameters are considered arbitrary.

### 4.2.2 Random order rate equation

In the random order binding, the free enzyme can first bind with either one of the substrates. The corresponding reaction mechanism is shown in figure 4.2. Kuby (1991) and Cornish-Bowden (1999) both derive the rate equation for this mechanism by assuming all steps but the interconversion of the ternary-complex (step 5) to be at equilibrium. Thus, reaction five in
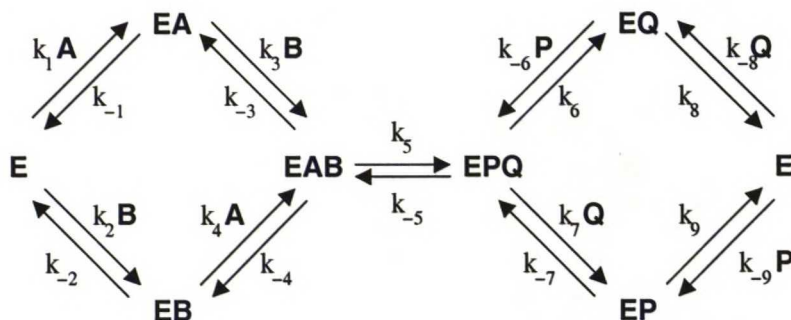
**Figure 4.2:** Diagram of the random order reaction mechanism. Redrawn following Kuby (1991).

figure 4.2 is the rate limiting step. The rate equation (Equation 4.9) would be somewhat more complicated if the rapid-equilibrium assumption did not hold. For example, second order terms would emerge. The rapid-equilibrium assumption also causes that the equation is more simple than the rate equation for the ordered binding mechanism (Equation 4.8).

$$v = \frac{\frac{V}{K_{iA}K_{mB}}\left(AB - \frac{PQ}{K_{eq}}\right)}{1 + \frac{A}{K_{iA}}\left(1 + \frac{B}{K_{mB}}\right) + \frac{B}{K_{iB}} + \frac{Q}{K_{iQ}}\left(1 + \frac{P}{K_{mP}}\right) + \frac{P}{K_{iP}}} \tag{4.9}$$

### 4.2.3 Ping-pong kinetics

The substituted enzyme mechanism propagates in two steps as shown in figure 4.3. The name "ping-pong" emphasizes that substrates react with the enzyme one at a time. That is, first substrate binds to the enzyme, reacts to give away the transferred group and is subsequently released. Only then the second substrate binds to the enzyme, receives the transferred group and is released. The binding site for both substrates is probably the same and no ternary-complex can occur. In the ping-pong mechanism it is not straightforward to define which product originates from which substrate. As figure 4.3 indicates, the mechanism only defines the order in which the substrates are associated and products are released. It could be asked whether a major or definitive part of A is covalently associated with the enzyme or released with P. The answer implies whether P should be considered the product of A or not. However, for the rate equation (Equation 4.10) that has no effect. The rate equation can be derived by the method of King and Altman (Kuby, 1991; Cornish-Bowden, 1999). The rate equation is shown below (Equation 4.10).

$$v = \frac{\frac{V}{K_{iA}K_{mB}}\left(AB - \frac{PQ}{K_{eq}}\right)}{\frac{A}{K_{iA}}\left(1 + \frac{B}{K_{mB}}\right) + \frac{K_{mA}B}{K_{iA}K_{mB}}\left(1 + \frac{Q}{K_{iQ}}\right) + \frac{P}{K_{iP}}\left(1 + \frac{A}{K_{iA}}\right) + \frac{Q}{K_{iP}K_{mQ}}\left(K_{mP} + P\right)} \tag{4.10}$$

It is noteworthy that the denominator does not have a constant term like the ternary-complex rate equations (Equations 4.8 and 4.9). Although detailed analysis of the rate equation at the limit of zero concentration is of little importance, the nature of the equation may well be different from the ternary-complex equations. For example, in the absence of constant term in the denominator one could argue that the rate equation is more sensitive to concentration changes
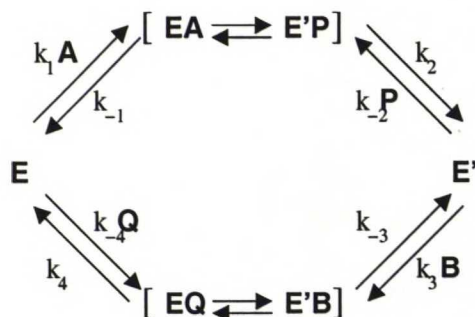
**Figure 4.3:** Diagram of the substituted enzyme (ping-pong) reaction mechanism. Redrawn following Kuby (1991).

especially when all concentrations are low relative to the corresponding Michaelis constants. This could be an important functional feature of ping-pong kinetics. It has been shown that pyruvate formate lyase (PFL), an important regulator of the fermentation mode, exhibits ping-pong kinetics (Knappe and Blaschkowski, 1975). Although this enzyme is allosterically regulated by GAP and DHAP (see section 2.5), ping-pong kinetics may also contribute to the onset of mixed-acid fermentation.

## 4.3 Miscellaneous rate equations

### 4.3.1 Cooperative kinetics

Cooperative kinetics arise from the assumption that multiple active sites of multi-subunit enzymes interact with each other. This type of enzyme activity is no longer described by the simple Michaelis-Menten kinetics. Instead, an equation 4.11, nowadays known as Hill equation, was developed (Cornish-Bowden, 1999). The irreversible one substrate Hill equation is shown below.

$$v = \frac{VA^h}{K_{0.5}^h + A^h} \tag{4.11}$$

This equation greatly resembles Michaelis-Menten kinetics but there are few important differences. The parameter $V$ has the same meaning as in the other rate equations. Parameter $K_{0.5}$ is similar to $K_m$; it indicates the substrate concentration causing reaction rate half of $V$. Therefore, $K_m$ should never be used instead of $K_{0.5}$ because $K_{0.5}$ specifically refers to the cooperative kinetics of the Hill equation. The parameter $h$, Hill coefficient, is a measure of cooperativity. Hill coefficient is purely empirical and it has no physical interpretation (Cornish-Bowden, 1999). Cooperation is said to be negative if $h < 1$ and positive if $h > 1$ although negative cooperation is rare. Figure 4.4 indicates the effects of positive and negative cooperation compared to the non-cooperative Michaelis-Menten kinetics. The Hill equation is introduced since pyruvate kinase (PK) from *Streptococcus lactis* exhibits cooperative kinetics (Crow and Pritchard, 1982).
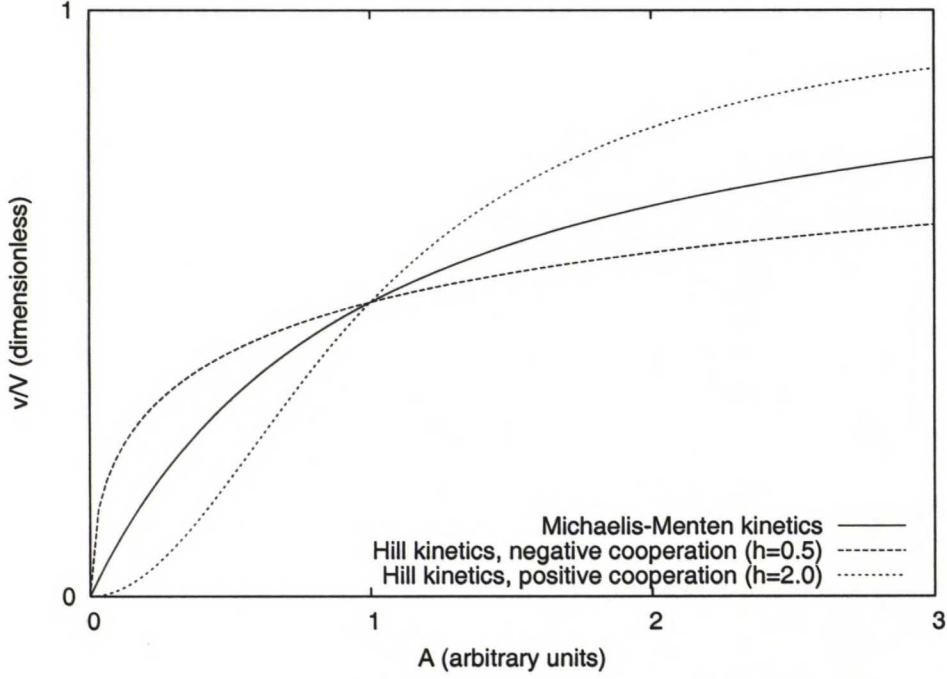
**Figure 4.4:** Comparison of Michaelis-Menten and cooperative Hill kinetics. Equation 4.11 was used with $K_{0.5} = 1.0$. Michaelis-Menten kinetics is a special case of Hill kinetics with $h = 1$ (no cooperation).

The two-substrate cooperative rate equation is presented in equation 4.12. The rate equation has 5 parameters altogether.

$$v = \frac{V A^{h_a} B^{h_b}}{\left(K_{0.5,a}^{h_a} + A^{h_a}\right)\left(K_{0.5,b}^{h_b} + B^{h_b}\right)} \tag{4.12}$$

### 4.3.2 Glyceraldehyde phosphate dehydrogenase rate equation

The rate equation for glyceraldehyde phosphate dehydrogenase (GAPDH) has imposed some problems. The whole reaction is

$$GAP + NAD^+ + P_i \rightleftharpoons 1,3PG + NADH \tag{4.13}$$

and it is catalyzed by a tetrameric enzyme with four active sites. The supposed reaction mechanism (Figure 4.5) by Nagradova (2001) does not correspond to any of the previously defined basic types. Therefore, a rate equation specific to that mechanism was derived by the method of King and Altman (Cornish-Bowden, 1999; Kuby, 1991). The details of the derivation are given in appendix A. The resulting rate equation is given below (Equation 4.14).

$$v = \frac{e_0\left(N_1[\text{GAP}][\text{NAD}^+][\text{Pi}] - N_2[1,3\text{PG}][\text{NADH}]\right)}{\begin{aligned}&[\text{GAP}]\left(D_1 + D_4[\text{NAD}^+] + D_5[\text{NADH}] + [\text{Pi}]\left(D_6 + D_{11}[\text{NAD}^+] + D_{12}[\text{NADH}]\right)\right) + \\ &[1,3\text{PG}]\left(D_2 + D_7[\text{NAD}^+] + D_8[\text{NADH}] + [\text{Pi}]\left(D_{13}[\text{NAD}^+] + D_{14}[\text{NADH}]\right)\right) + \\ &[\text{NADH}]\left(D_3 + D_{10}[\text{Pi}]\right) + D_9[\text{NAD}^+][\text{Pi}]\end{aligned}}$$

$$\begin{bmatrix} \text{E NAD}^+ \text{ GAP} \\ \text{E-GAP NADH} \end{bmatrix}$$

$k_1\text{GAP}$   $k_{-1}$   $k_2$   $k_{-2}\text{NADH}$

$\text{E NAD}^+$   $\text{E-GAP}$

$k_5$   $k_{-5}\text{1,3PG}$   $k_3\text{NAD}^+$   $k_{-3}$

$\text{E-GAPNAD}^+ \text{Pho}$   $\xrightarrow{k_4\text{Pho}}$ $\text{E-GAPNAD}^+$

$k_{-4}$
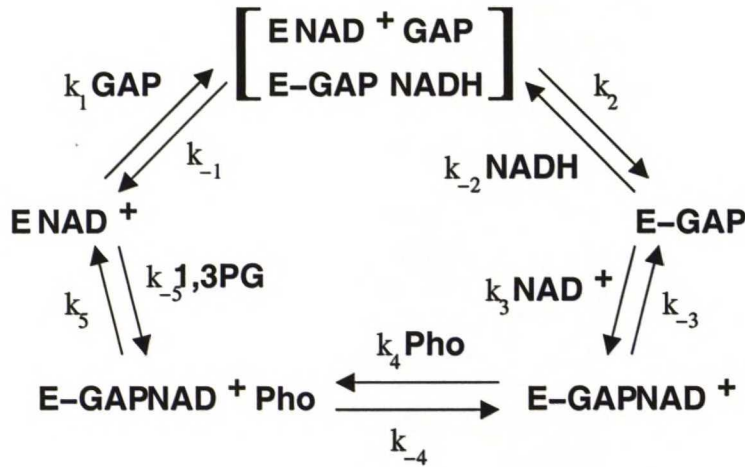
**Figure 4.5:** Reaction mechanism of GAPDH according to Nagradova (2001). Dash between E and GAP represents a covalent bond binding the enzyme with the transferred group.

$$(4.14)$$

where $e_0$ is the enzyme concentration ($mmol\ mg(CDW)^{-1}$) and $N_i$'s and $D_j$'s are coefficients of the numerator and the denominator, respectively (see Appendix A). Although this equation has 17 parameters ($e_0$, $N_i$'s and $D_j$'s) they depend only on 11 rate constants ($e_0$ and $k_i$'s). To ease the parameter estimation it is logical to calculate the values of $N_i$'s and $D_j$'s in terms of $k_i$'s as presented in appendix A. Thus, the equation has 11 free parameters.

This rate equation is structurally similar to the ping-pong rate equation in the sense that the denominator does not have a constant term. In fact, the mechanism is a substituted enzyme mechanism since it has an enzyme intermediate that is covalently bound to a transferred ligand (E-GAP). The mechanism is, though, more complicated than the one discussed in section 4.2.3.

Another important feature of the reaction mechanism is that there is no free enzyme in it. Obviously, if substrate and product concentrations approach zero the bound ligands would dissociate from the enzyme (except the covalently bound GAP) giving only the bare enzyme. Similarly, in an enzyme assay, for example, $NAD^+$ only binds to E-GAP, but binding of GAP requires bound $NAD^+$. This is a chicken and an egg problem. Therefore, it must be realized that this mechanism does not hold under all conditions, but describes the reaction cycle only.

Also another simplification is made. Nagradova (2001) has reported that GAPDH of *E. coli* exhibits so called half-of-the-sites reactivity, which means that the four catalytic sites of the enzyme are not equal due to structural asymmetry. Instead, they have negative cooperation; The first two $NAD^+$ bind with high affinity while the third and fourth have significantly lower affinity (Nagradova, 2001). However, Nagradova (2001) reports that the half-of-the-sites reactivity has not been observed in kinetic experiments which justifies the neglect of negative cooperation.

Despite the simplifications, this rate equation is more appropriate than the one previously

used. For example Teusink *et al.* (2000) used a reversible two-substrate two-product Michaelis-Menten equation for *S. cerevisiae* GAPDH and Hoefnagel *et al.* (2002a) have used presumably a three substrate two product Michaelis-Menten equation[1] for *L. lactis* GAPDH that is also different from equation 4.14. In order to make a model work correctly over a wide range of conditions, mechanistic rate equations should be used. This justifies the use of equation 4.14 although it has more parameters to be estimated than simple Michaelis-Menten equations.

## 4.4  Membrane transport rate equation

Passive carrier mediated membrane transport is modeled with a rate equation derived by Weiss (1995) for a standard four-state translocation membrane protein. This transport method is called facilitated diffusion because of the role of the carrier protein. In the derivation of the rate equation the author has assumed that the dissociation and association of the solute to the carrier is fast compared to the translocation step and it is thus assumed to be at equilibrium. Furthermore, the conformation change of the carrier from the outter to the inner surface is assumed to be the rate limiting step. A steady-state derived rate equations reads

$$v = \frac{V(A - P)}{K + A + P + \frac{A \cdot P}{K}} \tag{4.15}$$

where species $A$ and $P$ correspond to the intra- and extracellular pools of the same metabolite, $V$ is a limiting rate and $K$ is an equilibrium constant of the solute binding to the carrier. Even though only few of the transport processes modeled with this equations are known to be protein mediated, this equation is used for the sake of simplicity. However, purely diffusion driven transport of neutral species would have only one parameter (the membrane imposed transport barrier i.e. the rate constant term). This equation can mimic the non-enzymic behavior with $K$ much larger than $A$ or $P$. The parameter estimation process is thus responsible for choosing the right mode of operation.

## 4.5  Fructose bisphosphate aldolase rate equation

The FBA catalyzes the cleavage of six-carbon FBP into three-carbon products GAP and DHAP. No reference to the FBA rate equation could be found in the literature. However, implications of the mechanism have been introduced by Szwergold *et al.* (1995). The authors repoted that, based on inhibition experiments, the product release of *Escherichia coli* FBA is ordered and that GAP is released before DHAP. Because the nature of this mechanism greatly resembles the ordered two-substrate two-product mechanism it was decided that the rate equation would be derived from equation 4.8 by neglecting the second substrate. This method is suboptimal and was used due to time constraints.

---

[1]See http://jjj.biochem.sun.ac.za/database/test/index.html

The rate equation was derived by setting $B$ to zero for the denominator of equation 4.8 and to one for the numerator. This method is believed to save the ordered properties of the rate equation with minimal time consumption. The derived rate equation for FBA is given below (Equation 4.16).

$$v_{FBA} = \frac{\frac{V}{K_{m,FBP}}\left(FBP - \frac{GAP\,DHAP}{K_{eq}}\right)}{1 + \frac{FBP}{K_{m,FBP}}\left(1 + \frac{GAP}{K_{i,GAP}}\right) + \frac{DHAP}{K_{i,DHAP}} + \frac{K_{i,FBP}GAP}{K_{m,FBP}K_{i,GAP}}\left(1 + \frac{DHAP}{K_{m,DHAP}}\right)} \tag{4.16}$$

## 4.6 Rate equations by Hoefnagel *et al.*

The rate equations for the reactions of the pyruvate branch have been presented earlier by Hoefnagel *et al.* (2002b). The previously presented equations have been used where no other propositions for rate equations were available. The equations directly adopted from Hoefnagel *et al.* (2002b) are presented below.

### 4.6.1 Pyruvate dehydrogenase rate equation

The rate equation for PDH (Equation 4.17) is irreversible due to the production of $CO_2$ which evaporates rapidly from the medium and is hardly available for the reverse reaction. The rate equation is a Michaelis-Menten equation with non-competing substrate-product couples (Hoefnagel *et al.*, 2002b).

$$v_{PDH} = \frac{V\left(\frac{1}{1+K_i\frac{NADH}{NAD^+}}\right)\left(\frac{Pyr}{K_{m,Pyr}}\right)\left(\frac{NAD^+}{K_{m,NAD^+}}\right)\left(\frac{CoA}{K_{m,CoA}}\right)}{\left(1 + \frac{Pyr}{K_{m,Pyr}}\right)\left(1 + \frac{NAD^+}{K_{m,NAD^+}} + \frac{NADH}{K_{m,NADH}}\right)\left(1 + \frac{CoA}{K_{m,CoA}} + \frac{AcCoA}{K_{m,AcCoA}}\right)} \tag{4.17}$$

where $V$ is the limiting rate and the $K_m$'s are the Michaelis constants for the reactants. This rate equation also includes inhibition term $K_i$ for high $NADH/NAD^+$-ratio.

### 4.6.2 Acetaldehyde dehydrogenase rate equation

The ACDH rate equation (Equation 4.18) is a reversible Michaelis-Menten equation that is similar to the other rate equations with non-competing substrate-product couples. However, the $ACAL - CoA$ cross term in the denominator implies some interaction between the binding sites of the two reactants.

$$v_{ACDH} = \frac{V\left(\frac{1}{K_{m,AcCoA}K_{m,NADH}}\right)\left(AcCoA\,NADH - \frac{CoA\,NAD^+\,ACAL}{K_{eq}}\right)}{\left(1 + \frac{NAD^+}{K_{m,NAD^+}} + \frac{NADH}{K_{m,NADH}}\right)\left(1 + \frac{AcCoA}{K_{m,AcCoA}} + \frac{CoA}{K_{m,CoA}} + \frac{ACAL}{K_{m,ACAL}} + \frac{ACAL\,CoA}{K_{m,ACAL}K_{m,CoA}}\right)} \tag{4.18}$$

where the parameter have the usual meaning. $ACAL$ is abbreviation for acetaldehyde.

## 4.7 Complexity

Complexity is used in modeling to fulfill the Occam's razor principle[2] which states that the simplest theory that describes the observations should be chosen. Complexity is a penalty term that can be included in model class selection process. An example of such a process is to determine the proper degree for a polynome to model a given data set. Obviously, one does not want a fourth degree model if the data set is of third degree, even if the fourth degree model had smaller modeling error.

It this thesis, model class is a group of rate equations selected to model individual reactions. Thus, complexity could be used to select the simplest set of rate equations that describes the observed data. In practice, however, this approach may be computationally too demanding. It can be used to evaluate the effect of changing an individual rate equation. Especially, the statistical parameter estimation processes are significantly facilitated if the number of parameters can be decreased.

Defining an equation for complexity is a matter of choice. Of course, the equation should be tested in order to verify that it behaves as expected as a part of the model class selection process. In enzyme kinetics a rate equation could be compared to a normal type equation. This normal type could have, say, one parameter per each reactant, a parameter for enzyme activity and for equilibrium. Thus, the proposed equation for the relative complexity of biochemical rate equations $\mathbb{C}$ reads

$$\mathbb{C}(N_p, N_r) = \frac{N_p}{N_r + 2} \tag{4.19}$$

where $N_p$ and $N_r$ are the numbers of parameters and reactants, respectively. The denominator describes the number of parameters in a corresponding normal rate equation. Table 4.1 lists the complexities of the rate equations used in this thesis. Note that the normal equation corresponds to reversible Michaelis-Menten kinetics, regardless of the number of reactants. Also, the rate equations derived by the method of King and Altman have higher complexities because of the inhibition constants they have. The complexities of the ternary-complex mechanisms seem to indicate that low complexity implies more freedom in reaction mechanism; Complexities for the random order and ordered mechanisms are 1.33 and 1.67, respectively.

Notice also that reversible power-law kinetics are equally complex as Michaelis-Menten kinetics. As described in section 3.2.2 they have one parameter for each reactant and two kinetic parameters, one for each direction. Interestingly, not all inhibition schemes add another parameter in power-law kinetics. If one substrate inhibits the forward reaction by preventing another substrate from binding, the positive and negative order parameters can be combined to one. By analogy, the same holds for the reverse reaction with the obvious changes.

The relative complexity, as it is defined here, is a feature of a rate equation. For the whole model a total complexity or an average complexity can be calculated. Models with different

---

[2]See, for example, http://pespmc1.vub.ac.be/OCCAMRAZ.html

**Table 4.1:** Comparison of the complexities of the rate equations. Complexity is defined in equation 4.19. The rate equations are ordered according to increasing complexity.

| Rate equation | Number of | | Complexity |
|---|---|---|---|
| | parameters | reactants | |
| Carrier mediated transport | 2 | 2 | 0.50 |
| Reversible one-substrate one-product Michaelis-Menten | 4 | 2 | 1.00 |
| Reversible two-substrate two-product Michaelis-Menten | 6 | 4 | 1.00 |
| Reversible power-law kinetics, N reactants | N+2 | N | 1.00 |
| PDH rate equation | 7 | 5 | 1.00 |
| ACDH rate equation | 7 | 5 | 1.00 |
| Sigmoidal irreversible two-substrate kinetics | 5 | 2 | 1.25 |
| Random order bi-bi equation | 8 | 4 | 1.33 |
| FBA rate equation | 7 | 3 | 1.40 |
| Ping-pong bi-bi equation | 9 | 4 | 1.50 |
| GAPDH rate equation | 11 | 5 | 1.57 |
| Ordered sequential bi-bi equation | 10 | 4 | 1.67 |

pathways should be compared in total complexities while models with the same reactions but different rate equations can also be compared in average complexities. The total complexities of the mechanistic and power-law models used in this thesis are 30.23 and 27.50, respectively.

# Chapter 5

# Previous models of *L. lactis*

A kinetic model of glycolysis in *L. lactis* has been published earlier by Hoefnagel *et al.* (2002a). However, the report gives a rather superficial description of the model and the reader is directed to an Internet site for details (`http://jjj.biochem.sun.ca.az/`). This site gives the rate equations and the parameters used in the model, but no references are presented.

However, the pyruvate branches of the same model are properly presented in another article (Hoefnagel *et al.*, 2002b) and it is also available for testing on the aforementioned Internet site. This model studies the control of carbon flux at the level of pyruvate and it includes the lactate, butanediol, acetoin, acetate and ethanol branches (Figure 1 of Hoefnagel *et al.*, 2002b). Glycolytic reactions from glucose to pyruvate and anabolic ATP consumption are modeled as single reactions having parameters fitted to experimental data. Rest of the model has mechanistic rate equations with parameters from the literature. In addition to the wild-type organism, the authors have also modeled and characterized three mutants; a lactate dehydrogenase knockout mutant, a mutant overexpressing NADH oxidase and a combination of the two. The authors assumed that the culture is at pseudo steady-state at the exponential growth phase and compared the modeled fluxes of the pyruvate branches to the measured ones. They also used the kinetic model to calculate flow-control coefficients for each enzyme to determine which ones should be mutated.

The predicted fluxes of pyruvate branches match reasonably well with the measured ones, which is true for the mutants as well. This is interesting because no changes in transcription levels are assumed although genes have been knocked out or overexpressed. Intuitively, changing expression of one gene could have some effect on the expression of the others too, but changes to the expression of less significant genes may go unnoticed. It is possible this caused the noted differences between predictions and measurements. On the other hand, metabolism may be so well controlled, that it tries to maintain a constant state, i.e. homeostasis. Thus, unidentified control mechanisms may oppose changes in metabolic state although large genetic changes have been made.

However, even though the parameters are from the literature, there is a considerable number of

parameters for which no references are presented or which are said to be guessed. However, the origin of these parameters might as well be a fitting process; algorithmic or manual.

Hoefnagel *et al.* (2002b) presented a solid computational procedure to successfully apply a kinetic model in metabolic engineering of *L. lactis*. This procedure, using a kinetic model for flow-control coefficient calculations and mutant characterization, proves that *in silico* biology does provide great potential for the development of novel production hosts in biotechnological industry. Although many other successful examples exists, this one has most significance in the scope of this thesis.

Recently, a similar computational study of *L. lactis* metabolic networks was published by Nam *et al.* (2004). In this study the authors modeled the whole glycolytic and pyruvate pathways in order to find the most significant enzymes for lactate production. The model is an almost identical reproduction of the Hoefnagel *et al.* (2002b) and Hoefnagel *et al.* (2002a) models available from http://jjj.biochem.sun.ca.az/. The only difference is that some reactants and activators have been neglected from few reactions. With this kinetic model the authors solved steady-state fluxes with GEPASI (Mendes, 1993). In order to find the enzymes that had the greatest effect on lactate dehydrogenase (LDH) flux Nam *et al.* (2004) calculated the flux control coefficients of all enzymes for LDH. For the promising candidates, they plotted the LDH flux as a function of the enzyme activity of the candidates.

The flux control coefficients indicated that glucokinase (GLK), pyruvate kinase (PK), pyruvate dehydrogenase (PDH), lactate dehydrogenase (LDH), adenosine triphosphatase (ATPase) and NADH oxidase (NOX) had the greatest effect on the LDH flux. The figures that described the change in LDH flux in respect to the listed enzymes showed that actually only PDH, LDH and NOX had a significant effect on the LDH flux. This is an expected result since all these enzymes control the $NADH/NAD^+$ balance.

The study of Nam *et al.* (2004) has, however, several problems. The $NADH/NAD^+$ stoichiometries of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), LDH and acetoin dehydrogenase (ACETDH) catalyzed reactions are presented the wrong way around in figure 1 of Nam *et al.* (2004). The discussion in the text suggests that the GAPDH catalyzed reaction is just a typing error, but at least the stoichiometry of the LDH catalyzed reaction is incorrect. However, the rate equations copied from other publications are correct in terms of $NADH/NAD^+$ stoichiometry. Additionally, the ATP generation and consumption presented in the text do not meet the flux values of PYK, ACK and PFK presented in figure 2 of Nam *et al.* (2004). According to the values in the figure, ATP consumption and generation is not at steady-state. In general, the main problem of the work is that the authors have no experimental data to support their model. In addition, they have not questioned the *in vitro* parameters at any point. These two problems are linked; had there been data available Nam *et al.* (2004) could have fitted the parameters and seen whether the *in vitro* parameters gave a reasonably small modeling error. The Newton method applied for finding the steady-state finds only one steady-state depending on the initial conditions. Thus, the quality of the results depends heavily on the initial conditions, a fact that is not mentioned in the article. The authors should have presented the metabolite

concentrations of the steady-state to see if they are reasonable. Since these concentrations were not presented, it may be assumed that they do not give any additional confidence to the model.

Overall, the article of Nam *et al.* (2004) gives little input to the research of its field. The model is simply a reproduction of the previous work by others and the results are obvious. The concept and strategy of the work is, however, straightforward and serves as a good example of the applications that *in silico* studies may have.

# Part II

# Experimental part

# Chapter 6

# Materials and methods

## 6.1 Biological system

### 6.1.1 Organism

The simulations aim to mimic the behavior of *L. lactis* strains with and without xylose reductase (XR) activity. These strains are named LLXR and LLCON, respectively. In the simulated phase the bacteria are in resting state. It means that the bacteria are non-growing, this is achieved by feeding the culture with only a buffer solution with glucose and xylose. The benefit of this action is that it is not necessary to model reactions involved in cell growth. Specific details of the strains and the genetic modifications are given in appendix B.

The structure of the metabolic network used in this thesis is shown in figure 6.1. The enzymes catalyzing these reactions are listed in table 6.1. Organism and reference fields identify the source of reaction mechanism or the mechanistic rate equation used for each enzyme. In the model of Hoefnagel *et al.* (2002b) parameters from multiple organisms have been introduced to the same rate equation. Therefore, the organism could not be specified. The network consists of 34 metabolites and 26 reactions of the glycolytic pathway. In addition to the glycolytic pathway, there is also an engineered enzyme activity, XR, which additionally requires four metabolites ($Xylose, XLS, XLT$ and $Xylitol$) and three reactions (xylose transport, XR and xylitol transport).

### 6.1.2 Culture conditions

The detailed growth conditions are described in appendix B while a more generic description is given here listing the factors that somehow affect the mathematical procedures. The details of the devices and method used in the conversion experiments are also given in appendix B.

The experiments were conducted in a conversion medium that contains glucose and xylose and a minimum amount of other nutrients. The nutrient level was kept low enough to prevent
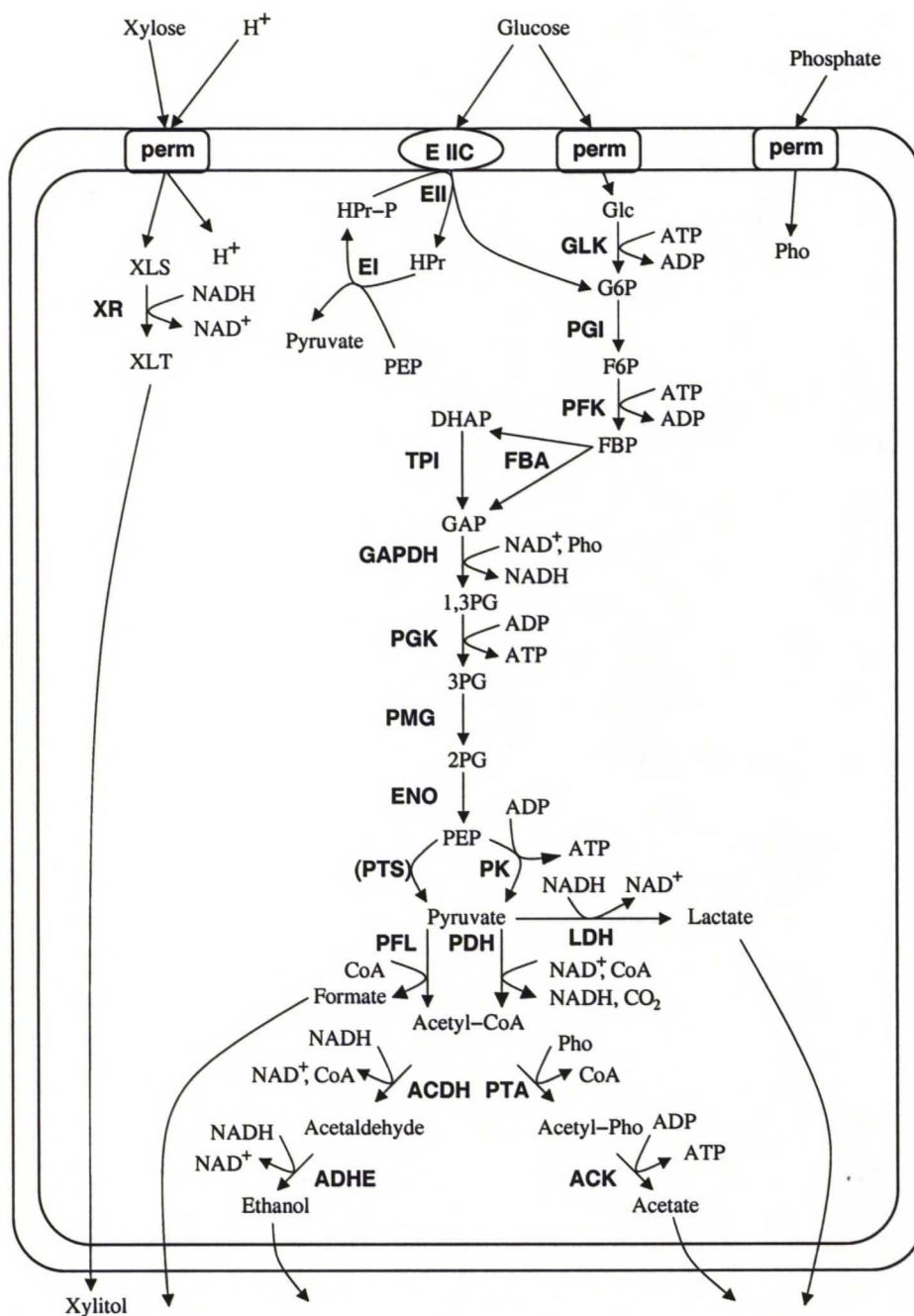
**Figure 6.1:** Target network of this study. Arrows represent the directions of the positive reaction rate the same way as they will be described in the stoichiometric matrix. They do not reflect the reversibility of the reactions. See Figure 2.2 for details of PTS. Abbreviations of the reactions are presented in table 6.1 and for metabolites; *Glucose*: external glucose, *Glc*: internal glucose, *G6P*: glucose-6-phosphate, *F6P*: fructose-6-phosphate, *FBP*: fructose-1,6-diphosphate, *DHAP*: dihydroxyacetonephosphate, *GAP*: glyceraldehyde-3-phosphate, $1,3PG$: 1,3-diphosphoglycerate, $3PG$: 3-phosphoglycerate, $2PG$: 2-phosphoglycerate, *PEP*: phosphoenolpyruvate, *CoA*: coenzyme-A, *Phosphate*: external phosphate, *Pho*: internal phosphate, $NAD^+$ and *NADH*: nicotinamide adenine dinucleotide oxidized and reduced forms, *ATP*: adenosine-triphosphate, *ADP*: adenosine-diphosphate, $CO_2$: carbon dioxide, *XLS*: internal xylose, *Xylose*: external xylose, *XLT*: internal xylitol and *Xylitol*: external Xylitol.

**Table 6.1:** Enzymes included in the model and source of kinetic properties. The Hoefnagel model used parameters from different organism in the same rate equation. Therefore the organism is not specified here. Xylose reductase is not a part of the core glycolytic model. Permease transporters for glucose, xylose, phosphate, xylitol, lactate, formate, acetate and ethanol are excluded from this table.

| Abbr. | Enzyme | EC | Organism | Reference |
|---|---|---|---|---|
| EI | Mannose/Glucose PTS Enzyme I | 2.7.3.9 | *Streptococcus salivarius* | Vadeboncoeur *et al.* (1983) |
| EII | Mannose/Glucose PTS Enzyme II | 2.7.1.69 | *Streptococcus faecalis* | Hüdig and Hengstenberg (1980) |
| GLK | Glucokinase | 2.7.1.2 | *Streptococcus mutans* | Porter *et al.* (1982) |
| PGI | Glucose phosphate isomerase | 5.3.1.9 | *Lactobacillus casei* | Pradhan and Nadkarni (1980) |
| PFK | Phosphofructokinase | 2.7.1.11 | *Streptococcus thermophilus* | Simon and Hofer (1981) |
| FBA | Fructose bisphosphate aldolase | 4.1.2.13 | *E. coli* | Callens *et al.* (1991), Szwergold *et al.* (1995), Baldwin *et al.* (1978) |
| TPI | Triosephosphate isomerase | 5.3.1.1 | Brewer's Yeast | Krietsch (1975) |
| GAPDH | Glyceraldehyde phosphate dehydrogenase | 1.2.1.12 | *E. coli* | Eyschen *et al.* (1999) |
| PGK | Phosphoglycerate kinase | 2.7.2.3 | Yeast | Scopes (1978) |
| PMG | Phosphoglyceromutase | 5.4.2.1 | *Saccharomyces cerevisiae* | Anonymous (2003) |
| ENO | Enolase | 4.2.1.11 | *E. coli* | Spring and Wold (1971) |
| PK | Pyruvate kinase | 2.7.1.40 | *Streptococcus lactis* and yeast | Crow and Pritchard (1982), Fenton and Blair (2002) |
| LDH | Lactate dehydrogenase | 1.1.1.27 | see original paper | Hoefnagel *et al.* (2002b) |
| PFL | Pyruvate formate lyase | 2.3.1.54 | *E. coli* | Knappe and Blaschkowski (1975) |
| ACDH | Acetaldehyde dehydrogenase | 1.2.1.10 | see original paper | Hoefnagel *et al.* (2002b) |
| ADHE | Alcohol dehydrogenase | 1.1.1.1 | see original paper | Hoefnagel *et al.* (2002b) |
| PDH | Pyruvate dehydrogenase | 1.2.1.51 | see original paper | Hoefnagel *et al.* (2002b) |
| PTA | Phosphotransacetylase | 2.3.1.8 | see original paper | Hoefnagel *et al.* (2002b) |
| ACK | Acetate kinase | 2.7.2.1 | see original paper | Hoefnagel *et al.* (2002b) |
| XR | Xylose reductase | 1.1.1.21 | *Pichia stipitis* | Verduyn *et al.* (1985) |

cells from growing but allowing maintenance and repair activities, thus preventing the culture from dying. The amount of biomass was assumed to be constant during the experiments based on earlier experiments with these strains (Nyyssölä *et al.*, 2005). The atmosphere was kept anaerobic during the conversion experiment with nitrogen gassing into the head-space of the reactor. At biomass growth stage atmosphere may have been aerobic. Temperature and pH of the batch were also maintained constant (at 30°C and 6.5, respectively) to standardize the conditions.

## 6.2 Computational setup

### 6.2.1 ODE-system

Simulating the kinetic behavior of a metabolic network involves solving a system of ordinary differential equations (ODE). Each equation describes the rate at which the concentrations of that metabolite is changing. The total rate is a sum of reaction rates of the reactions that consume and produce the said metabolite. This system is written more easily by matrices (Equation 6.1). First, a stoichiometric matrix $S$ gives the stoichiometric coefficients of all

reactions arranged so that the coefficients of the same reaction are in one column on the rows that correspond to the metabolites (Table C.1 on page 83). Second, the reaction rates calculated with whatever method (mechanistic or power-law, in this thesis) are written into a vector $v$ in the same order as the columns of the stoichiometric matrix are. Multiplying the stoichiometric matrix by the rate vector yields a vector that has the molar rate of change of each metabolite. Thus the figures have to be divided by the volume of the compartment to give the rate of change of concentration. The system is written mathematically in equation 6.1.

$$\frac{dC}{dt} = \frac{k}{V} \cdot S \times v(C) \iff \begin{pmatrix} \frac{dC_1}{dt} \\ \vdots \\ \frac{dC_n}{dt} \end{pmatrix} = \frac{k}{V} \cdot \begin{pmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nm} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}, \tag{6.1}$$

In equation 6.1 $n$ is the number of metabolites in the model and $m$ is the number of reactions. Term $k$ is used for unit conversion and $V$ is the volume of the corresponding compartment. Since the enzyme activity is defined per cell dry weight (CDW) also volumes have to be given relative to 1 mg CDW. The cell internal volume is reported to be $V_{in} = 1.5 \cdot 10^{-3} mL/mg\ CDW$ (Kashket and Wilson, 1973). Given the cell density $cd$ (in mg CDW / mL of batch volume), the fraction of unit volume occupied by the cell internal volume is $V_{in} \cdot cd$. Naturally, rest of the unit volume is cell external space, that is $V_{out} = 1 - V_{in}/cd$ in mL.

The solvers are commonly numerical integrators that adopt special algorithms in order to deal with different kinds of systems. In addition to the above ODE-system, an initial condition for the concentrations has to be determined in order to get a unique solution. Solving this system gives the concentrations for all metabolites between the initial and final time. With the rate equations, the reaction rates can also be calculated from the concentrations. The time course of concentration (rate) is also called a concentration (rate) profile.

### 6.2.2 Initial conditions

Initial condition is an essential part that has to be defined in order to solve a ODE-system. In this case the initial condition means the initial concentrations of all the metabolites. As only some of these values were measured, the others were given a guessed value of 1.0 mM. The measured initial concentrations are given in table 6.2. For the metabolites that could not be detected from the initial samples, a very small non-zero value was given as an initial concentration.

### 6.2.3 Solver

The model was coded with Fortran 95. Particularly, the ode-solver was selected from a public domain library ODEPACK (Hindmarsh, 1983) that is coded in Fortran 77 and available online from Netlib (http://www.netlib.org/odepack). The solver used, dlsode, is designed to solve stiff ordinary differential equations systems. The solver has a number of parameters, most important of which are discussed next.

**Table 6.2:** Initial condition of the simulation. Other metabolites were given a value of 1.0 mM.

| Metabolite | Initial concentration (mM) |
|---|---|
| Glucose (external) | 484.0 |
| G6P | 1.62 |
| F6P | 0.40 |
| FBP | 3.47 |
| 1,3PG | $1.0 \cdot 10^{-15}$ |
| 3PG | 2.30 |
| PEP | 0.78 |
| Pyr | 0.59 |
| Lactate (external) | 47.1 |
| Formate (external) | $1.0 \cdot 10^{-15}$ |
| Ethanol (external) | 0.39 |
| Acetate (external) | $1.0 \cdot 10^{-15}$ |
| Xylose (external) | 163.0 |
| Xylitol (external) | 1.74 |
| Phosphate (external) | 25.0 |
| ATP | 0.42 |
| NADH | 0.46 |
| ADP | 0.41 |
| CoA | 20.0 |

The solver needs the function to evaluate the right hand side of equation 6.1. This is a user supplied function that uses a fixed parameter set to calculate the reaction rates at given metabolite concentrations. Other important solver parameters are relative and absolute tolerances. They do not give the error in the results, but the maximum allowed local error. They affect the step length chosen by the solver. The final error may be much larger due to accumulation. The values for the tolerances are chosen to give stable behavior without needlessly slowing the calculation due to excess accuracy. Here an absolute tolerance of $10^{-10}$ and a relative tolerance of $10^{-8}$ were used. The solver has also additional parameters which are less important.

### 6.2.4 Compiler and libraries

The source code was compiled with the Intel® Fortran Compiler (Version 8.1, Build 20040803Z, Intel Corporation, Santa Clara, USA) that is optimized for Pentium processors. In addition, an Intel-optimized math library Intel® Math Kernel Library (Version 7.0.1, Intel Corporation, Santa Clara, USA) was used for linear algebra and power function evaluations.

## 6.3 Parameter estimation

The problem of parameter estimation for the kinetic metabolic network model is difficult and no direct analytical method exists that could find the correct parameters. The parameter esti-

mation is based on minimization of an objective function. The objective function in this case represents the squared difference of the model prediction and the measured concentration profiles. There are many routines that minimize the objective function and they can be divided into deterministic and stochastic methods. Deterministic methods always find the same solution as long as the initial conditions are the same whereas stochastic methods use randomness to find a better parameter set and thus may give different solutions on different runs.

In this thesis a stochastic methods have been used. They are robust methods in the sense that a lower value of the objective function is nearly always available with a simple code. Deterministic methods are more complicated. A multidimensional Newton-Raphson method (Press *et al.*, 1992, pp. 376-381) with some obvious changes was tested but it did not give much promise. It seems that deterministic optimization methods are not applicable for this complex problems.

### 6.3.1 Random search method

Random search is a very simple method. Basically, parameter sets are selected randomly within certain limits and the value of the objective function is evaluated. The parameters that give the lowest value are the best. The problem with random search method is that there is no clear rules for the stop condition. Continuing the run will most probably lead to a lower value of the objective function, no matter how many iterations are already made. However, there is no way of knowing whether the lower value found is the lowest possible in the parameter range that is being searched. On the other hand, there is usually a certain level of accuracy that is sufficient and no further iterations should be made in order to enhance it. Therefore, two conditions are commonly adopted: sufficiently low value of objective function and an upper bound to the number of iterations. These values are subjective and they have to be balanced according to the required accuracy and the available computational resources.

Selection of the parameter limits is another problem. For a new system the proper scale of the parameters is usually unknown. Thus, initial runs have to be made with wide parameter ranges in order to find the relevant minima of the objective function. This applies especially to the mechanistic rate equations that may have parameter ranges of several decades. Power-law functions are more simple in this sense since the biologically relevant values of the indices are quite limited (less than a decade). The rate constants have naturally more freedom.

### 6.3.2 Monte-Carlo method

Monte-Carlo methods are used to refine the parameter sets obtained from random search to further minimize the corresponding values of the objective function. The idea in Monte-Carlo type parameter estimation is simple. At each iteration the value of a randomly chosen parameter is changed by a random amount and the objective function is evaluated. If the value is lower than the previous lowest value, the new parameter value is accepted. If the value is higher,

41

the new parameter value is rejected and given its original value. A term Monte-Carlo step is used in this thesis to describe a number of iteration equal to the number of parameters in the model. Thus, during each MC-step an attempt to change every parameter is made once on average. This way the number of iteration steps can be directly compared between models with different amount of parameters.

The Monte-Carlo method only finds a local minimum. On subsequent runs from the same initial parameter values it may, however, find different local minima due to the randomness incorporated into the algorithm. Thus, a number of Monte-Carlo runs are required to get a reliable estimate for the global minimum of the objective function.

The details of the Monte-Carlo algorithm for parameter estimation that is used in this thesis is described below.

1. Get the initial parameters either randomly or take the best parameters from a random search.

2. Select a random parameter. Take a uniformly distributed random number, scale it to the number of parameters in the model and force it to integer. This is the index of the parameter to be changed.

3. Change the value of the selected parameter. For example, get a random number from a Gaussian distribution with mean one and variance of your choice and multiply the old parameter value with that random number. Do not allow the random number to be negative. Then multiply the old value with this random number in order to get the new value. In the present study a Gaussian random variable with $\mu=1$ and $\sigma=0.01$ was used.

4. Evaluate the objective function.

5. Accept or reject the new parameter. If the new value of the objective function is lower than before, accept the new parameter value. If the value is higher, the new parameter value is rejected.

6. Check the stop condition. The number of iterations of the Monte-Carlo simulation should have a limit. If the limit is met, stop the simulation. In addition to the iteration count, it may be beneficial to stop the simulation if no progress is made for some time. This indicates that the simulation is stuck in a local minimum and is rejecting all attempts to change the parameter values.

7. If stop condition is not met continue simulation from step 2.

### 6.3.3 The combination method

The parameter estimation was carried out with a combination of the random search and Monte-Carlo minimization methods. First, a random set of parameters were selected from the parameter space given in table 6.3 and checked that they did not produce ODE-problems. Then

**Table 6.3:** The scale of parameter space in random search.

| Model | Parameter | Scale |
|---|---|---|
| Power-law model | Rate constants | $0.001\ldots10.0$ |
| | Kinetic orders | $-0.5\ldots1.0$ |
| Mechanistic model | Michaelis constants ($K_m, K_i$) | $10^{-6}\ldots10^{3}$ |
| | Limiting rates ($V_{max}$) | $10^{-6}\ldots10^{3}$ |
| | Equilibrium constants ($K_{eq}$) | $10^{-6}\ldots10^{3}$ |
| | Rate constants | $10^{-6}\ldots10^{3}$ |

the parameters were passed on to the Monte-Carlo minimization algorithm. The length of MC minimization was limited to 5000 MC-steps. However, if the decrease in error was less than 0.01 in 100 MC-steps, the minimization was stopped, because it was considered to be too slow or close to a local minimum. The combination of a random search and a subsequent MC minimization is here called a *random run*. In order to gain more confidence in the results, random runs were performed repeatedly to see if they ended up in the same minimum. Such a method is often referred to as a multistart method.

From every random run a number of variables were recorded for further analysis. The most important figures are the final error between simulations and data, the final parameter values, the number of MC-steps taken during the random run and the time consumption of the parameter estimation. In addition, the number of accepted and rejected changes, and changes that ended prematurely (ODE-problems), were calculated for each random run.

# Chapter 7

# Results and discussion

First in this chapter the stability and performance of the mechanistic and power-law simulation models will be evaluated. Then the parameter estimation routine and its applicability to the current models are assessed. Finally, the metabolism of the LLCON and LLXR strains are simulated and the concentration profiles are studied. Simulations are carried out with both the mechanistic and power-law models with parameters giving the lowest error for each combination of model and strain.

## 7.1 Model

### 7.1.1 Model stability

The model itself appears to have some stability problems. It was noted multiple times that with parameters given by a parameter estimation method the calculation was not finished on subsequent runs. In most of these cases the solver was stopped on purpose because it produced negative concentrations, which is chemically impossible but should be allowed from the mathematical point of view. However, because behavior of the rate equations on negative concentrations could lead to many problems, it was decided that the calculation should be ended on such an event. Integration tolerances could have been tightened to avoid negative concentrations, but this would have been done on the expense of speed.

In addition to the stability problems, the models were also very sensitive to their parameter values. This property was not investigated on a large scale, but it was observed during the parameter estimation runs. It may be perfectly normal for such a complex ODE-systems but it may also indicate a floating point accuracy problem.
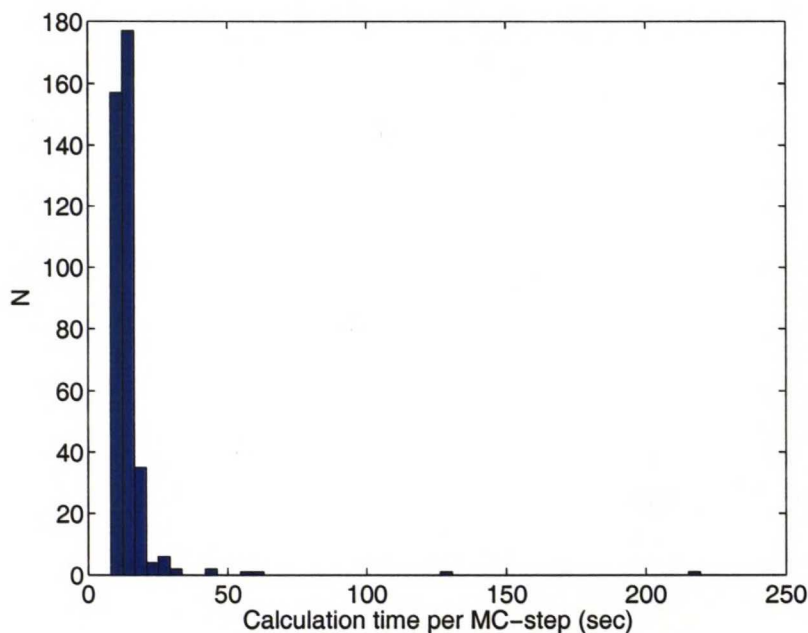
**Figure 7.1:** The distribution of the calculation time of one Monte-Carlo step in the parameter estimation with the mechanistic model. The average value of the 387 random runs was 14.7 s.

### 7.1.2 Model performance

The time taken by the ODE-system evaluation depends greatly on the initial conditions and the parameter values. In case the parameters opposed rapid changes in the concentration profiles the evaluation was slower than if the changes were more gradual. The speed of the algorithm is directly proportional to the number of evaluations of the right-hand side of equation 6.1. Therefore, the evaluation of the rate equations of the two models is probably not equally fast.

The distributions of MC-step calculation times for the mechanistic and power-law models are shown in figures 7.1 and 7.2, respectively. The values are averages obtained from the overall Monte-Carlo minimization time and the total number of MC-steps taken in one random run. Because of this, the number includes also the failed ODE-system evaluations that have been aborted prematurely. Thus, these numbers are underestimates of the real calculation times of one MC-step. The calculations were carried out on a variety of modern computers including 2.4 to 3.2 GHz Intel Pentium 4's and 2.0 GHz AMD Opterons.

The distribution for the mechanistic model (Figure 7.1) appears to be narrow with an average of 14.7 seconds. In a few instances the calculation time has been greatly increased which may have been due to a condition where the randomly selected parameter values have imposed fast reaction rates and the ODE-system evaluation has been slow, or the parameters have been selected from a part of the parameter space where the variations from the original values have not posed premature abortion of the ODE-system evaluation. In the latter cases, the values are considered to be better estimates of the real MC-step calculation time.

The power-law model has a broader distribution with considerably higher average of 65.7 sec-
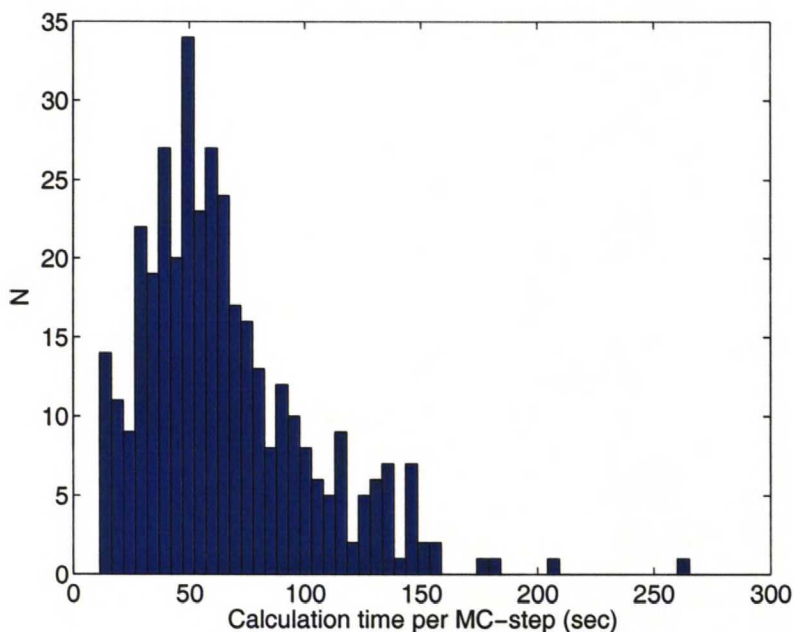
**Figure 7.2:** The distribution of the calculation time of one Monte-Carlo step in the parameter estimation with the power-law model. The average value of the 370 random runs was 65.7 s.

onds (Figure 7.2). This is almost five times that of the mechanistic model. Even though the mechanistic model had some runs with long calculation time, only few of them were higher than the average of the power-law model. The power-law model is therefore significantly slower than the mechanistic model. As previously speculated, this is probably due to the longer evaluation times of the power functions compared to addition, subtraction, multiplication and division.

## 7.2 Parameter estimation

In the early stage of the study it became very clear that the *in vitro* parameters reported in the literature did not describe the kinetics of the network correctly. It turned out that with the literature parameters the ODE-solver aborted prematurely. The fundamental problem seemed to be that the literature parameters defined such a complex kinetic state that it was not possible to solve it with any practical error tolerance. Therefore, the literature parameters were abandoned and no prior information of the parameter values was used. Instead, the parameter values were estimated with the combination method presented earlier. This yielded a more conventional parameter estimation scheme where the parameters can have virtually any values.

The number of random runs made for each model and strain are given in table 7.1. Target was 200 random runs for each model-strain combination but they were limited by time constraints. In any case, thorough statistical analysis would have required thousands of random runs which was not possible to achieve in the present study.

**Table 7.1:** The number of random runs carried out for each model and strain. Target was 200 runs for each model-strain combination.

| Model | Random runs | | |
|---|---|---|---|
| | LLCON | LLXR | Sum |
| Mechanistic | 187 | 200 | 387 |
| Power-law | 102 | 268 | 370 |
| Total | | | 757 |

## 7.2.1 Performance of the algorithm

The performance of the algorithm can be measured in many ways such as the rate of minimization or the minimization per MC-step. Because here the issue has been stability, the amount of ODE-problems (premature abortion of the run due to negative concentration) in one MC-step has been selected for the measure of algorithm performance. Strictly speaking, the amount of ODE-problems is a sum of many factors such as the stability of the ODE-system, the solver and the parameters of the parameter estimation algorithm.

Histograms of the amount of ODE-problems per one MC-step for the mechanistic and power-law models are shown in figures 7.3 and 7.4, respectively. The values are averages of each random run. The maximum value is the amount of proposed changes per one MC-step, 164 for the mechanistic model and 146 for the power-law model. The figures show that both models create a great deal of ODE-problems but the power-law model creates more than the mechanistic model. It is important to note that not all ODE-problems are caused by inabilities of the solver, but also because of the parameters of the minimization algorithm. For example, an ODE-problem could have been caused by a too large change in a parameter value. As a result, the algorithm is driven off the track. It seems that the solvable paths in parameter space are quite narrow.

Further analysis of the histograms of ODE-problems (Figures 7.3 and 7.4) gives more insight into the ODE-problems related to the algorithmic performance. The histogram for the mechanistic model (Figure 7.3) showed that the distribution is bimodal. For many random runs the average amount of ODE-problems per MC-step was close to zero. However, in the majority of the random runs more than half of the proposed changes faced problems. Figure 7.5 gives a possible explanation to the distribution of ODE-problems. In that figure the number of ODE-problems has been plotted versus the final error of the minimization algorithm separately for both strains. The figure clearly shows that in the low ODE-problem runs not as low errors were achieved as in those with more ODE-problems. There is a clear gap between the lowest error close to $4.15 \cdot 10^5$ and the $4.35 \cdot 10^5$ error range. The high number of ODE-problems seem to be the price to pay for crossing that gap for the mechanistic model. Additionally, the results with the LLCON strain seem to be concentrated to the higher errors and lower number of ODE-problems whereas the LLXR strain concentrated on the low error and high ODE-problem side of the plot.
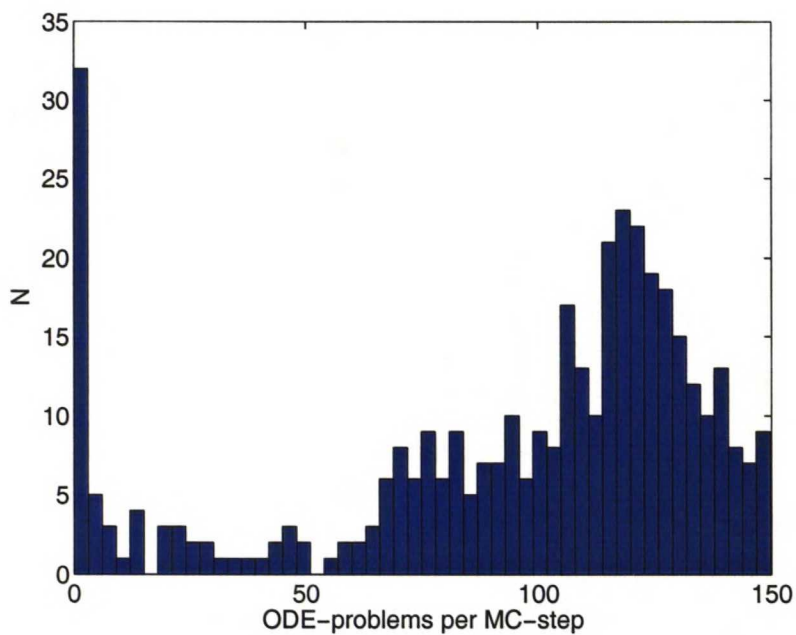
**Figure 7.3:** A histogram of the average number of ODE-problems per one MC-step for 387 random runs for the mechanistic model.
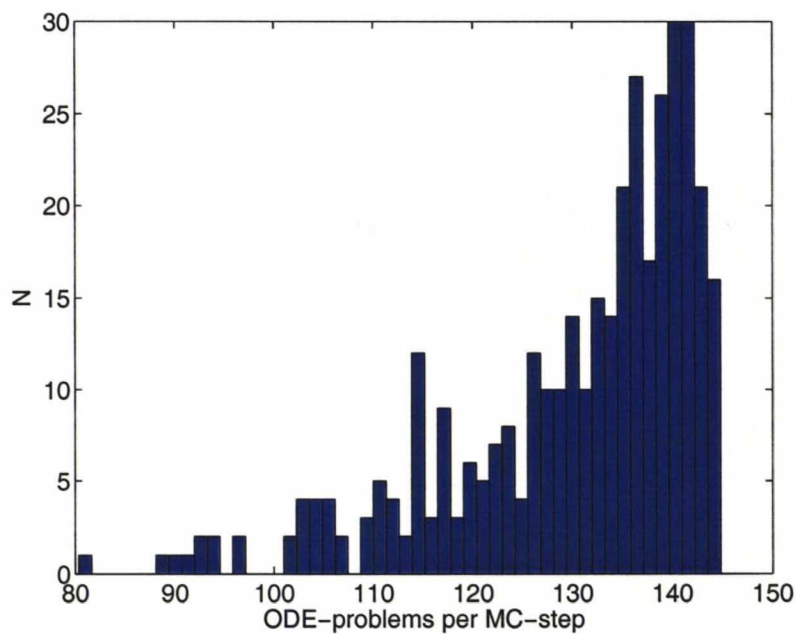


**Figure 7.4:** A histogram of the average number of ODE-problems per one MC-step for 370 random runs for the power-law model.
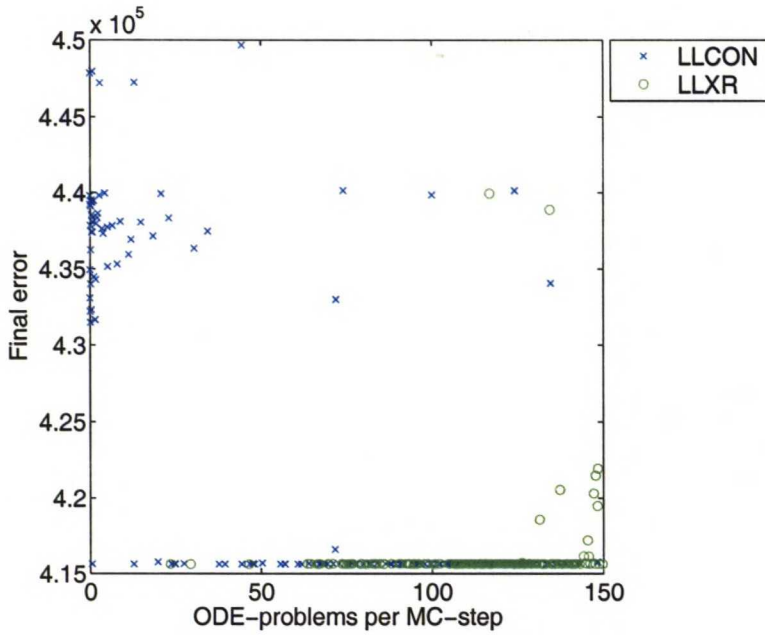
**Figure 7.5:** A scatter plot of the average number of ODE-problems per MC-step versus the final error after minimization for the mechanistic model. Results for the LLCON and LLXR strains are marked with blue crosses and green circles, respectively.

For the power-law model a great majority of proposed changes ended prematurely (Figure 7.4, note that the horizontal scales in figures 7.3 and 7.4 are different). This behavior is probably caused by too large changes to the parameters in attempting to find a lower minimum. If this is true, the power-law model is more sensitive to changes in its parameters than the mechanistic model. It was surprising that despite the similar amount of random runs, the average of ODE-problems per MC-step for the power-law model was never below 80, whereas the mechanistic model had considerable probability mass below that value (Figure 7.3). This indicates that the size of changes in parameter estimation of the power-law model should be revised and optimized for lower percentage of ODE-problems. This may also be true for the mechanistic model. However, these questions could not have been answered in the scope of the current study. The simulations would have taken about 14 months on one computer but they were actually performed on multiple hosts within 3 months.

In addition to the different distribution in ODE-problems, the high value of ODE-problems for the power-law model does not correlate with low error as it did for the mechanistic model (Figure 7.6). In fact, a close-up in figure 7.7 shows that the band of lowest errors has approximately same distribution as the whole sample in figure 7.4. Additionally, high number of ODE-problems resulted in significantly higher errors for the LLXR strain but not for the LLCON strain, as indicated by figure 7.6. This is completely opposite from what was observed for the mechanistic model. However, it is unlikely that there would be any theoretical reasons for the two models to give such different results. These observations are probably only due to the unoptimized size of changes in the parameter estimation routine for the power-law model.

The gap visible in figure 7.5 was an example of the fact that in this kind of parameter estimation
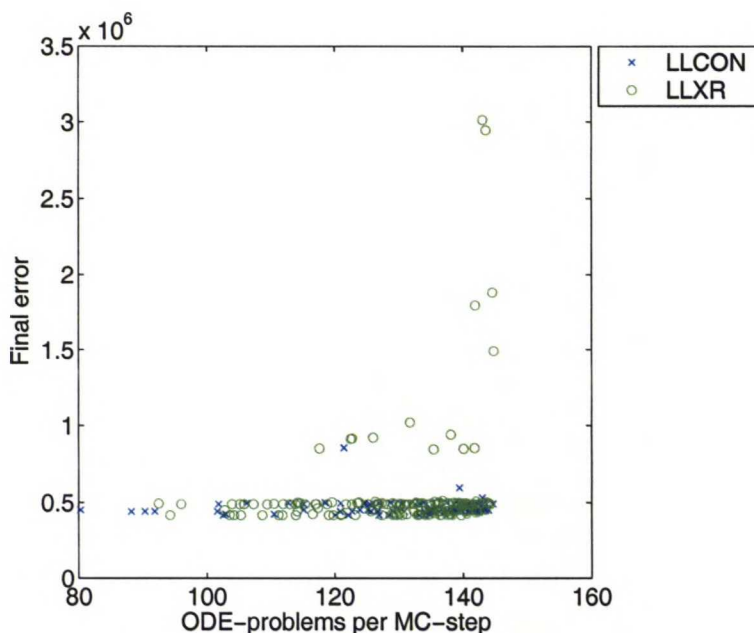
**Figure 7.6:** A scatter plot of the average number of ODE-problems per MC-step versus the final error after minimization for the power-law model. Results for the LLCON and LLXR strains are marked with blue crosses and green circles, respectively.

many runs end up in the same local minimum. The two bands around the gap represent two such minima in the parameter space. A gap can also be found in figure 7.6 just below error of $0.5 \cdot 10^6$. A close-up of that region in figure 7.7 reveals a fine structure of local minima that are represented by horizontal bands. It is important to note that the power-law model has more visible bands than the mechanistic model. Additionally many of these bands were common for both strains. It may be considered a benefit for a model to gain more information of the situation around a minimum. The visibility of higher level of details may facilitate the finding of lower errors and thus better parameters by the parameter estimation algorithm. However, closely situated local minima may prevent the algorithm from finding a global minimum, since the runs are trapped to the first local minima they can find. A simulated annealing method would be more successful in this case because it can pass a local minimum. However, it may be slower in the beginning of the parameter estimation run.

A stacked bar graph summary of the average outcomes of the 387 random runs for the mechanistic model is shown in figure 7.8. This figure also highlights the bimodal distribution of the ODE-problems (red area). Note that the first 187 random runs in figure 7.8 are performed for the LLCON strain and the rest 200 for the LLXR strain. Thus, the parameter estimation algorithm gave more ODE-problems for the mutated strain than for the control strain. This could have been due to the introduction of xylose reductase in the LLXR strain. This reaction utilizes the NADH and $NAD^+$ pools and puts more strain on them, possibly facilitating the depleting of the pools too fast. The result was that a concentration got negative due to the inability of the ODE-solver to follow the rapid depletion. This suspicion is enforced by the fact that all random runs with close to zero fraction of ODE-problems were carried out for the control strain.
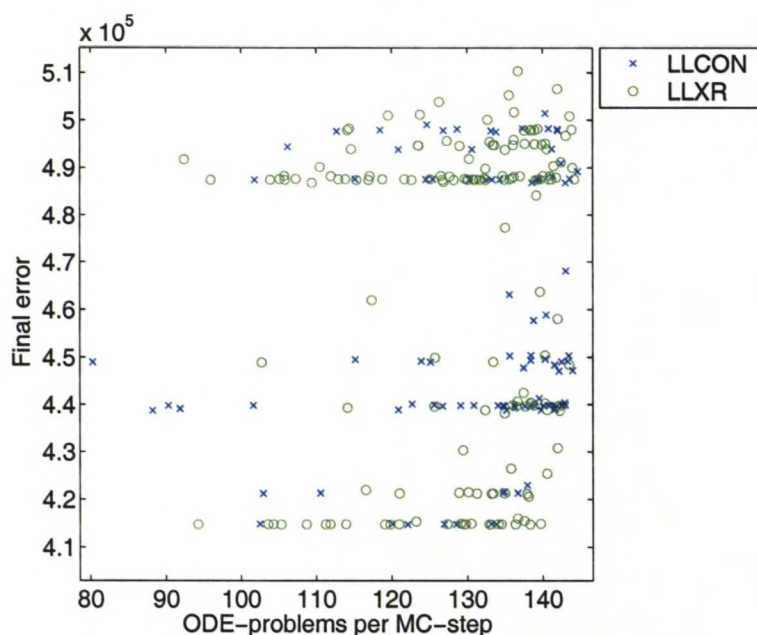
**Figure 7.7:** A close-up of the scatter plot in figure 7.6 showing the fine structure of the local minima. Results for the LLCON and LLXR strains are marked with blue crosses and green circles, respectively.
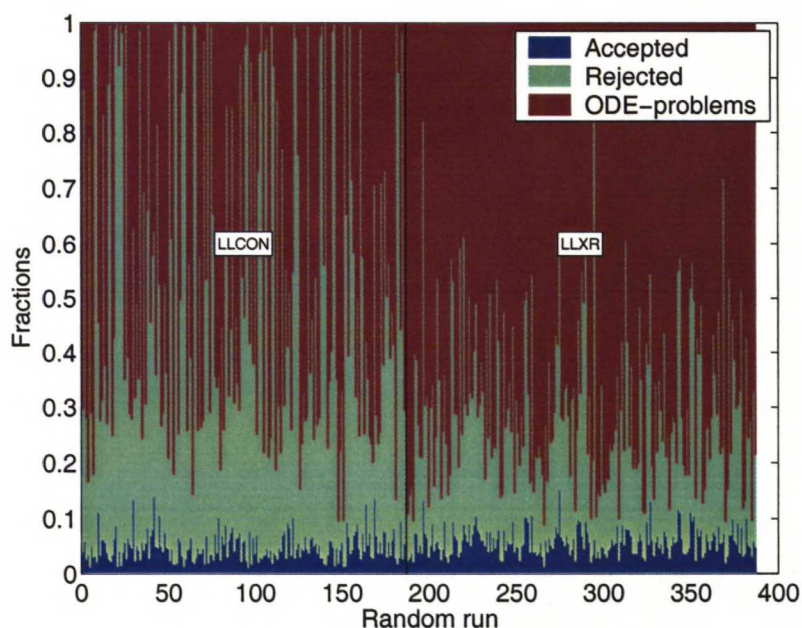


**Figure 7.8:** The fraction of each possible outcome of a proposed change for the mechanistic model. The 187 first random runs correspond to the LLCON strain and the rest 200 to the LLXR strain.
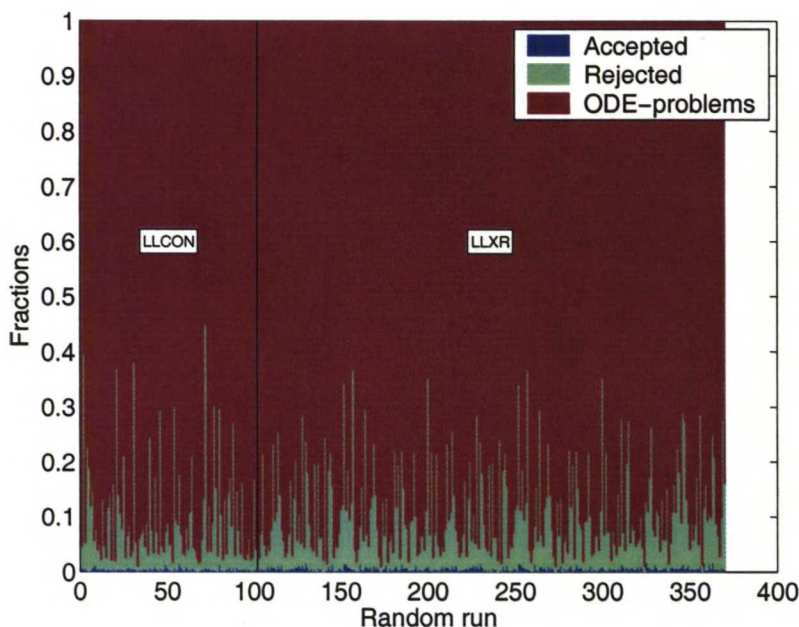
**Figure 7.9:** The fraction of each possible outcome of a proposed change for the power-law model. The 102 first random runs correspond to the LLCON strain and the rest 268 to the LLXR strain.

The same summary for the power-law model is shown in figure 7.9. It emphasizes the high amount of ODE-problems (red area). The number of accepted and rejected changes was also much less for the power-law model than for the mechanistic model. There were no differences between the strains as for the mechanistic model in figure 7.8. For the power-law model the strains had an equal amount of ODE-problems.

## 7.3 Comparison of the models

### 7.3.1 Mechanistic model

The mechanistic models seemed to have an overhead error of around 415000 that could not be passed in any of the 387 attempts. The minimum errors for the LLCON and LLXR strains were 415596.39 and 415593.36, respectively. Concentration profiles of the LLCON strain internal metabolites corresponding to the minimum error are presented in figure 7.10 and external metabolites in figure 7.11. Corresponding concentration profiles for the LLXR strain are presented in figures 7.12 and 7.13, respectively. The results of the two different strains indicate that the simulation results are very close to each other as could have been assumed on the basis of the errors. The figures 7.12 and 7.13 indicate that the concentrations remained constant for most of the latter part of the simulation. A closer inspection on the numeric results confirms that the concentrations did not change after 30 minutes (data not shown). This implies that both simulations ended up in a halted equilibrium state.

There may be many causes for the halted state. Clearly, a halted state is not observed in
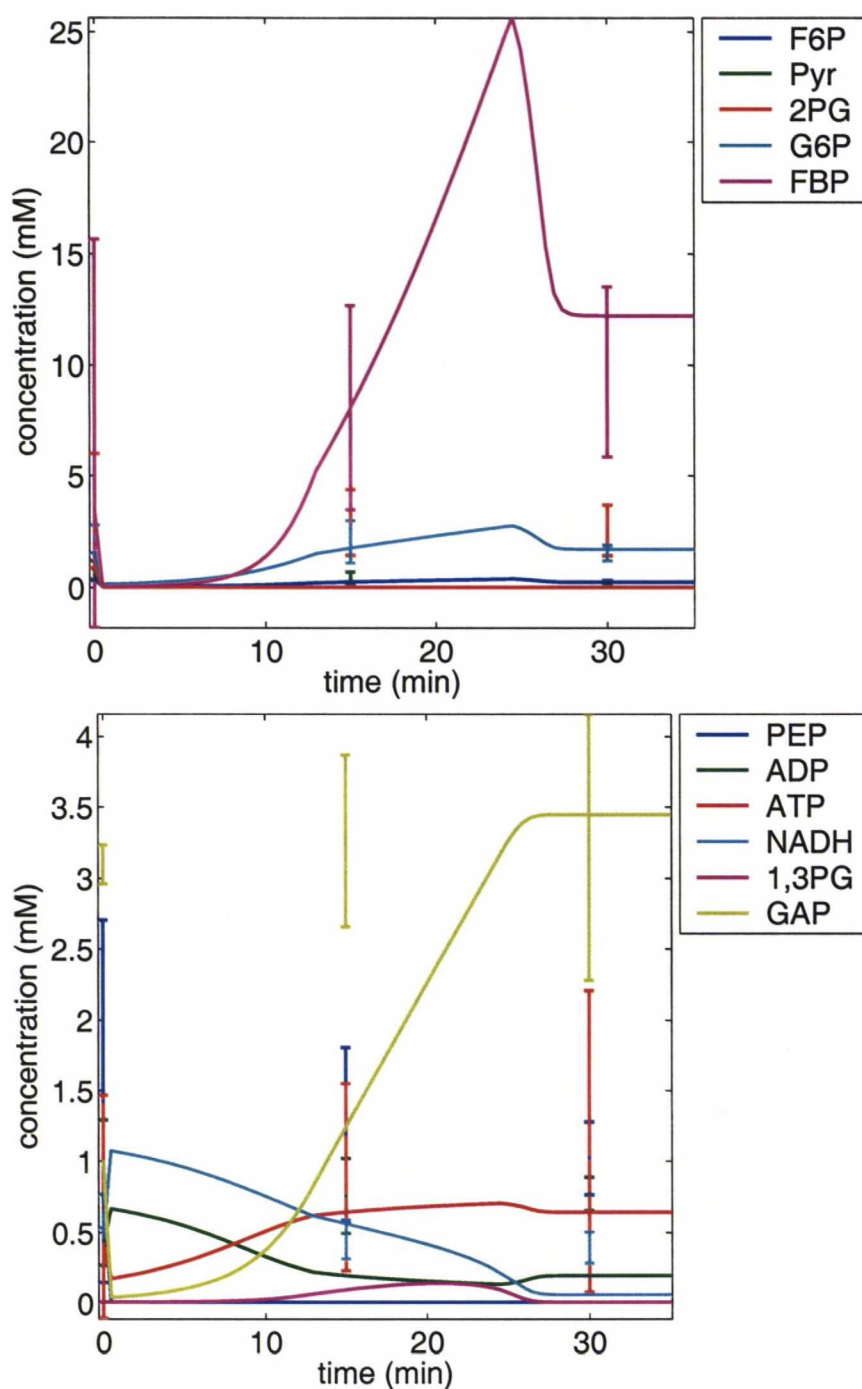
**Figure 7.10:** Concentration profiles for the internal metabolites giving lowest error for the LLCON strain simulated with the mechanistic model. Error bars represent the standard deviation in the measurements.
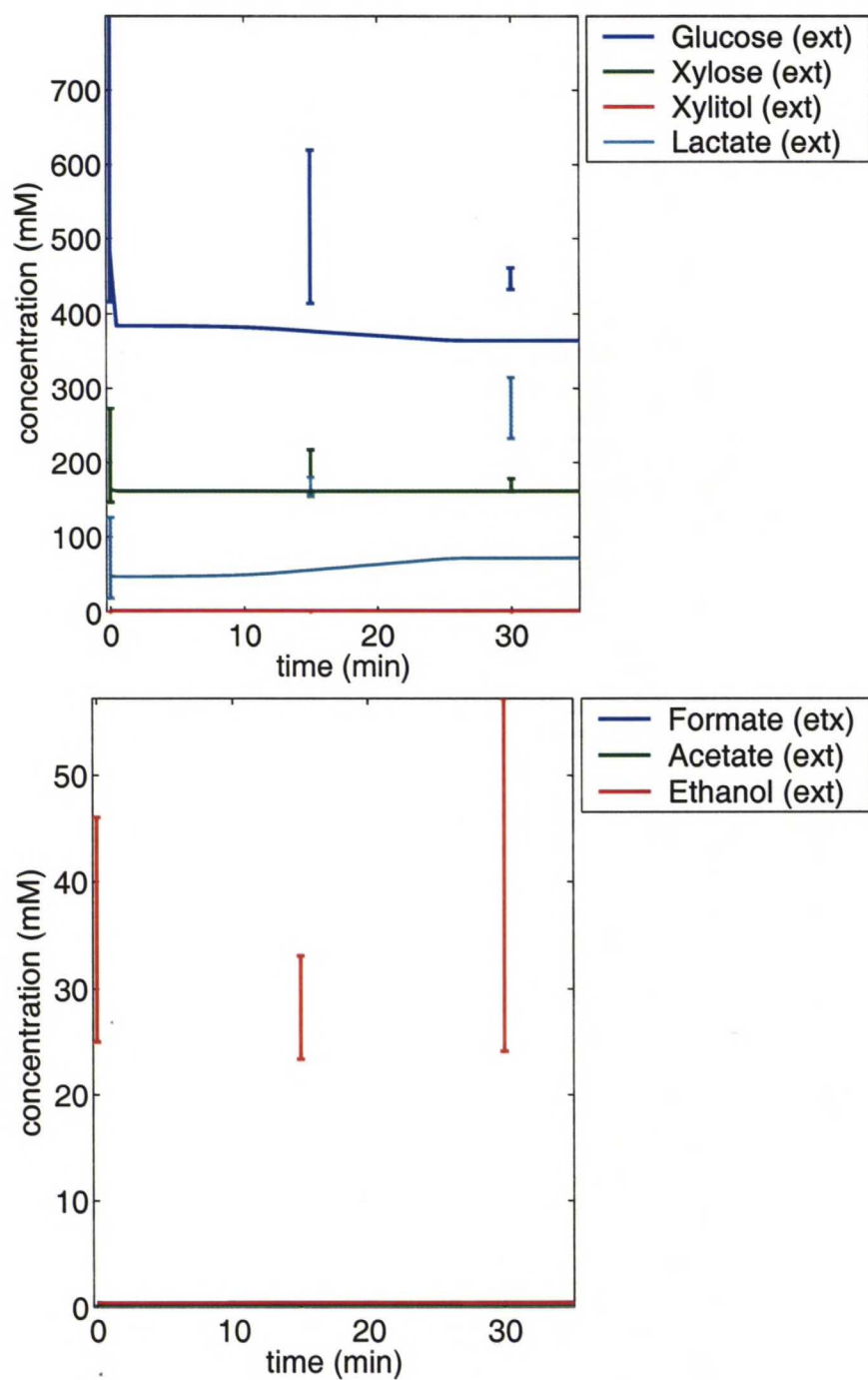
**Figure 7.11:** Concentration profiles for the external metabolites giving lowest error for the LLCON strain simulated with the mechanistic model. Error bars represent the standard deviation in the measurements.
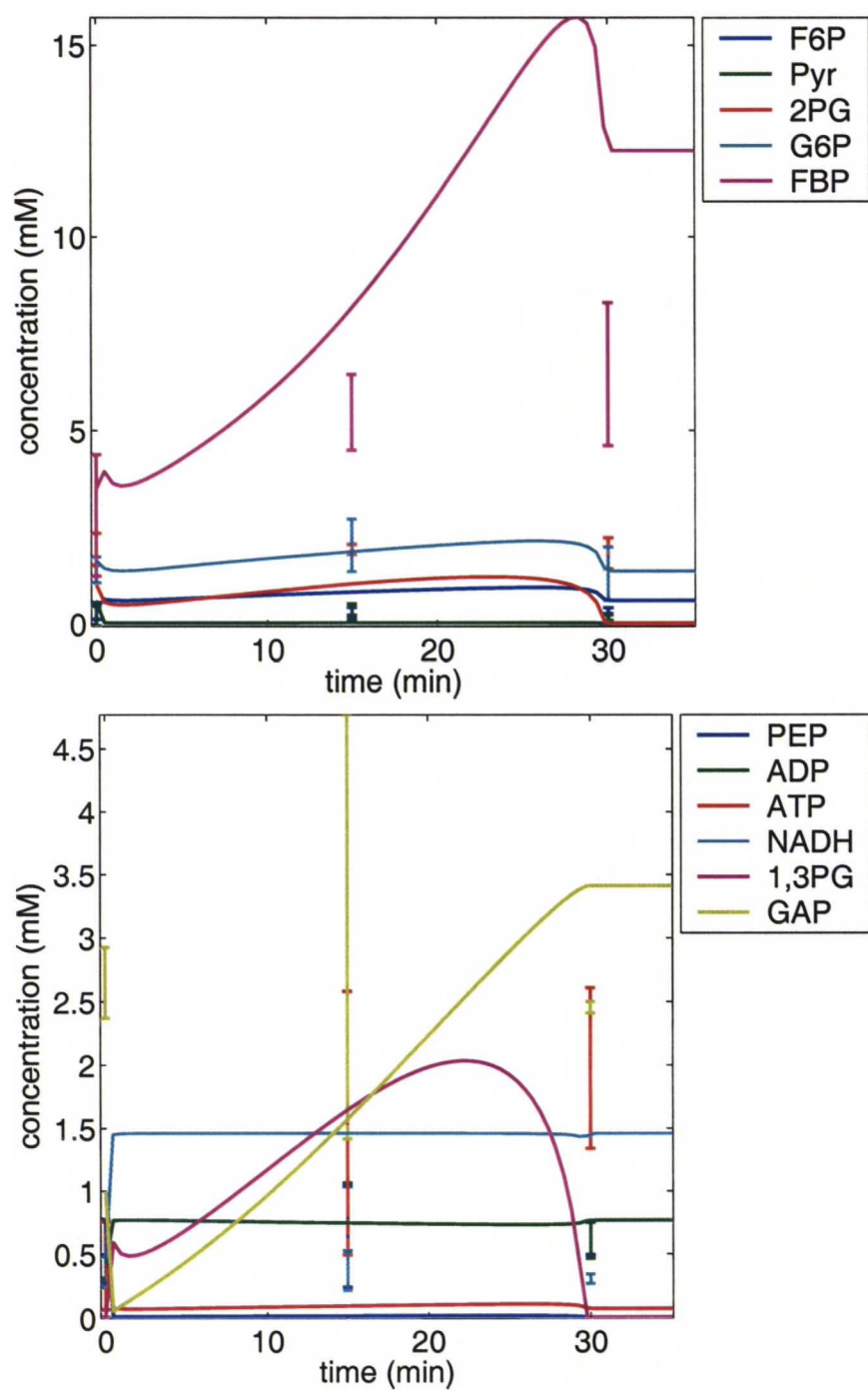
**Figure 7.12:** Concentration profiles for the internal metabolites giving lowest error for the LLXR strain simulated with the mechanistic model. Error bars represent the standard deviation in the measurements.

**Figure 7.13:** Concentration profiles for the external metabolites giving lowest error for the LLXR strain simulated with the mechanistic model. Error bars represent the standard deviation in the measurements.

the *in vivo* data. Although internal metabolites may operate around a preferable value, the external metabolites should show continuous utilization of glucose and production of lactate in both strains. Thus, the halted state is an anomaly of the mathematical setup. It could have been caused by depletion of an important metabolite. The simulation data shows that 1,3PG, 3PG, 2PG, PEP, Pyr, acetyl-P and internal and external formate, acetate and phosphate had practically zero concentrations on both strains at the halted state. Out of these metabolites, two candidates emerge for causing the halted state; PEP and internal phosphate. PEP has influence on the PTS transport system. Depletion of the metabolites previous to PEP in the reaction network indicates that the PTS system has been drained in order to keep glucose transport on. The result of this is that PTS can not translocate any more carbohydrates into the cell. On the other hand, the lack of phosphate completely stopped GAPDH, which is a more probable cause for halted state. This theory is supported in part by the fact that FBP and GAP have higher concentration in simulations than in measurements for the LLXR strain, which implies accumulation of metabolites preceding GAPDH. This, in turn, suggests that GAPDH is rate limiting. The initial phosphate concentration in the experiments was 20 mM. The initial concentration of external phosphate used in the simulation was increased to 25 mM due to possibility of phosphate in the M17 powder. It seems, however, that this was not enough. Obviously, more phosphate is needed to convert the given batch of glucose to lactate. In the current model GAPDH and PTA utilize phosphate that is eventually passed on to ADP to form ATP. There is not a single reaction that would release phosphate. Such a maintenance reaction consuming ATP and producing ADP and phosphate could be introduced but its specifications would be completely arbitrary. In addition, the amount of maintenance in resting cells is low and the reaction would cause more parameters to be introduced into the model. In its current state, the system is completely dependent on the extracellular phosphate pool.

A short series of single runs were conducted with increased phosphate concentrations. The runs were made with 50, 100 and 200 mM of external phosphate and the parameters corresponding to the minimum error parameters for both strains (i.e. two sets of parameters). All runs ended prematurely indicating the sensitivity of the model to concentrations. Thus, no additional support on the postulate of phosphate depletion causing the halted state was gained.

The parameters of the mechanistic model could have been compared with the parameters reported in literature. Because of the poor results and vast number of parameters, this was not done. If the simulations had performed better the comparison would have been sensible. In the current situation there is only a random chance of finding similar parameter values.

Overall, the simulations were not able to follow the conversion experiments past 30 minutes. After that the simulations got stuck in a halted state presumably caused by phosphate depletion. This could have been prevented by making the initial concentrations subject to the parameter estimation. Obviously, real phosphate concentration was not high enough to give correct simulation results.

### 7.3.2 Power-law model

The power-law model also suffered from an error overhead of around 414000. The minimum error for the LLCON and LLXR strains were 414689.78 and 414682.66, respectively. Concentration profiles of the LLCON strain simulated internal metabolites, together with the measured results and their standard error, are presented in figure 7.14 and simulated external metabolites in figure 7.15. Corresponding concentration profiles for the LLXR strain are presented in figures 7.16 and 7.17, respectively. These results are fairly close to one another and also close to the results for the mechanistic model.

A halted state was also observed for the power-law model. Furthermore, the halted equilibrium state was achieved much faster for the power-law model than for the mechanistic model. A closer inspection on the simulation results showed that no changes in concentrations were observed after 1.0 minute for the LLCON strain and after 0.5 minutes for the LLXR strain. It seems that the time scale of reaction rates is much shorter here than in the mechanistic model or in the measured data. It is possible that the scale of reaction rate constants should be narrowed from the top end forcing slower reactions to occur.

Most of the speculation of the halted state for the mechanistic model also applies to the power-law model. Also the results from the power-law simulations support the theory of halted state being caused by phosphate depletion. This is suggested by the fact that many metabolites previous to GAPDH in the pathway accumulated during the simulation. The GAP concentrations were higher than what was measured for both strains and F6P and FBP concentrations were higher than what was measured for the LLXR strain only.

A series of simulations with increased phosphate concentrations was carried out with the power-law model as well. The test included six simulations with 50, 100 and 200 mM of external phosphate and parameter sets giving the minimum error for both strains. Four of the six tests finished while two ended prematurely (data not shown). The changes in results were evaluated on the basis of the concentrations of external glucose, lactate, xylose and xylitol. The simulations with 50 mM of phosphate did not exhibit any visible changes. The simulations with 100 mM of phosphate together with the control strain parameters and 200 mM of phosphate together with LLXR strain parameters did show visible differences to the original simulations of the power-law model. The concentration of glucose at the halted state was slightly lower (250 - 300 mM) and the lactate concentration was slightly higher (150 - 250 mM). This behavior was expected. However, these results were still far away from the measured concentrations, and furthermore, the halted state still began after 1 minute of simulation, as it did in the original simulations. Therefore, these additional simulations could not confirm the cause of the halted state.

The power-law model was five times slower than the mechanistic model in terms of calculation time of one MC-step. In this model, Intel optimized code was used for power function evaluations. Despite that, the mechanistic model outperformed the power-law model. However, it is possible that there is a more optimized approach. In the case of power-law modeled enzyme
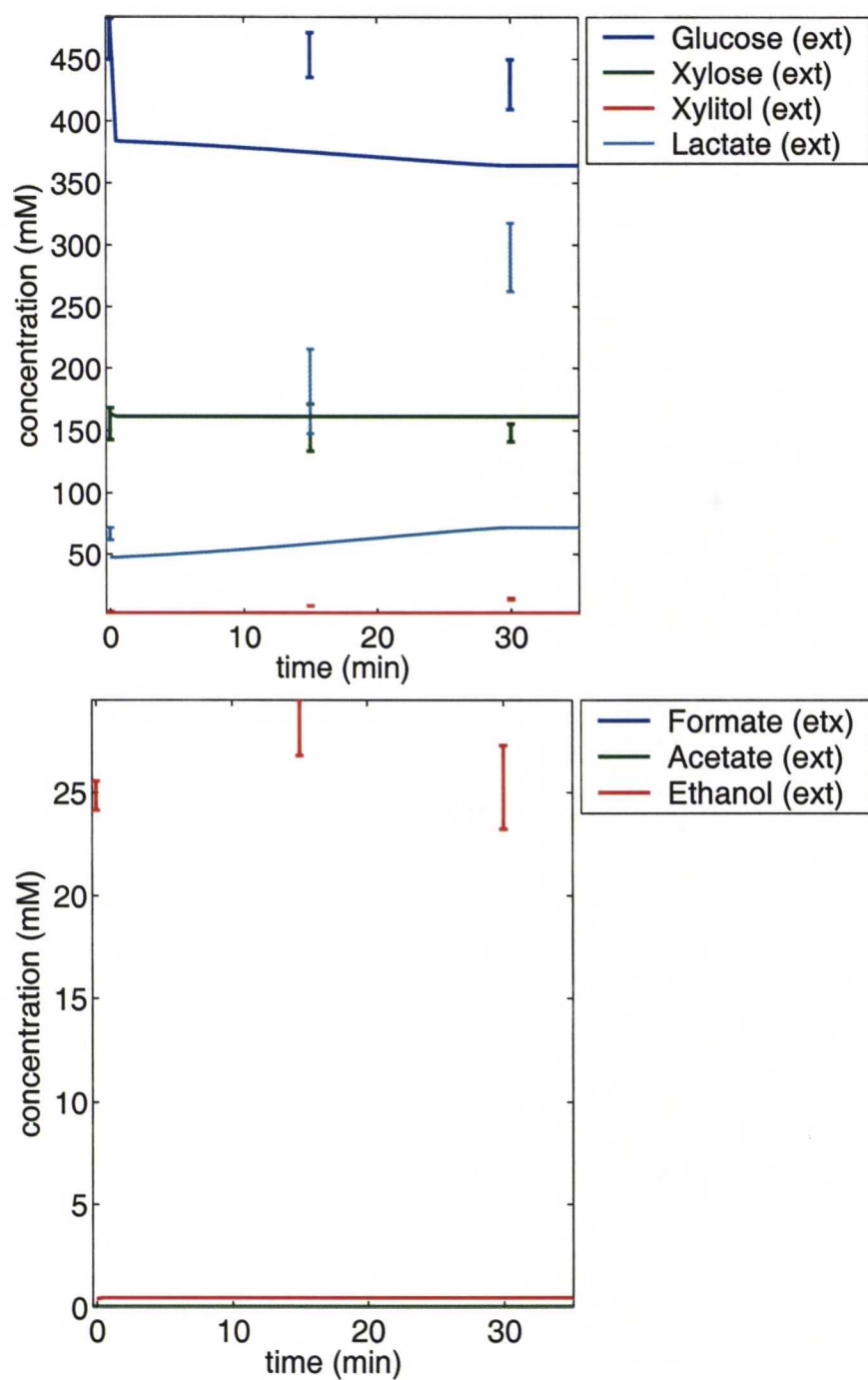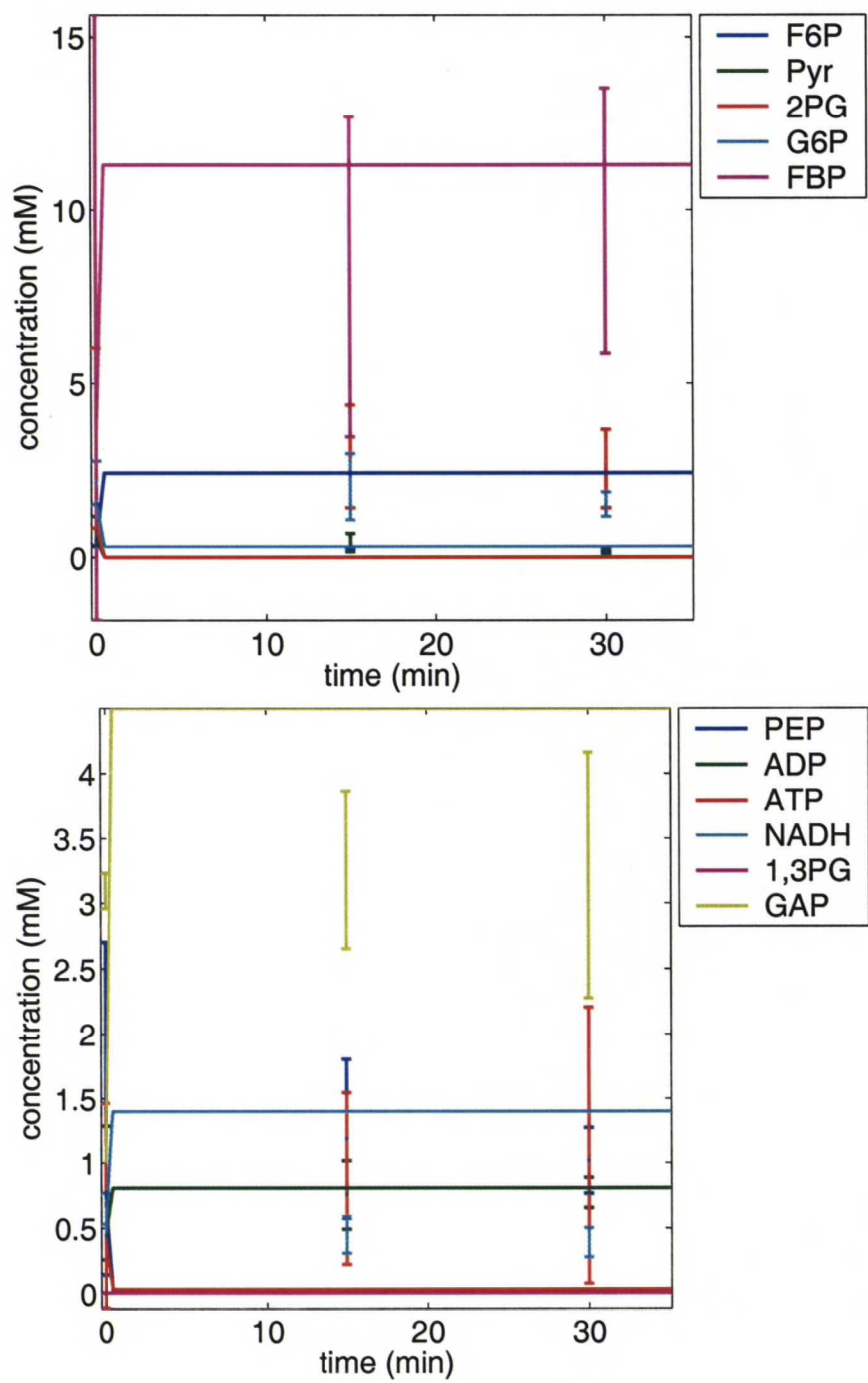
**Figure 7.14:** Concentration profiles for the internal metabolites giving lowest error for the LLCON strain simulated with the power-law model. Error bars represent the standard deviation in the measurements.
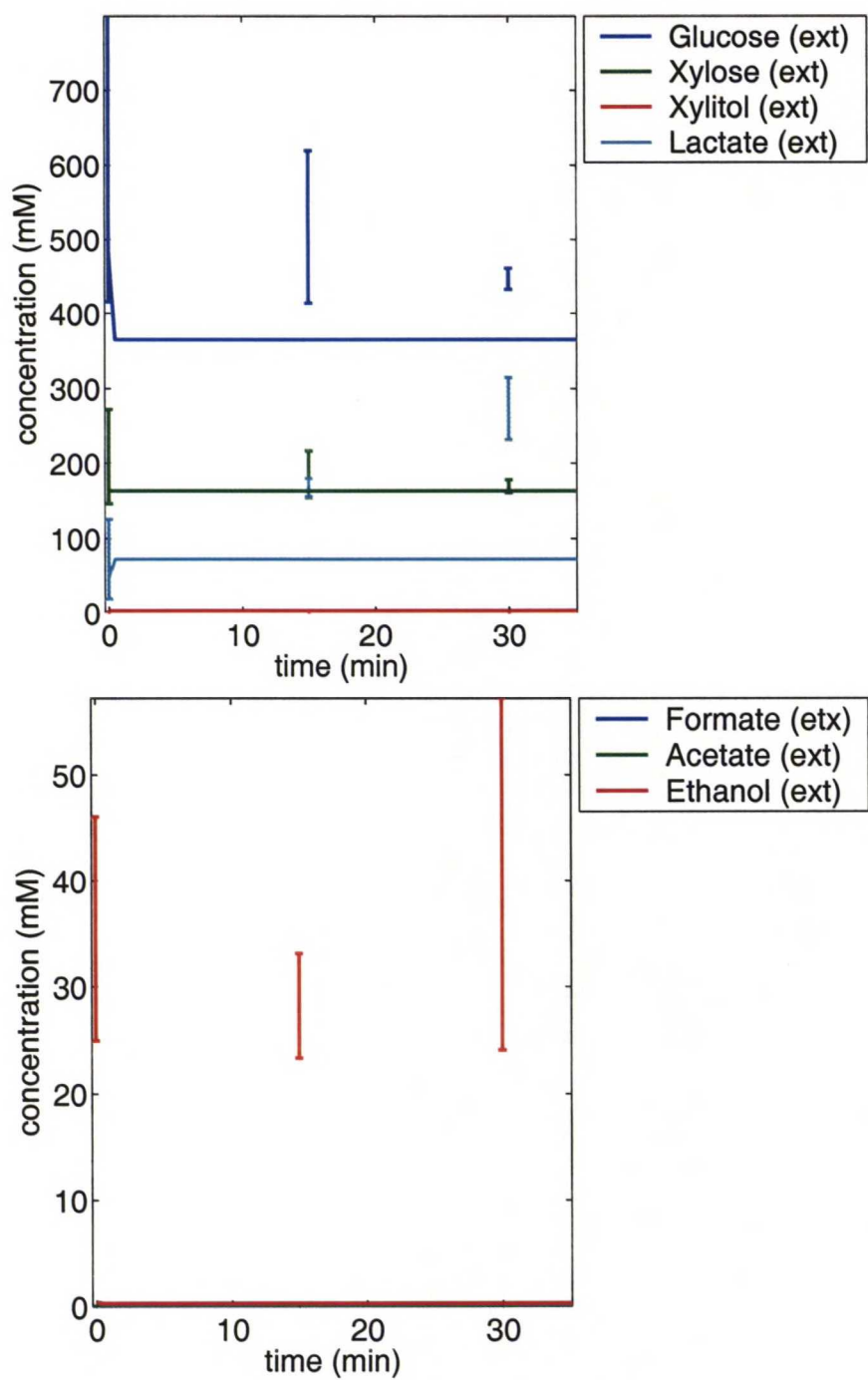
**Figure 7.15:** Concentration profiles for the external metabolites giving lowest error for the LLCON strain simulated with the power-law model. Error bars represent the standard deviation in the measurements.
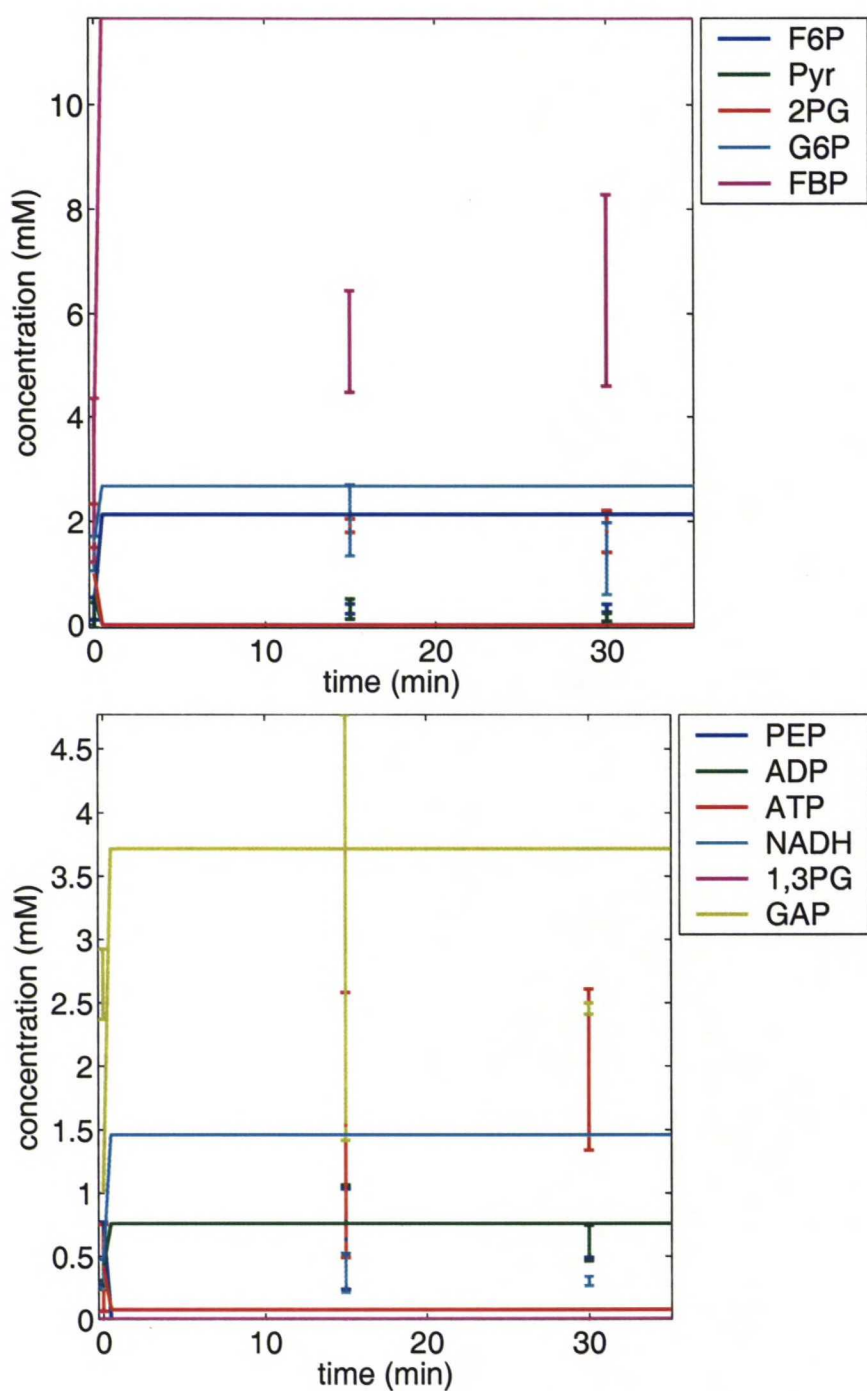
**Figure 7.16:** Concentration profiles for the internal metabolites giving lowest error for the LLXR strain simulated with the power-law model. Error bars represent the standard deviation in the measurements.
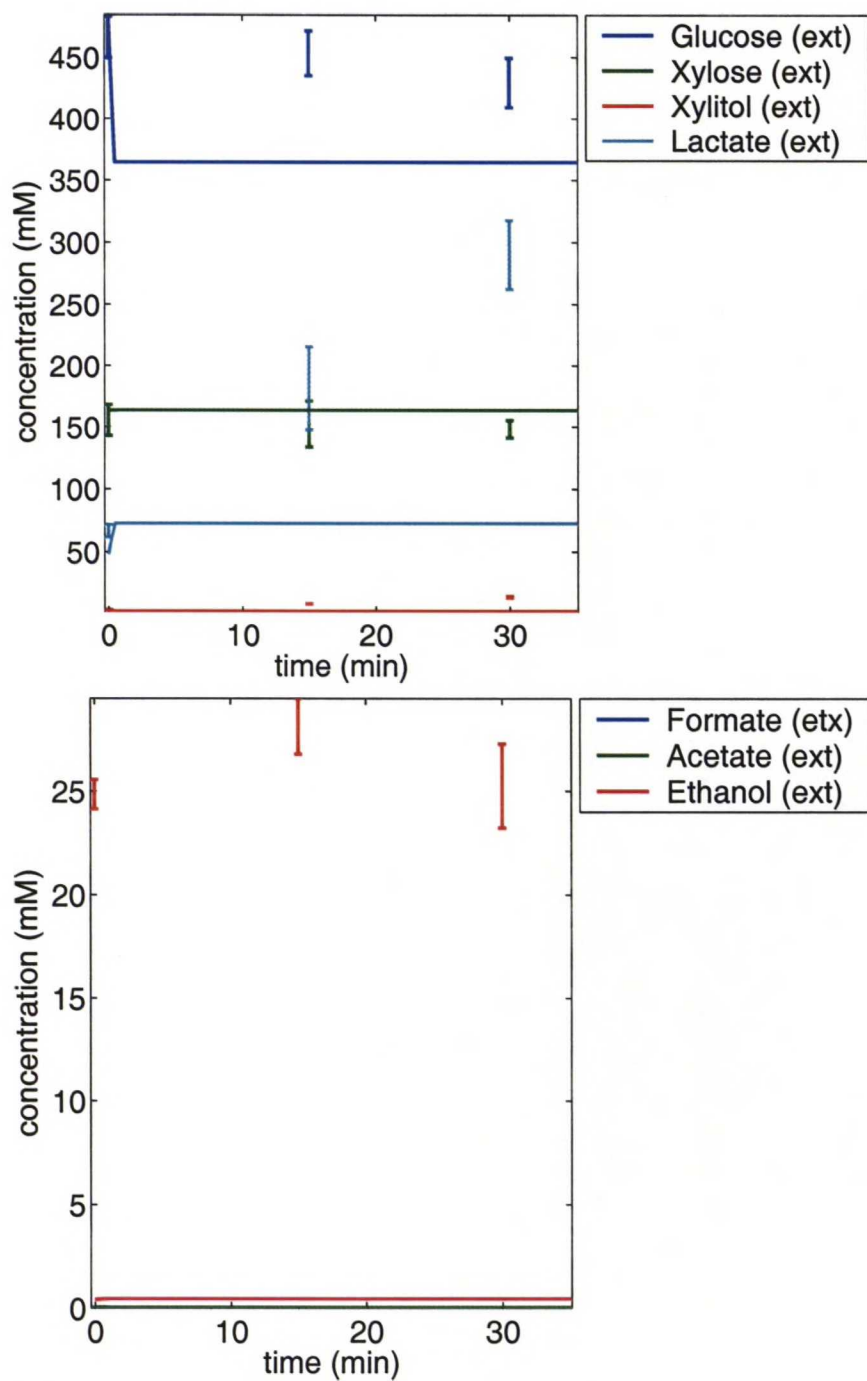
**Figure 7.17:** Concentration profiles for the external metabolites giving lowest error for the LLXR strain simulated with the power-law model. Error bars represent the standard deviation in the measurements.

kinetics, the order terms lie on a relatively narrow range from -0.5 to 1.0. In this range the algorithm for power function evaluation may be much faster than the one for a wide range of order terms. In addition, the accuracy of a single power function term should not be critical. Therefore, the full 15 digit accuracy of double precision numbers is an overkill and probably the 6 digits of a single precision value may be sufficient for a single term.

# Chapter 8

# Conclusions

In this thesis, the metabolism of two genetically modified strains of lactic acid bacterium *L. lactis* were modeled with mechanistic and power-law rate equations. The model based on mechanistic rate equations represents the current knowledge of the kinetic properties of the enzymes catalyzing the reactions of the network. The power-law model was used as a rival model that is more general in nature and does not assume any particular reaction mechanism to take place. This generality makes the power-law model less complex than the mechanistic model, which is demonstrated by the fact that the power-law model has only 146 parameters whereas the mechanistic model has 164 parameters. In the beginning of the study it became clear that the parameter values reported in scientific literature could not be used successfully in the model. Therefore, a combination of random search method and Monte-Carlo minimization was used to fit the parameters to experimental measurements.

The parameter estimation method was unable to find a satisfactory parameter set and the squared error remained above 414000. The results show that the simulations got stuck in an equilibrium state where no reactions occurred. This was due to depletion of phosphate. The reason that led to this condition is that the initial concentrations of the metabolites were not a subject to the parameter estimation but they were fixed based on measurements and arbitrary values. Because the solution to a differential equation system is defined by the initial condition, the parameter estimation could not have saved the situation in which the starting point was ill-conditioned. Thus, the initial concentrations should have been varied to find a lower squared error, regardless of the measured or biologically feasible values. However, increased difference to the measured initial concentrations causes increase to the value of the objective function which limits the acceptable variations from measurements.

Besides the previous issue, there were also other factors that decreased the performance of the parameter estimation process. The parameter estimation speed is dependent on the speed of the ODE-system evaluation and the speed of the parameter estimation algorithm. Both of these terms could be improved.

First, the models seemed to be excessively complex compared to the level of detail in the mea-

sured data. The total relative complexities for the mechanistic and power-law models were 30.23 and 27.50, respectively. The power-law model had relative complexity of 1.0 for every reaction except PDH. The difference to the generally more complex mechanistic model was narrow because the mechanistic model had many transport reactions with low complexities (0.5). Thus, the difference between the models was not as large as was expected. Additionally, many of the parameters had very small effect on the objective function. These parameters should have been located and removed from the models at least in the initial phase of parameter estimation. Reducing the amount of parameters greatly speeds up the evaluation of the ODE-system and reduces the time to calculate one MC-step. The increase in computational performance is essential in order to make longer minimization runs or to gain more samples in a multistart method. The best way to reduce the number of parameters is to group less important reactions together. This applies specifically to the linear non-branching parts of the reaction network. In this way the functionality of the network is not constrained, but the number of parameters is reduced.

Second, the parameter estimation algorithm used in this study was suboptimal. Although stochastic methods are preferred over deterministic methods, there are better stochastic choices than the random search Monte-Carlo combination used in this thesis. It was suggested that simulated annealing would perform better for the power-law model with which the problem was it getting stuck in local minima. However, the method of parameter estimation was not a subject in this study, but it definitely is a target for development in future studies of the subject.

Even though the models were excessively complex, some differences between them were identified. It was shown that the power-law model was five times slower than the mechanistic model in terms of calculation time of an MC-step. This is probably due to the long evaluation time of a power function. In addition, it is presumed that the mechanistic model would be more adaptable to a large range of operation conditions because it is based on real reaction mechanisms. However, the power-law model has an important benefit over the mechanistic model; the rate equations can be written as algorithms given the stoichiometrics of the reaction network. This benefit makes the power-law model more scalable and more suitable to computational studies in systems biology. The mechanistic rate equations and their derivatives are written separately by hand for each reaction. As a result the mechanistic model is static, rigid and prone to typing or calculation errors.

The current models could not be used to predict any behavior of the modeled organisms as the models did not work as hoped. However, if these models were eventually fixed with changes proposed earlier in this chapter (lumped reactions, estimated initial concentrations, re-evaluation of parameter ranges and simulated annealing parameter estimation for the power-law model), the models could be used to design modifications to the network. For example, maximizing the xylitol production of XR by screening every deletion of other NADH consuming reactions and their combinations. On the other hand, a glycolytic backbone model could be used in combination with an *in vitro* model of an integrated pathway. This way the efficiency of a non-native pathway could be tested before implementation.

Finally, it should be noted that the eventual model is only as good as the data that was used in parameter estimation. Even though the cultivation experiments were not a part of the present study, the cultivation techniques, sampling and sample preparation should be improved in order to get accurate data. Particularly, the time scales of changes and consequently the sampling points should be re-evaluated. In the present study some details may have been lost in the beginning of the cultivation experiment due to too long sampling intervals.

# Bibliography

Anonymous, BRENDA: The Comprehensive Enzyme Information System, http://www.brenda.uni-koeln.de/, July 29., 2003.

Anonymous, GenomeNet, http://www.genome.ad.jp/, 2004, visited June 21, 2004.

Asanuma, N. and Hino, T., Effects of pH and Energy Supply on Activity and Amount of Pyruvate Formate-Lyase in *Streptococcus bovis*, *Appl. Microbiol. Biotechnol.* **66** (2000) 3773–3777.

Axelsson, L., *Lactic Acid Bacteria: Microbiology and Functional Aspects*, chapter Lactic Acid Bacteria: Classification and Physiology, 2nd edition, Marcel Dekker Inc., New York, 1998, pp. 1–72.

Baldwin, S. A., Perham, R. N. and Stribling, D., Purification and Characterization of the Class-II D-Fructose 1,6-Bisphosphate Aldolase from *Escherichia coli* (Crookes' Strain), *Biochem. J.* **169** (1978) 633–641.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, D. S. and Sorokin, A., The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis* ssp. *lactis* IL1403, *Genome Res.* **11** (2001) 731–753.

Buchholz, A., Hurlebaus, J., Wandrey, C. and Takors, R., Metabolomics: quantification of intracellular metabolite dynamics, *Biomol. Eng.* **19** (2002) 5–15.

Callens, M., Kuntz, D. A. and Opperdoes, F. R., Kinetic properties of fructose bisphosphate aldolase from *Trypanosoma brucei* compared to aldolase from rabbit muscle and *Staphylococcus aureus*, *Mol. Biochem. Parasitol.* **47** (1991) 1–9.

Cleland, W. W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products, I Nomenclature and rate equations, *Biochim. Biophys. Acta* **67** (1963a) 104–137.

Cleland, W. W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products, II Inhibition: nomenclature and theory, *Biochim. Biophys. Acta* **67** (1963b) 173–187.

Cleland, W. W., The kinetics of enzyme-catalyzed reactions with two or more substrates or products, III Prediction of initial velocity and inhibition patterns by inspection, *Biochim. Biophys. Acta* **67** (1963c) 188–196.

Cocaign-Bousquet, M., Even, S., Lindley, N. D. and Loubière, P., Anaerobic sugar metabolism in *Lactococcus lactis*: genetic regulation and enzyme control over pathway flux, *Appl. Microbiol. Biotechnol.* **60** (2002) 24–32.

Cornish-Bowden, A., *Fundamentals of Enzyme Kinetics*, 2nd edition, Portland Press, London, 1999, 343 p.

Crow, V. L. and Pritchard, G. G., Pyruvate Kinase from *Streptococcus lactis*, *Methods Enzymol.* **90** (1982) 165–170.

de Ruyter, P. G. G. A., Kuipers, O. P. and de Vos, W. M., Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin, *Appl. Environ. Microbiol.* **62** (1996) 3662–3667.

de Vos, W. M. and Hugenholtz, J., Engineering metabolic highways in Lactococci and other lactic acid bacteria, *Trends Biotechnol.* **22** (2004) 72–79.

Eyschen, J., Vitoux, B., Michel Marraud, M. T. C. and Branlant, G., Engineered Glycolytic Glyceraldehyde-3-Phosphate Dehydrogenase Binds the *Anti* Conformation of $NAD^+$ Nicotinamide but Does Not Experience A-Specific Hydride Transfer, *Arch. Biochem. Biophys.* **364** (1999) 219–227.

Fenton, A. W. and Blair, J. B., Kinetic and Allosteric Consequences of Mutations in the Subunit and Domain Interfaces and the Allosteric Site of Yeast Pyruvate Kinase, *Arch. Biochem. Biophys.* **397** (2002) 28–39.

Garrigues, C., Loubiere, P., Lindley, N. D. and Cocaign-Bousquet, M., Control of the shift from homolactic acid to mixed acid fermentation in *Lactococcus lactis*: Predominant role of the $NADH/NAD^+$ ratio, *J. Bacteriol* **179** (1997) 5282–5287.

Garrigues, C., Mercade, M., Cocaign-Bousquet, M., Lindley, N. D. and Loubiere, P., Regulation of Pyruvate Metabolism in *Lactococcus lactis* Depends on the Imbalance Between Catabolism and Anabolism, *Biotechnol. Bioeng.* **74** (2001) 108–115.

Goryanin, I., Hodgman, T. C. and Selkov, E., Mathematical simulation and analysis of cellular metabolism and regulation, *Bioinformatics* **15** (1999) 749–758.

Hindmarsh, A. C., ODEPACK, A Systematized Collection of ODE Solvers, in *Scientific Computing*, Elsevier Science, North-Holland, 1983, pp. 55–64.

Hoefnagel, M. H. N., Hugenholtz, J. and Snoep, J. L., Time dependent responses of glycolytic intermediates in a detailed glycolytic model of *Lactococcus lactis* during glucose run-out experiments, *Mol. Biol. Rep.* **29** (2002a) 157–161.

Hoefnagel, M. H. N., Starrenburg, M. J. C., Martens, D. E., Hugenholtz, J., Kleerebezem, M., van Swam, I. I., Bongers, R., Westerhoff, H. V. and Snoep, J. L., Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis, *Microbiology* **148** (2002b) 1003–1013.

Hüdig, H. and Hengstenberg, W., The bacterial phosphoenolpyruvate dependent phosphotransferase system (PTS), *FEBS Letters* **114** (1980) 103–106.

Hudson, J., *Suurin tiede*, Art House, 1994, translation of *The History of Chemistry*, Chapman and Hall 1992, 400 p.

Hugenholtz, J., Sybesma, W., Groot, M. N., Wissenlink, W., Ladero, V., Burgess, K., van Sinderen, D., Piard, J.-C., Eggink, G., Smid, A. J., Savoy, G., Sesma, F., Jansen, T., Hols, P. and Kleerebezem, M., Metabolic engineering of lactic acid bacteria for the production of nutraceuticals, *Antonie van Leeuwenhoek* **82** (2002) 217–235.

Kashket, E. R. and Wilson, T. H., Proton-Coupled Accumulation of Galactoside in *Streptococcus lactis* 7962, *Proc. Nat. Acad. Sci. USA* **70** (1973) 2866–2869.

Kauffman, K. J., Prakash, P. and Edwards, J. S., Advances in flux balance analysis, *Curr. Opin. Biotechnol.* **14** (2003) 491–496.

Knappe, J. and Blaschkowski, H. P., Pyruvate formate-lyase from *Escherichia coli* and its activation system, *Methods Enzymol.* **41B** (1975) 508–518.

Krietsch, W. K. G., Triosephosphate Isomerase from Yeast, *Methods Enzymol.* **41B** (1975) 434–438.

Kuby, S. A., *Enzyme Catalysis, Kinetics and Substrate Binding*, volume 1 of *A Study of Enzymes*, CRC Press, USA, 1991.

Melchiorsen, C. R., Jensen, N. B., Christensen, B., Jokumsen, K. V. and Villadsen, J., Dynamics of Pyruvate Metabolism in *Lactococcus lactis*, *Biotechnol. Bioeng.* **74** (2001) 271–279.

Mendes, P., GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems, *Comput. Appl. Biosci.* **9** (1993) 563–571.

Mendes, P. and Kell, D. B., Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation, *Bioinformatics* **14** (1998) 869–883.

Nagradova, N. K., Study of the Properties of Phosphorylating D-Glyceraldehyde-3-Phosphate Dehydrogenase, *Biochemistry (Moscow)* **66** (2001) 1323–1334.

Nam, J. W., Han, K. H., Yoon, E. S., Shina, D. I., Jin, J. H., Lee, D. H., Lee, S. Y. and Lee, J., In silico analysis of lactate producing metabolic network in *Lactococcus lactis*, *Enzyme Microb. Technol.* **35** (2004) 654–662.

Nicholls, D. G. and Ferguson, S. J., *Bioenergetics3*, 2nd edition, Academic Press, London, 2001, 297 p.

Nielsen, J., Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates, *Biochem. J.* **321** (1997) 133–138.

Novák, L. and Loubiere, P., The Metabolic Network of *Lactococcus lactis*: Distribution of $^{14}$C-Labeled Substrates between Catabolic and Anabolic Pathways, *J. Bacteriol.* **182** (2000) 1136–1143.

Nyyssölä, A., Pihlajaniemi, A., Palva, A., von Weymarn, N. and Leisola, M., Production of Xylitol from D-Xylose by Recombinant *Lactococcus lactis*, *J. Biotechnol.* (2005), submitted.

Pitkänen, J.-P., Törmä, A., Alff, S., Huopaniemi, L., Mattila, P. and Renkonen, R., Excess mannose limits the growth of phosphomannose isomerase PMI40 deletion strain of *Saccharomyces cerevisiae*, *J. Biol. Chem.* **279** (2004) 55737–55743.

Poolman, B., Transporters and their role on LAB cell physiology, *Antonie van Leeuwenhoek* **82** (2002) 147–164.

Porter, E. V., Chassy, B. M. and Holmlund, C. E., Purification and kinetic characterization of a specific glucokinase from *Streptococcus mutans* OMZ70 cells, *Biochim. Biophys Acta* **709** (1982) 178–186.

Pradhan, P. G. and Nadkarni, G. B., Functional multiplicity of phosphoglucose isomerase from *Lactobacillus casei*, *Biochim. Biophys. Acta* **615** (1980) 465–473.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flanery, B. P., *Numerical Recipes in Fortran 77, The Art of Scientific Computing*, volume 1 of *Fortran Numerical Recipes*, 2nd edition, Cambridge University Press, New York, 1992, 973 p.

Rizzi, M., Baltes, M., Theobald, U. and Reuss, M., *In Vivo* Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: II. Mathematical Model, *Biotechnol. Bioeng.* **55** (1997) 592–608.

Saier, Jr., M. H., Chauvaux, S., Cook, G. M., Deutscher, J., Paulsen, I. T., Reizer, J. and Ye, J.-J., Catabolite repression and inducer control in Gram-positive bacteria, *Microbiology* **142** (1996) 217–230.

Scopes, R. K., The Steady-State Kinetics of Yeast Phosphoglycerate Kinase, Anomalous Kinetic Plots and the Effects of Salts on Activity, *Eur. J. Biochem.* **85** (1978) 503–516.

Simon, W. A. and Hofer, H. W., Phosphofructokinases from *Lactobacteriaceae*: II. Purification and properties of phosphofructokinase from *Streptococcus thermophilus*, *Biochim. Biophys. Acta* **661** (1981) 158–163.

Spring, T. G. and Wold, F., The Purification and Characterization of *Escherichia coli* Enolase, *J. Biol. Chem.* **246** (1971) 6797–6802.

Stephanopoulos, G. N., Aristidou, A. A. and Nielsen, J., *Metabolic Engineering Principles and Methodologies*, Academic Press, San Diego, 1998, 725 pp.

Stryer, L., *Biochemistry*, 4 edition, W.H. Freeman and Company, New York, 1995, 1064 pp.

Szwergold, B. S., Ugurbil, K. and Brown, T. R., Properties of Fructose-1,6-bisphosphate Aldolase from *Escherichia coli*: An NMR Analysis, *Arch. Biochem. Biophys.* **317** (1995) 244–252.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., van Dam, K., Westerhoff, H. V. and Snoep, J. L., Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry, *Eur. J. Biochem.* **267** (2000) 5313–5329.

Teusink, B., Walsh, M. C., van Dam, K. and Westerhoff, H. V., The danger of metabolic pathways with turbo design, *Trends Biochem. Sci.* **23** (1998) 162–169.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M. and Reuss, M., *In Vivo* Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: I. Experimental Observations, *Biotechnol. Bioeng.* **55** (1997) 305–316.

Vadeboncoeur, C., Proulx, M. and Trahan, L., Purification of proteins similar to HPr and EI from the oral bacterium *Streptococcus salivarius*. Biochemical and immunochemical properties, *Can. J. Microbiol.* **29** (1983) 1694–1705.

Verduyn, C., van Kleef, R., Frank, Jzn., J., Schreuder, H., van Dijken, J. P. and Scheffers, W. A., Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast *Pichia stipitis*, *Biochem. J.* **226** (1985) 669–677.

Visser, D. and Heijnen, J. J., Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics, *Metab. Eng.* **5** (2003) 164–176.

Visser, D., van der Heijden, R., Mauch, K., Reuss, M. and Heijnen, S., Tendency Modeling: A New Approach to Obtain Simplified Kinetics Models of Metabolism Applied to *Saccharomyces cerevisiae*, *Metab. Eng.* **2** (2000) 252–275.

Visser, D., van Zuylen, G. A., van Dam, J. C., Oudshoorn, A., Eman, M. R., Ras, C., van Gulik, W. M., Frank, J., van Dedem, G. W. K. and Heijnen, J. J., Rapid sampling for Analysis of *in vivo* Kinetics Using the BioScope: A System for Continuous-Pulse Experiments, *Biotechnol. Bioeng.* **76** (2002) 674–681.

Voit, E. O., *Computational Analysis of Biochemical Systems*, Cambridge University Press, Cambridge, 2000, 531 p.

Weiss, T. F., *Transport*, volume 1 of *Cellular Biophysics*, MIT Press, Cambridge, USA, 1995.

Wiechert, W., Modeling and simulation: tools for metabolic engineering, *J. Biotechnol.* **94** (2002) 37–63.

Yun, M., Park, C.-G., Kim, J.-Y. and Park, H.-W., Structural Analysis of Glyceraldehyde 3-Phosphate Dehydrogenase from *Escherichia coli*: Direct Evidence of Substrate Binding and Cofactor-Induced Conformational Changes, *Biochemistry* **39** (2000) 10702–10710.

# Part III

# Appendices

# Appendix A

# GAPDH rate equation

The rate equation of glyceraldehyde phosphate dehydrogenase (GAPDH) was not available in the literature. Instead the reaction mechanism of *Escherichia coli* GAPDH has been studied by Nagradova (2001) and Yun *et al.* (2000). The rate equation is derived by the method of King and Altman (Cornish-Bowden, 1999; Kuby, 1991) from the reaction mechanism proposed by Nagradova (2001) shown in figure A.1. The whole reaction is

$$GAP + NAD^+ + P_i \rightleftharpoons 1,3PG + NADH \qquad (A.1)$$
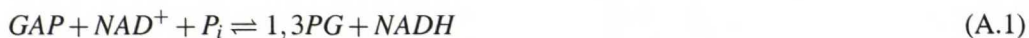
The enzyme complexes are denoted as

$E_1$ = E NAD$^+$

$E_2$ = E NAD$^+$ GAP $\rightleftharpoons$ E-GAP NADH

$E_3$ = E-GAP

$E_4$ = E-GAP NAD$^+$ and

$E_5$ = E-GAP NAD$^+$ Pho.

Hence, the concentration term of each enzyme complex constitutes of a sum of five terms according to the King-Altman patterns (not shown). Here concentration of species A is denoted as [A] to enhance the readability of the equations and $\sum$ denotes the sum of all enzyme complex terms below.
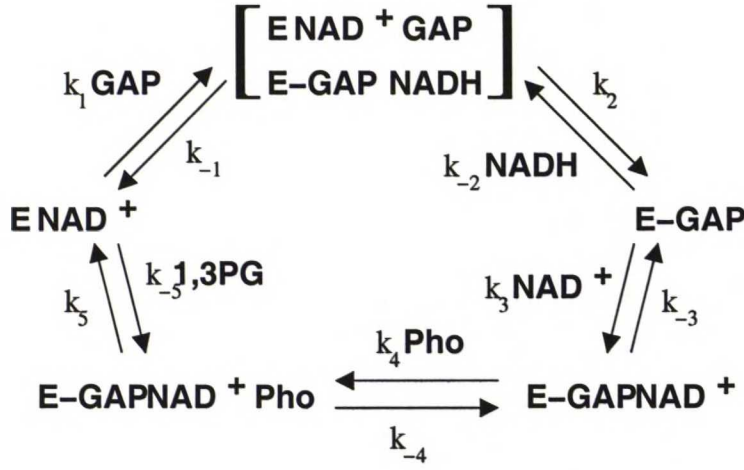
**Figure A.1:** Reaction mechanism of GAPDH according to Nagradova (2001). Dash between E and GAP represents a covalent bond binding the enzyme with the transferred group.

$$\frac{\sum \cdot [\text{E}_1]}{e_0} = k_{-1}k_{-2}[\text{NADH}]k_{-3}k_{-4} + k_2k_3[\text{NAD}^+]k_4[\text{Pi}]k_5 +$$

$$k_{-1}k_3[\text{NAD}^+]k_4[\text{Pi}]k_5 + k_{-1}k_{-2}[\text{NADH}]k_4[\text{Pi}]k_5 +$$

$$k_{-1}k_{-2}[\text{NADH}]k_{-3}k_5 \tag{A.2}$$

$$\frac{\sum \cdot [\text{E}_2]}{e_0} = k_1[\text{GAP}]k_{-2}[\text{NADH}]k_{-3}k_{-4} + k_{-2}[\text{NADH}]k_{-3}k_{-4}k_{-5}[1,3\text{PG}] +$$

$$k_1[\text{GAP}]k_3[\text{NAD}^+]k_4[\text{Pi}]k_5 + k_1[\text{GAP}]k_{-2}[\text{NADH}]k_4[\text{Pi}]k_5 +$$

$$k_1[\text{GAP}]k_{-2}[\text{NADH}]k_{-3}k_5 \tag{A.3}$$

$$\frac{\sum \cdot [\text{E}_3]}{e_0} = k_1[\text{GAP}]k_2k_{-3}k_{-4} + k_2k_{-3}k_{-4}k_{-5}[1,3\text{PG}] +$$

$$k_{-1}k_{-3}k_{-4}k_{-5}[1,3\text{PG}] + k_1[\text{GAP}]k_2k_4[\text{Pi}]k_5 +$$

$$k_1[\text{GAP}]k_2k_{-3}k_5 \tag{A.4}$$

$$\frac{\sum \cdot [\text{E}_4]}{e_0} = k_1[\text{GAP}]k_2k_3[\text{NAD}^+]k_{-4} + k_2k_3[\text{NAD}^+]k_{-4}k_{-5}[1,3\text{PG}] +$$

$$k_{-1}k_3[\text{NAD}^+]k_{-4}k_{-5}[1,3\text{PG}] + k_{-1}k_{-2}[\text{NADH}]k_{-4}k_{-5}[1,3\text{PG}] +$$

$$k_1[\text{GAP}]k_2k_3[\text{NAD}^+]k_5 \tag{A.5}$$

$$\frac{\sum \cdot [\text{E}_5]}{e_0} = k_1[\text{GAP}]k_2k_3[\text{NAD}^+]k_4[\text{Pi}] + k_2k_3[\text{NAD}^+]k_4[\text{Pi}]k_{-5}[1,3\text{PG}] +$$

$$k_{-1}k_3[\text{NAD}^+]k_4[\text{Pi}]k_{-5}[1,3\text{PG}] + k_{-1}k_{-2}[\text{NADH}]k_4[\text{Pi}]k_{-5}[1,3\text{PG}] +$$

$$k_{-1}k_{-2}[\text{NADH}]k_{-3}k_{-5}[1,3\text{PG}] \tag{A.6}$$

Then the reaction rate is set equal to the formation of 1,3PG.

$$v = \frac{d[1,3\text{PG}]}{dt} = k_5[\text{E}_5] - k_{-5}[1,3\text{PG}][\text{E}_1] \tag{A.7}$$

Substituting $[E_5]$ and $[E_1]$ from equations A.6 and A.2, respectively, and denoting the coefficients of the numerator as $N_i$ yields

$$v = \frac{e_0(N_1[\text{GAP}][\text{NAD}^+][\text{Pi}] - N_2[1,3\text{PG}][\text{NADH}])}{\Sigma} \tag{A.8}$$

where $N_1 = k_1 k_2 k_3 k_4 k_5$ and $N_2 = k_{-1} k_{-2} k_{-3} k_{-4} k_{-5}$. Similarly the denominator term $\Sigma$ is written as a sum of the right hand sides of equations A.2 thru A.6. Denoting the coefficients of the denominator as $D_i$ yields

$$
\begin{aligned}
\Sigma = \ & D_1[\text{GAP}] + D_2[1,3\text{PG}] + D_3[\text{NADH}] + \\
& D_4[\text{GAP}][\text{NAD}^+] + D_5[\text{GAP}][\text{NADH}] + \\
& D_6[\text{GAP}][\text{Pi}] + D_7[1,3\text{PG}][\text{NAD}^+] + \\
& D_8[1,3\text{PG}][\text{NADH}] + D_9[\text{NAD}^+][\text{Pi}] + D_{10}[\text{NADH}][\text{Pi}] + \\
& D_{11}[\text{GAP}][\text{Pi}][\text{NAD}^+] + D_{12}[\text{GAP}][\text{Pi}][\text{NADH}] + \\
& D_{13}[1,3\text{PG}][\text{Pi}][\text{NAD}^+] + D_{14}[1,3\text{PG}][\text{Pi}][\text{NADH}]
\end{aligned}
\tag{A.9}
$$

where

$$
\begin{aligned}
D_1 &= k_1 k_2 k_{-3}(k_{-4} + k_5); & D_2 &= k_{-3} k_{-4} k_{-5}(k_{-1} + k_2) \\
D_3 &= k_{-1} k_{-2} k_{-3}(k_{-4} + k_5); & D_4 &= k_1 k_2 k_3(k_{-4} + k_5) \\
D_5 &= k_1 k_{-2} k_{-3}(k_{-4} + k_5); & D_6 &= k_1 k_2 k_4 k_5 \\
D_7 &= k_3 k_{-4} k_{-5}(k_{-1} + k_2); & D_8 &= k_{-2} k_{-5}(k_{-3} k_{-4} + k_{-1} k_{-3} + k_{-1} k_{-4}) \\
D_9 &= k_3 k_4 k_5(k_{-1} + k_2); & D_{10} &= k_{-1} k_{-2} k_4 k_5 \\
D_{11} &= k_1 k_3 k_4(k_2 + k_5); & D_{12} &= k_1 k_{-2} k_4 k_5 \\
D_{13} &= k_3 k_4 k_{-5}(k_{-1} + k_2); & D_{14} &= k_{-1} k_{-2} k_4 k_{-5}.
\end{aligned}
$$

The rate equation has 11 free parameters (the $k$'s) altogether since the numerator and denominator coefficients depend on each other. Eventually it could be advantageous to write equation A.9 to similar form as the ternary-complex rate equations. Furthermore it could give more insight to the meaning of the parameters if they were written in terms of $K_m$'s and $K_i$'s instead of $N$'s and $D$'s. However, this may not be possible for all rate equations since Michaelis constants require some mechanistic properties of the reaction. If the mechanisms are fundamentally different, writing the parameters in Michaelis constants would be misleading.

# Appendix B

# Experimental measurements

*Co-authors: Antti Nyyssölä and Kristiina Kiviharju, Laboratory of Bioprocess Engineering, Helsinki university of Technology*

## B.1   Materials and methods

### B.1.1   Microbial strains, growth media and cultivation conditions

The *Lactococcus lactis* strains used in the study were grown at 30°C in M17 broth (Difco) containing 10 g/L glucose and 8 mg/L chloramphenicol. The construction of the expression plasmids is described elsewhere (Nyyssölä *et al.*, 2005). *L. lactis* LLXR carries the gene of the xylose reductase from *Pichia stipitis* CBS 5773 cloned into the expression vector pNZ8032. *L. lactis* LLCON carrying the plasmid pNZ8037 was used as the control strain. The host strain *L. lactis* NZ9800 and the vectors pNZ8032 and pNZ8037 belong to the nisin inducible expression system (NICE) described previously (de Ruyter *et al.*, 1996).

### B.1.2   Preparation of cell suspensions for bioconversion experiments

5 L of growth medium was inoculated with 25 mL of recombinant *L. lactis* cell culture propagated to the stationary phase. The cells were grown for 2 h and 0.8 g/L nisin was added as the inducer. The cultures were further grown for 11 h and the cells were harvested by centrifugation for 5 min at 5000 g at +4°C.

The conversion buffer used in the experiments contained 25 g/L xylose, 75 g/L glucose, 2.2 g/L M17, 20 mM potassium phosphate buffer pH 6.5, 2 mM MgSO4, 8 mg/L chloramphenicol and 0.8 g/L nisin. The harvested cells were washed with 500 mL of 1.2× concentrated conversion buffer without M17 and glucose. The washed cells were suspended into 500 mL of the same buffer, transferred into the bioreactor and 100 mL of a solution containing M17 and glucose was added at the final concentrations.

### B.1.3 Bioreactor experiments

A Biostat Q (B. Braun Biotech International, Melsungen, Germany) bioreactor with magnetic stirring was used in the bioconversion experiments. The working volume was 550-600 mL, temperature 30°C and stirring rate 300 rpm. The pH was controlled at 6.5 with 4 M KOH. A slow stream of nitrogen was passed into the headspace of the bioreactor in order to minimize the effects of oxygen. MFCS program (version 2.1, Sartorius BBI Systems, Göttingen, Germany) was used for monitoring the conditions in the bioreactor.

4 mL samples were drawn from the bioreactor intermittently using an automatic sampling system (Fermentation Sampling Control 1.2, Medicel Oy, Helsinki, Finland). The samples were pumped into 15 mL Greiner tubes containing 8 mL 70 % methanol. The tubes were kept at -37°C in a cooling bath. The cell suspensions were centrifuged for 10 min at 7200 g at -5°C and the supernatants and the cell pellets were separated. The samples were stored at -80°C.

### B.1.4 Analytical techniques

Cell concentrations were analyzed at the beginning and the end of the cultivation from 5 mL samples. Cells were separated by centrifugation for 5 min at 5000 g, washed once with 3 mL saline and once with 3 mL $H_2O$. The cell dry weight was measured after drying the samples for 16 h at 75°C.

The analysis of internal and external metabolites is explained elsewhere (Pitkänen *et al.*, 2004). Briefly, external metabolites are quantitated with HPLC from the supernatants using an Aminex column (HPX-87H, Bio-Rad Inc., Hercules, USA) and Waters 2414 refractive index detector (Waters). Internal metabolites were extracted from the cell pellets in boiling ethanol. The samples were first dried and then diluted in MQ water. Concentrations were analyzed with an anion exchange LC-MSMS method.

## B.2 Results

The concentration profiles of the internal metabolites show some differences between the strains (Figure B.1). The internal metabolites with visually different profiles are 2PG, FBP, PEP and ATP (but not ADP). The LLXR strain has lower concentrations of glycolytic intermediates 2PG, FBP and PEP, which indicates a higher overall glycolytic reaction rate, possibly caused by the extra demand for NADH by XR. The higher glycolytic rate would also explain the high concentration of ATP in the LLXR strain. However, the confidence to these results is undermined by the fact that despite the changes in ATP concentration between the strain, no changes in ADP were detected. It is possible, though, that the overall pool of ATP and ADP is larger in the LLXR strain.
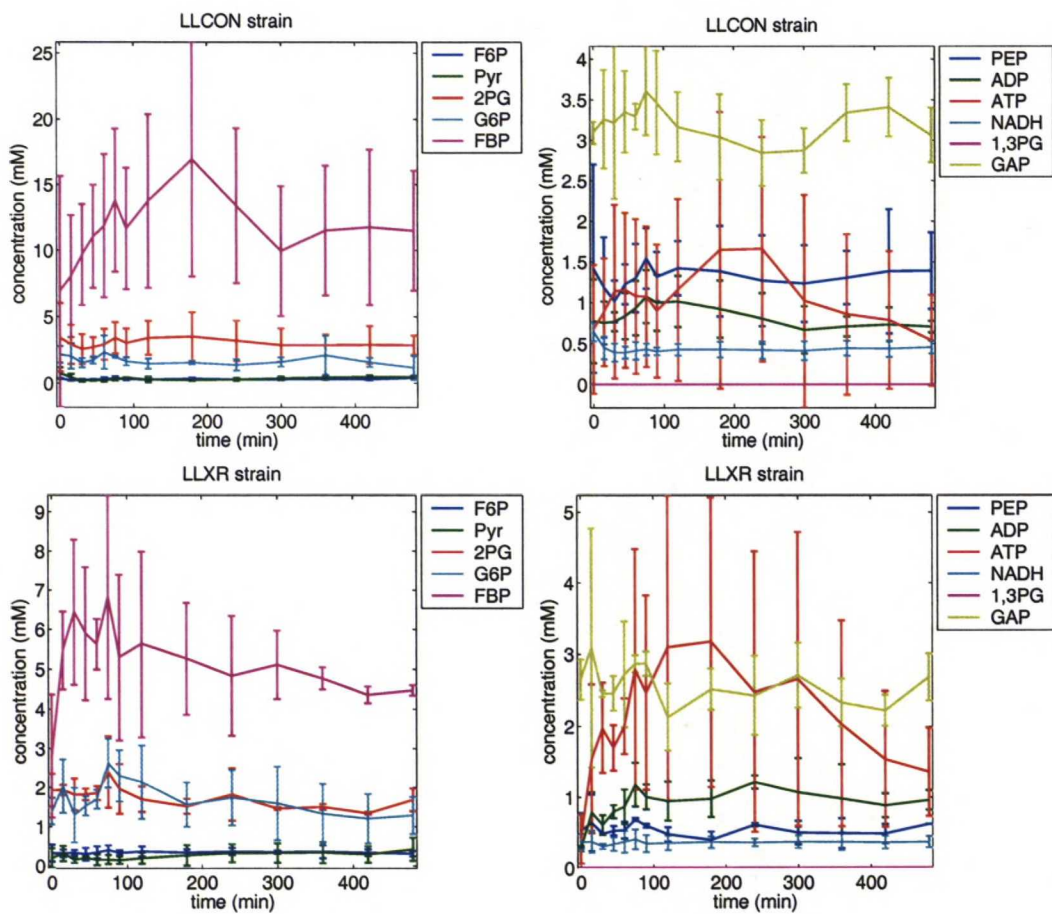
**Figure B.1:** Comparison of the measured internal concentrations for the LLCON-strain (upper figures) and the LLXR-strain (lower figures) together with the standard deviation of the conversion experiments.
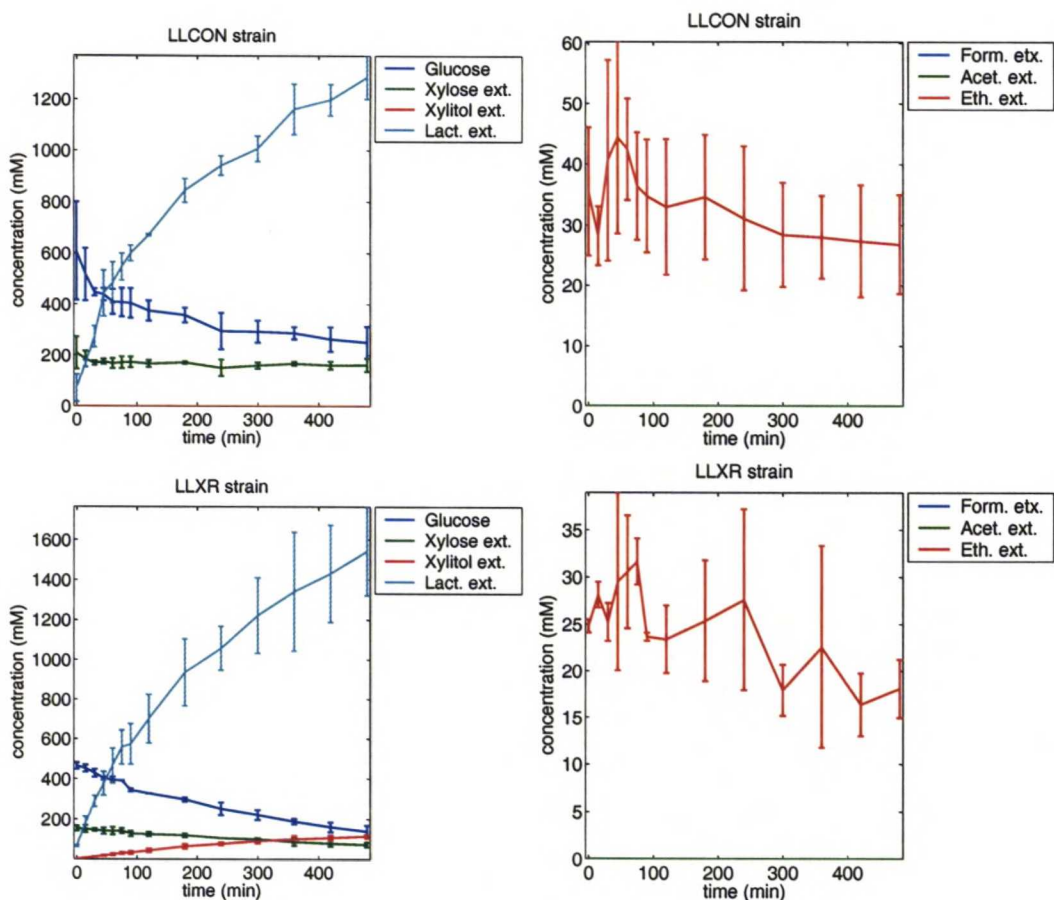
**Figure B.2:** Comparison of the measured external concentrations for the LLCON-strain (upper figures) and the LLXR-strain (lower figures) together with the standard deviation of the conversion experiments.

The external metabolite concentration profiles also have some visual differences (Figure B.2). Naturally, only the LLXR strain utilizes xylose and produces xylitol. Additionally the LLXR strain seems to use more glucose and produce more lactate than the LLCON strain. This observation also suggests that the LLXR strain has higher glycolytic rate. The figures also indicate that the LLCON strain has higher ethanol concentration. This result is controversial since the equipment was rinsed with ethanol between samples. It is highly likely that the ethanol is from a non-bacterial source because there was a lot of variance between the samples and no clear trend was observed.

The visual differences are confirmed by the numerical data shown in tables B.2 and B.2. It is noteworthy that no formate or acetate was detected from the medium. Therefore, given the controversial results for ethanol, it seems that the only active branch of pyruvate metabolism was the LDH branch. This indicates purely homolactic fermentation mode for both strains.

**Table B.1:** Concentrations for the internal metabolites of the LLCON strain. Concentrations are determined as an average of three parallel experiments.

| Time (min) | G6P (mmol/L) | F6P (mmol/L) | FBP (mmol/L) | GAP (mmol/L) | 1,3PG (mmol/L) | 2PG (mmol/L) | PEP (mmol/L) | Pyr (mmol/L) | ATP (mmol/L) | NADH (mmol/L) | NAD$^+$ (mmol/L) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **LLXR strain** | | | | | | |
| 0 | 1.39±0.34 | 0.32±0.22 | 2.79±1.57 | 2.64±0.28 | 0.00±0.00 | 1.92±0.42 | 0.52±0.25 | 0.19±0.25 | 0.41±0.34 | 0.36±0.11 | 0.27±0.04 |
| 15 | 2.02±0.68 | 0.31±0.10 | 5.46±0.98 | 3.09±1.68 | 0.00±0.00 | 1.92±0.13 | 0.63±0.40 | 0.31±0.20 | 1.53±1.04 | 0.37±0.16 | 0.78±0.28 |
| 30 | 1.28±0.70 | 0.32±0.07 | 6.44±1.84 | 2.45±0.05 | 0.00±0.00 | 1.81±0.41 | 0.48±0.01 | 0.14±0.08 | 1.97±0.63 | 0.30±0.04 | 0.60±0.14 |
| 45 | 1.52±0.25 | 0.29±0.16 | 5.89±1.68 | 2.46±0.24 | 0.00±0.00 | 1.81±0.15 | 0.52±0.07 | 0.19±0.03 | 1.69±0.32 | 0.33±0.09 | 0.79±0.09 |
| 60 | 1.69±0.24 | 0.33±0.16 | 5.61±0.64 | 2.72±0.73 | 0.00±0.00 | 1.85±0.18 | 0.53±0.12 | 0.15±0.14 | 2.00±0.39 | 0.37±0.15 | 0.86±0.24 |
| 75 | 2.61±0.63 | 0.39±0.12 | 6.83±2.59 | 2.86±0.12 | 0.00±0.00 | 2.38±0.91 | 0.68±0.02 | 0.15±0.11 | 2.79±1.69 | 0.40±0.14 | 1.17±0.31 |
| 90 | 2.28±0.66 | 0.32±0.07 | 5.29±2.09 | 2.87±0.17 | 0.00±0.00 | 1.95±0.64 | 0.59±0.03 | 0.13±0.09 | 2.46±1.36 | 0.34±0.12 | 1.00±0.18 |
| 120 | 2.12±0.94 | 0.36±0.13 | 5.63±2.36 | 2.12±0.47 | 0.00±0.00 | 1.69±0.33 | 0.47±0.10 | 0.18±0.14 | 3.09±2.14 | 0.35±0.10 | 0.94±0.28 |
| 180 | 1.55±0.57 | 0.33±0.08 | 5.26±1.41 | 2.51±0.28 | 0.00±0.00 | 1.51±0.19 | 0.39±0.12 | 0.26±0.26 | 3.17±2.03 | 0.36±0.08 | 0.97±0.26 |
| 240 | 1.73±0.71 | 0.35±0.09 | 4.82±1.51 | 2.43±0.55 | 0.00±0.00 | 1.81±0.67 | 0.61±0.03 | 0.31±0.23 | 2.48±1.97 | 0.36±0.06 | 1.21±0.09 |
| 300 | 1.59±0.94 | 0.34±0.04 | 5.10±0.85 | 2.70±0.45 | 0.00±0.00 | 1.45±0.03 | 0.49±0.17 | 0.32±0.25 | 2.65±2.06 | 0.37±0.09 | 1.06±0.48 |
| 360 | 1.32±0.77 | 0.35±0.17 | 4.76±0.28 | 2.33±0.33 | 0.00±0.00 | 1.50±0.08 | 0.49±0.21 | 0.33±0.27 | 2.02±1.44 | 0.36±0.12 | 0.97±0.48 |
| 420 | 1.19±0.63 | 0.32±0.09 | 4.35±0.21 | 2.22±0.22 | 0.00±0.00 | 1.33±0.03 | 0.48±0.14 | 0.28±0.21 | 1.53±0.96 | 0.36±0.09 | 0.88±0.17 |
| 480 | 1.28±0.47 | 0.30±0.07 | 4.46±0.13 | 2.68±0.32 | 0.00±0.00 | 1.68±0.29 | 0.62±0.00 | 0.41±0.30 | 1.35±0.63 | 0.36±0.08 | 0.95±0.14 |
| | | | | | **LLCON strain** | | | | | | |
| 0 | 2.15±0.62 | 0.33±0.02 | 6.91±8.73 | 3.09±0.14 | 0.00±0.00 | 3.42±2.58 | 1.42±1.28 | 0.74±0.43 | 0.67±0.79 | 0.64±0.12 | 0.77±0.51 |
| 15 | 2.02±0.95 | 0.21±0.04 | 8.06±4.61 | 3.26±0.61 | 0.00±0.00 | 2.90±1.47 | 1.19±0.61 | 0.42±0.26 | 0.88±0.66 | 0.45±0.14 | 0.75±0.26 |
| 30 | 1.52±0.35 | 0.23±0.07 | 9.68±3.83 | 3.22±0.94 | 0.00±0.00 | 2.54±1.12 | 1.02±0.26 | 0.17±0.09 | 1.14±1.07 | 0.39±0.11 | 0.77±0.12 |
| 45 | 1.72±0.28 | 0.25±0.04 | 11.06±3.92 | 3.35±0.50 | 0.00±0.00 | 2.69±0.75 | 1.22±0.25 | 0.23±0.06 | 1.15±0.95 | 0.39±0.07 | 0.84±0.28 |
| 60 | 2.31±1.26 | 0.31±0.11 | 11.87±5.43 | 3.29±0.16 | 0.00±0.00 | 2.91±1.15 | 1.30±0.42 | 0.21±0.12 | 1.08±0.95 | 0.41±0.11 | 0.93±0.33 |
| 75 | 1.97±0.10 | 0.34±0.15 | 13.83±5.45 | 3.61±0.55 | 0.00±0.00 | 3.40±1.19 | 1.55±0.37 | 0.30±0.11 | 1.07±0.86 | 0.43±0.07 | 1.07±0.32 |
| 90 | 1.63±0.31 | 0.28±0.07 | 11.66±4.62 | 3.46±0.64 | 0.00±0.00 | 3.01±1.08 | 1.32±0.30 | 0.39±0.04 | 0.90±0.81 | 0.40±0.05 | 1.00±0.34 |
| 120 | 1.46±0.39 | 0.26±0.03 | 13.76±6.61 | 3.17±0.43 | 0.00±0.00 | 3.40±1.28 | 1.42±0.33 | 0.26±0.20 | 1.16±1.12 | 0.43±0.07 | 1.01±0.31 |
| 180 | 1.53±0.13 | 0.32±0.08 | 16.92±8.91 | 3.04±0.53 | 0.00±0.00 | 3.50±1.82 | 1.39±0.56 | 0.23±0.08 | 1.65±1.70 | 0.43±0.10 | 0.92±0.35 |
| 240 | 1.35±0.43 | 0.26±0.06 | 13.40±5.89 | 2.84±0.41 | 0.00±0.00 | 3.18±1.51 | 1.27±0.56 | 0.26±0.10 | 1.66±1.38 | 0.42±0.11 | 0.81±0.31 |
| 300 | 1.57±0.35 | 0.27±0.06 | 9.95±4.93 | 2.88±0.28 | 0.00±0.00 | 2.85±1.26 | 1.24±0.47 | 0.34±0.12 | 1.02±1.30 | 0.41±0.12 | 0.67±0.29 |
| 360 | 2.09±1.47 | 0.29±0.04 | 11.52±4.93 | 3.34±0.35 | 0.00±0.00 | 2.85±0.79 | 1.31±0.33 | 0.41±0.17 | 0.86±0.99 | 0.44±0.09 | 0.71±0.12 |
| 420 | 1.55±0.35 | 0.27±0.05 | 11.77±5.91 | 3.41±0.36 | 0.00±0.00 | 2.87±1.42 | 1.39±0.76 | 0.45±0.12 | 0.79±0.84 | 0.44±0.10 | 0.73±0.21 |
| 480 | 1.15±0.89 | 0.39±0.15 | 11.51±4.56 | 3.07±0.34 | 0.00±0.00 | 2.84±0.73 | 1.39±0.47 | 0.44±0.04 | 0.54±0.56 | 0.46±0.08 | 0.71±0.07 |

**Table B.2:** Concentrations for the external metabolites of both strains and the standard deviation of the parallel conversion experiments. Concentrations are determined as an average of two and three parallel experiments for the LLXR and LLCON strains, respectively. Notice that the samples have been diluted with the KOH solution used in pH adjustment during the conversion experiments. Therefore, these volume in different samples is not the same.

| Time (min) | Glucose (mmol/L) | Lactate (mmol/L) | Formate (mmol/L) | Ethanol (mmol/L) | Acetate (mmol/L) | Xylose (mmol/L) | Xylitol (mmol/L) |
|---|---|---|---|---|---|---|---|
| | | | **LLXR strain** | | | | |
| 0 | 465.44±16.58 | 66.12±5.10 | 0.00±0.00 | 24.81±0.71 | 0.00±0.00 | 154.99±12.75 | 1.72±0.80 |
| 15 | 452.51±18.07 | 181.14±33.81 | 0.00±0.00 | 28.12±1.36 | 0.00±0.00 | 152.08±18.75 | 7.48±0.02 |
| 30 | 428.75±19.96 | 289.39±27.80 | 0.00±0.00 | 25.23±2.03 | 0.00±0.00 | 147.88±7.10 | 12.99±0.61 |
| 45 | 402.43±7.83 | 377.19±59.46 | 0.00±0.00 | 29.50±9.44 | 0.00±0.00 | 142.30±16.20 | 18.20±3.20 |
| 60 | 394.31±15.28 | 468.20±84.44 | 0.00±0.00 | 30.54±5.99 | 0.00±0.00 | 142.06±18.79 | 24.17±5.29 |
| 75 | 391.30±0.50 | 558.38±85.54 | 0.00±0.00 | 31.64±2.44 | 0.00±0.00 | 143.21±13.61 | 30.68±6.06 |
| 90 | 344.19±8.21 | 574.04±101.88 | 0.00±0.00 | 23.67±0.42 | 0.00±0.00 | 128.71±13.62 | 33.22±7.35 |
| 120 | 327.35±0.80 | 701.33±123.04 | 0.00±0.00 | 23.39±3.61 | 0.00±0.00 | 125.74±11.13 | 44.04±8.94 |
| 180 | 297.68±10.07 | 935.94±169.49 | 0.00±0.00 | 25.35±6.43 | 0.00±0.00 | 120.27±9.24 | 64.66±11.76 |
| 240 | 251.90±31.47 | 1057.99±110.26 | 0.00±0.00 | 27.57±9.62 | 0.00±0.00 | 106.88±0.38 | 76.96±6.55 |
| 300 | 221.81±24.83 | 1222.12±188.43 | 0.00±0.00 | 17.94±2.76 | 0.00±0.00 | 100.93±5.30 | 90.80±10.69 |
| 360 | 188.77±13.59 | 1341.73±297.16 | 0.00±0.00 | 22.54±10.76 | 0.00±0.00 | 87.41±18.74 | 100.38±17.36 |
| 420 | 161.81±24.38 | 1431.68±242.68 | 0.00±0.00 | 16.39±3.39 | 0.00±0.00 | 79.48±15.03 | 106.95±11.84 |
| 480 | 139.98±29.87 | 1543.25±221.27 | 0.00±0.00 | 18.12±3.15 | 0.00±0.00 | 73.19±11.74 | 114.60±8.58 |
| | | | **LLCON strain** | | | | |
| 0 | 607.36±192.10 | 71.59±53.50 | 0.00±0.00 | 35.47±10.53 | 0.00±0.00 | 208.85±63.16 | 0.00±0.00 |
| 15 | 516.16±102.88 | 167.35±12.07 | 0.00±0.00 | 28.18±4.87 | 0.00±0.00 | 185.20±31.30 | 0.00±0.00 |
| 30 | 446.04±14.27 | 272.79±41.40 | 0.00±0.00 | 40.64±16.56 | 0.00±0.00 | 168.96±8.68 | 0.00±0.00 |
| 45 | 437.20±24.18 | 441.61±90.28 | 0.00±0.00 | 44.39±15.83 | 0.00±0.00 | 175.30±10.72 | 0.00±0.00 |
| 60 | 408.73±48.99 | 483.64±81.16 | 0.00±0.00 | 42.43±8.37 | 0.00±0.00 | 167.84±19.66 | 0.00±0.00 |
| 75 | 406.41±55.95 | 545.78±53.26 | 0.00±0.00 | 36.37±8.84 | 0.00±0.00 | 170.89±23.02 | 0.00±0.00 |
| 90 | 403.65±57.95 | 599.45±30.44 | 0.00±0.00 | 34.74±9.27 | 0.00±0.00 | 172.91±20.96 | 0.00±0.00 |
| 120 | 373.72±39.74 | 670.46±4.22 | 0.00±0.00 | 32.93±11.18 | 0.00±0.00 | 166.49±13.74 | 0.00±0.00 |
| 180 | 356.53±28.97 | 843.41±46.04 | 0.00±0.00 | 34.57±10.25 | 0.00±0.00 | 171.43±4.82 | 0.00±0.00 |
| 240 | 294.29±71.04 | 940.23±37.80 | 0.00±0.00 | 31.05±11.85 | 0.00±0.00 | 150.63±32.49 | 0.00±0.00 |
| 300 | 292.13±42.87 | 1005.30±49.79 | 0.00±0.00 | 28.36±8.60 | 0.00±0.00 | 159.61±12.53 | 0.00±0.00 |
| 360 | 286.53±24.18 | 1159.67±99.60 | 0.00±0.00 | 27.98±6.83 | 0.00±0.00 | 166.79±7.16 | 0.00±0.00 |
| 420 | 262.61±47.67 | 1196.03±61.46 | 0.00±0.00 | 27.31±9.23 | 0.00±0.00 | 160.57±15.41 | 0.00±0.00 |
| 480 | 250.50±61.90 | 1284.63±85.71 | 0.00±0.00 | 26.78±8.16 | 0.00±0.00 | 161.61±26.01 | 0.00±0.00 |

# Appendix C

# Computational details

## C.1 Stoichiometric matrix

Table C.1 presents the stoichiometric matrix of the model network as describes earlier in figure 6.1. The stoichiometric matrix lists all 29 reactions and 38 metabolites incorporated into the model. However, the proton concentrations (H_in and H_ext, metabolites 30 and 31, respectively) and the proton pump ($H^+$ATPase, reaction 25) included in the stoichiometric matrix are not included in the current version of the model.

## C.2 Mechanistic model

Further details of the mechanistic model are listed in table C.2. The table gives an interpretation to the parameter number used later on. It also includes references to the equations that have been used to model the reactions and tells which metabolites have been used in each position in the equation. It is noteworthy that these position are important; they suggest a role for the metabolite in the mechanism for some reactions.

**Table C.1:** The stoichiometric matrix of the model network. Proton concentrations and proton pump are not yet implemented. Abbreviations: perm., permease; transp., transport. Other abbreviations are presented previously.

| | Glucose perm. | EI | EII | GLK | PGI | PFK | FBA | TPI | GAPDH | PGK | PMG | ENO | PK | LDH | PFL | PTA | ACK | ACDH | ADHE | Pho perm. | Lactate transp. | Acetate transp. | Ethanol transp. | Formate transp. | H⁺ ATPase | PDH | Xylose transp. | XR | Xylitol transp. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glucose | -1 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLC | 1 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G6P | 0 | 0 | 1 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F6P | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FBP | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DHAP | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GAP | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1,3PG | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3PG | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2PG | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PEP | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pyr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 |
| Lactate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lactate_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Fomate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 |
| Formate_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| AcCoA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Acetaldehyde | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ethanol | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ethanol_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| AcetylP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acetate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acetate_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| XLS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | 0 |
| Xylose_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 |
| XLT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 |
| Xylitol_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pho | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pho_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| H_in | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| H_ext | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| ATP | 0 | 0 | 0 | -1 | 0 | -1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| NADH | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | -1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -1 | 0 |
| HPr | 0 | -1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ADP | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | -1 | 0 | 0 | -1 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| NAD⁺ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 1 | 0 |
| HPr-P | 0 | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CoA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 |

**Table C.2:** Parameters of the mechanistic model. Abbreviations: Acetald., Acetaldehyde. Other abbreviations are presented previously.

| # | Term | Reaction | Rate Eqn. | Notes |
|---|------|----------|-----------|-------|
| 1 | Km_glucose | Glucose_perm | 4.3 | A=Glucose, P=Glc |
| 2 | Km_glc | | | |
| 3 | Vmax(forw) | | | |
| 4 | Vmax(rev) | | | |
| 5 | Km_pep | PTS I | 4.10 | A=PEP, B=HPr, P=HPr-P, Q=Pyr |
| 6 | Km_hpr | | | |
| 7 | Km_pyr | | | |
| 8 | Km_hprp | | | |
| 9 | Ki_pep | | | |
| 10 | Ki_pyr | | | |
| 11 | Ki_hprp | | | |
| 12 | Vmax(forw) | | | |
| 13 | Vmax(rev) | | | |
| 14 | Km_hprp | PTS II | 4.10 | A=HPr-P, B=Glucose, P=G6P, Q=HPr |
| 15 | Km_glucose | | | |
| 16 | Km_hpr | | | |
| 17 | Km_g6p | | | |
| 18 | Ki_hprp | | | |
| 19 | Ki_hpr | | | |
| 20 | Ki_g6p | | | |
| 21 | Vmax(forw) | | | |
| 22 | Vmax(rev) | | | |
| 23 | Km_glc | GLK | 4.8 | A=Glc, B=ATP, P=ADP, Q=G6P |
| 24 | Km_atp | | | |
| 25 | Km_g6p | | | |
| 26 | Km_adp | | | |
| 27 | Ki_glc | | | |
| 28 | Ki_atp | | | |
| 29 | Ki_g6p | | | |
| 30 | Ki_adp | | | |
| 31 | Vmax(forw) | | | |
| 32 | Vmax(rev) | | | |
| 33 | Km_g6p | PGI | 4.3 | A=G6P, P=F6P |
| 34 | Km_f6p | | | |
| 35 | Vmax(forw) | | | |
| 36 | Vmax(rev) | | | |
| 37 | Km_f6p | PFK | 4.8 | A=F6P, B=ATP, P=ADP, Q=FBP |
| 38 | Km_atp | | | |
| 39 | Km_fbp | | | |
| 40 | Km_adp | | | |
| 41 | Ki_f6p | | | |
| 42 | Ki_atp | | | |

| # | Term | Reaction | Rate Eqn. | Notes |
|---|------|----------|-----------|-------|
| 43 | Ki_fbp | | | |
| 44 | Ki_adp | | | |
| 45 | Vmax(forw) | | | |
| 46 | Vmax(rev) | | | |
| 47 | Km_A | FBA | 4.16 | A=FBP, P=GAP, Q=DHAP |
| 48 | Km_B2 | | | |
| 49 | Ki_A | | | |
| 50 | Ki_B1 | | | |
| 51 | Ki_B2 | | | |
| 52 | K_eq | | | |
| 53 | V_max | | | |
| 54 | Km_dhap | TPI | 4.3 | A=DHAP, P=GAP |
| 55 | Km_gap | | | |
| 56 | Vmax(forw) | | | |
| 57 | Vmax(rev) | | | |
| 58 | k1 | GAPDH | 4.14 | see Appendix A |
| 59 | k2 | | | |
| 60 | k3 | | | |
| 61 | k4 | | | |
| 62 | k5 | | | |
| 63 | k-1 | | | |
| 64 | k-2 | | | |
| 65 | k-3 | | | |
| 66 | k-4 | | | |
| 67 | k-5 | | | |
| 68 | e0 | | | |
| 69 | Km_13pg | PGK | 4.8 | A=1,3PG, B=ADP, P=ATP, Q=3PG |
| 70 | Km_adp | | | |
| 71 | Km_3pg | | | |
| 72 | Km_atp | | | |
| 73 | Ki_13pg | | | |
| 74 | Ki_adp | | | |
| 75 | Ki_3pg | | | |
| 76 | Ki_atp | | | |
| 77 | Vmax(forw) | | | |
| 78 | Vmax(rev) | | | |
| 79 | Km_3pg | PMG | 4.3 | A=3PG, P=2PG |
| 80 | Km_2pg | | | |
| 81 | Vmax(forw) | | | |
| 82 | Vmax(rev) | | | |
| 83 | Km_2pg | ENO | 4.3 | A=2PG, P=PEP |
| 84 | Km_pep | | | |
| 85 | Vmax(forw) | | | |

| #   | Term         | Reaction | Rate Eqn. | Notes                                                    |
|-----|--------------|----------|-----------|----------------------------------------------------------|
| 86  | Vmax(rev)    |          |           |                                                          |
| 87  | Km_pep       | PK       | 4.12      | A=PEP, B=ADP                                             |
| 88  | Km_adp       |          |           |                                                          |
| 89  | h_pep        |          |           |                                                          |
| 90  | h_adp        |          |           |                                                          |
| 91  | Vmax(forw)   |          |           |                                                          |
| 92  | Km_pyr       | LDH      | 4.7       | A=Pyr, B=NADH, P=NAD$^+$, Q=Lactate$_{in}$              |
| 93  | Km_nadh      |          |           |                                                          |
| 94  | Km_nad       |          |           |                                                          |
| 95  | Km_lactate   |          |           |                                                          |
| 96  | K_eq         |          |           |                                                          |
| 97  | Vmax(forw)   |          |           |                                                          |
| 98  | Km_pyr       | PFL      | 4.10      | A=Pyr, B=CoA, P=AcCoA, Q=Formate$_{in}$                 |
| 99  | Km_coa       |          |           |                                                          |
| 100 | Km_formate   |          |           |                                                          |
| 101 | Km_accoa     |          |           |                                                          |
| 102 | Ki_pyr       |          |           |                                                          |
| 103 | Ki_formate   |          |           |                                                          |
| 104 | Ki_accoa     |          |           |                                                          |
| 105 | Vmax(forw)   |          |           |                                                          |
| 106 | Vmax(rev)    |          |           |                                                          |
| 107 | Km_pho       | PTA      | 4.9       | A=AcCoA, B=Pho, P=CoA, Q=Acetyl-P                       |
| 108 | Km_coa       |          |           |                                                          |
| 109 | Ki_accoa     |          |           |                                                          |
| 110 | Ki_pho       |          |           |                                                          |
| 111 | Ki_acetylp   |          |           |                                                          |
| 112 | Ki_coa       |          |           |                                                          |
| 113 | Vmax(forw)   |          |           |                                                          |
| 114 | Vmax(rev)    |          |           |                                                          |
| 115 | Km_acetylp   | ACK      | 4.7       | A=Acetyl-P, B=ADP, P=ATP, Q=Acetate$_{in}$              |
| 116 | Km_adp       |          |           |                                                          |
| 117 | Km_atp       |          |           |                                                          |
| 118 | Km_acetate   |          |           |                                                          |
| 119 | K_eq         |          |           |                                                          |
| 120 | Vmax(forw)   |          |           |                                                          |
| 121 | Km_AcCoA     | ACDH     | 4.18      |                                                          |
| 122 | Km_nadh      |          |           |                                                          |
| 123 | Km_nad       |          |           |                                                          |
| 124 | Km_CoA       |          |           |                                                          |
| 125 | Km_acetald.  |          |           |                                                          |
| 126 | Keq          |          |           |                                                          |
| 127 | Vmax         |          |           |                                                          |
| 128 | Km_acetald.  | ADHE     | 4.7       | A=Acetaldehyde, B=NADH, P=NAD$^+$, Q=Ethanol$_{in}$     |

*Continues on next page...*

| # | Term | Reaction | Rate Eqn. | Notes |
|---|------|----------|-----------|-------|
| 129 | Km_nadh | | | |
| 130 | Km_nad | | | |
| 131 | Km_ethanol | | | |
| 132 | K_eq | | | |
| 133 | Vmax(forw) | | | |
| 134 | Vm | Phosphate_perm. | 4.15 | A=Phosphate, P=Pho |
| 135 | K | | | |
| 136 | Vm | Lactate_transp. | 4.15 | A=Lactate$_{in}$, P=Lactate$_{out}$ |
| 137 | K | | | |
| 138 | Vm | Acetate_transp. | 4.15 | A=Acetate$_{in}$, P=Acetate$_{out}$ |
| 139 | K | | | |
| 140 | Vm | Ethanol_transp. | 4.15 | A=Ethanol$_{in}$, P=Ethanol$_{out}$ |
| 141 | K | | | |
| 142 | Vm | Formate_transp. | 4.15 | A=Formate$_{in}$, P=Formate$_{out}$ |
| 143 | K | | | |
| 144 | Km pyr | PDH | 4.17 | |
| 145 | Km nad | | | |
| 146 | Km coa | | | |
| 147 | Km nadh | | | |
| 148 | Km accoa | | | |
| 149 | Ki | | | |
| 150 | Vmax | | | |
| 151 | Vm | Xylose_transp. | 4.15 | A=Xylose, P=XLS |
| 152 | K | | | |
| 153 | Km nadh | XR | 4.8 | A=NADH, B=XLS, P=XLT, Q=NAD$^+$ |
| 154 | Km xylose | | | |
| 155 | Km nad | | | |
| 156 | Km xylitol | | | |
| 157 | Ki nadh | | | |
| 158 | Ki xylose | | | |
| 159 | Ki nad | | | |
| 160 | Ki xylitol | | | |
| 161 | Vmax(forw) | | | |
| 162 | Vmax(rev) | | | |
| 163 | Vm | Xylitol_transp. | 4.15 | A=XLT, P=Xylitol |
| 164 | K | | | |

## C.3  Power-law model

The evaluation of the reaction rates in the power-law model is an algorithmic process rather than a set of fixed rate equations. The evaluation is based on the stoichiometric matrix. The rules governing the rate evaluation process and parameter naming are listed below. Note that

the $H^+$ATPase is not included in the model.

- There are forward and reverse rate constants for each reaction.

- The number of order parameters is equal to the number of reactants in a reaction. In other words, the number of non-zero elements in the corresponding column of the stoichiometric matrix. PDH is the only exception which is modeled as irreversible and has only three order parameters.

- The total number of order parameters is 88. The total number of non-zero elements in the stoichiometric matrix is 94. Subtracting the excluded $H^+$ATPase (four elements) and the reverse term of PDH (two elements) results in 88 order parameters.

- The total number of parameters is twice the number of reactions plus the number of order parameters, that is, $2 \times 29 + 88 = 146$.

- The parameters are designated from $P_1$ to $P_{146}$. Parameters $P_1 \ldots P_{29}$ correspond to the forward rate constants of the reactions in the same order as the columns of the stoichiometric matrix are listed. Similarly, parameters $P_{30} \ldots P_{58}$ correspond to the reverse rate constants. Parameters $P_{59} \ldots P_{146}$ are order parameters. The order parameters within a reaction are ordered according to table C.3.

For example, the rate equation for PFK (reaction 6) is

$$v_{PFK} = P_6 [F6P]^{P_{74}} [ATP]^{P_{75}} - P_{35} [FBP]^{P_{76}} [ADP]^{P_{77}}, \tag{C.1}$$

where the brackets refer to the concentration of the reactant.

**Table C.3:** Order parameters of each reaction of the power-law model.

| Reaction | | Order of order parameters |
| --- | --- | --- |
| No. | Name | (metabolite numbers) |
| 1 | Glucose_perm. | 1,2 |
| 2 | PTS I | 11,34,12,37 |
| 3 | PTS II | 1,37,3,34 |
| 4 | GLK | 2,32,3,35 |
| 5 | PGI | 3,4 |
| 6 | PFK | 4,32,5,35 |
| 7 | FBA | 5,6,7 |
| 8 | TPI | 6,7 |
| 9 | GAPDH | 7,36,28,8,33 |
| 10 | PGK | 8,35,9,32 |
| 11 | PMG | 9,10 |
| 12 | ENO | 10,11 |
| 13 | PK | 11,35,12,32 |
| 14 | LDH | 12,33,13,36 |
| 15 | PFL | 12,38,17,15 |
| 16 | PTA | 17,28,21,38 |
| 17 | ACK | 21,35,22,32 |
| 18 | ACDH | 17,33,18,36,38 |
| 19 | ADHE | 18,33,19,36 |
| 20 | Phosphate_perm. | 29,28 |
| 21 | Lactate_tranps. | 13,14 |
| 22 | Acetate_tranps. | 22,23 |
| 23 | Ethanol_tranps. | 19,20 |
| 24 | Formate_tranps. | 15,16 |
| 25 | $H^+$ATPase | - |
| 26 | PDH | 12,36,38 |
| 27 | Xylose_tranps. | 25,24 |
| 28 | XR | 24,33,36,26 |
| 29 | Xylitol_tranps. | 26,27 |