Forschungszentrum Jülich

Institut für Biotechnologie

A Pathway Modeling Tool for Metabolic Engineering

Jochen Hurlebaus

Jül-3912

A Pathway Modeling Tool for Metabolic Engineering

Jochen Hurlebaus

Berichte des Forschungszentrums Jülich ; 3912
ISSN 0944-2952
Institut für Biotechnologie Jül-3912
D 5 (Diss., Bonn, Univ., 2001)

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek
52425 Jülich · Bundesrepublik Deutschland
☎ 02461/61-5220 · Telefax: 02461/61-6103 · e-mail: zb-publikation@fz-juelich.de

Abstract

Identification of metabolic regulation is a key point in metabolic engineering. Metabolic regulation phenomena depend on intracellular compounds such as enzymes, metabolites, nucleotides and cofactors. Quantitative knowledge about these compounds in combination with the known network of metabolic pathways allows the construction of mathematical models that describe the dynamic changes in metabolite concentrations over time. The models are high-dimensional systems of ordinary, non-linear differential equations. To solve the problems of this approach (setup of the equations that describe the metabolic pathways in form of kinetic rate equations and the parameter identification of the system parameters), a variety of pathway modeling software has been developed over the last years.

This work describes a Metabolic Modeling Tool (MMT) software for pathway analysis that is based on a relational database. It has been developed and applied to experimental data during this Ph.D. thesis. The software can be used to define and analyze dynamic pathway models based on non-linear kinetic rate equations. The design of the MMT software differs from other available pathway modeling tools in a variety of things. With respect to model definition, the tool has no limitations in dimension and it is based on a relational database that stores all model information and results. In order to allow definition of a large number of models the tool has a structured approach that allows to build groups of models and groups of parameter fittings. Groups share e.g. the metabolic pathway or kinetic equations.

With respect to simulation and parameter fitting, MMT provides high flexibility. Parameters can e.g. be optimized or set to experimental data values or initial conditions can be optimized. Stoichiometric coefficients can also be subject to parameter fitting. Additionally simulation time is variable and artificial noise can be added to the calculated values to simulate experimental data. The software creates a C-Code with the system of differential equations and also with the explicit derivatives with respect to initial conditions and parameters, thus providing a complete set of sensitivity functions.

The tool allows not only visualization of the solution of the differential equations but also of each individual flux between compounds. These values correspond to the kinetic equations over time and are helpful in analyzing simulation and parameter fitting results.

MMT was applied to the central metabolic pathways of *Escherichia coli*. Data from fast sampling experiments (Glucose pulse experiments, 4 samples per second over 40 seconds) was used for parameter fitting. Some metabolic regulation phenomena in relation to these experiments are discussed in this thesis.

This thesis is based on research at the Unit of Theoretical Biology of the Mathematics and Natural Sciences Department at the Rheinisch Friedrich-Wilhelms-University Bonn, Germany. The work was conducted at the facilities of the Institute of Biotechnology 2 of the Research Center Jülich, Germany.

Special thanks to

- Prof. W. Alt, University of Bonn, for supervising this work.
- Prof. W. Wiechert, University of Siegen, for his extensive help during this work and his critical input in the process of writing this thesis.
- Prof. C. Wandrey, University of Bonn, for providing the opportunity to work on this challenging project.
- Dr. Ing. R. Takors in his position as the leader of the Fermentation Group at the Institute of Biotechnology where this work was conducted.
- Dr. U. Schäfer and Dipl. Chem. A. Buchholz for their extensive work in the field of analytics during their Ph.D. thesis that provided the experimental data for this work.
- Dr. K. Schmidt for discussions about this work.
- W. Boos, A. Franz, H. Kiehl for their computer system administration and help with programming problems.
- The people at the Institutes of Biotechnology 1 and 2, especially the people of the Fermentation Group, that helped in many ways for the success of this work.
- My family and Heike Guist for their understanding and support.

Contents

1. Introduction

Imagine a civilization that uses cars as a general mean of transport but never figured out why cars drive. At some point, an increasing amount of people starts to rip old cars apart, take a look at single parts, e.g. rolling tires down a hill trying to approximate how fast a car can possibly run. Other scientists look at running cars, estimating how much gas and other resources are entering the car, and into what substances the car transforms these resources. Through the years, the civilization has found all relevant parts needed to build cars and even possesses the tools to put the parts together. Also, conditions could be found that are good for certain cars and not so good for others. Now some people are very enthusiastic and claim that they are able to improve existing cars to make them perfect or even build new superior cars from scratch - all that with their huge amount of knowledge they gained by looking at all the single parts of cars, their resource consumption and seeing them drive. Would you believe these people? Would you give them a chance to make a specific improvement on a car?

Looking at the history of microbiology, a very similar situation to the above story exists. For a long time cell behavior has been examined, growth conditions have been optimized. Over the years, cells have been taken apart and substances as e.g. enzymes have been analyzed in detail. Then techniques for DNA recombination became available. Almost 20 years ago, in the year of 1981, some people started to talk about directed pathway modifications in cells, having in mind better cells e.g. for the production of amino acids [53]. This process nowadays is often referred to as metabolic engineering, a term coined by Bailey in 1991 [3]. A widely used definition of metabolic engineering is the one of Stephanopoulus et al in [96]: *Metabolic engineering means the directed improvement of product formation or cellular properties through the modification of specific biochemical reaction(s) or introduction of new one(s) with the use of recombinant DNA technology.* Nowadays, a variety of additional tools is available for creation of new strains for metabolic engineering.

The chances that our civilization mentioned above will improve any car are rather low. Fortunately some people in this civilization started a new approach, trying to analyze a combination of parts of the car like e.g. a complete engine. Even better, they developed methods to examine engines and other parts while the car is driving!

Therefore interactions between single parts and the importance of single parts for the functioning of the car could be analyzed. Single parts could be analyzed in a working environment which resulted in completely different results than considering the parts in an isolated environment in the garage. Using these techniques it was e.g. possible to give hints what limits the speed or the resource consumption of the available cars. Now the part specialists could use their knowledge to replace or improve parts that limited overall car performance.

In the field of metabolic engineering, new measurement techniques focus on intracellular, *in vivo* cell behavior - trying to figure out what controls and limits intracellular behavior. Now the intracellular evolution of living cells is analyzed under various conditions ranging from steady drive at high or low speed (steady state analysis) to dynamic drive under rapidly changing conditions (e.g. pulse experiments, severely limited cells, temperature shift). Because an increasing number of intracellular and extracellular substances can be quantified with increasing accuracy during such processes, the dream metabolic engineering has become closer.

1.1. Targets and methods of metabolic engineering

Metabolic engineering is of interest for industrial purposes because it increases efficiency of production processes [96]. The main purpose here is to decrease the amount of used substrates, cofactors, and energy, to decrease the amount of waste (by-products), and to increase the amount and quality of desired product, thus increasing the product yield (see Figure 1.1).

On the other hand, the way to achieve the goals of metabolic engineering are of great interest from a scientific point of view because they give insight in the regulation and control of cellular evolution. Successful metabolic engineering requires a thorough understanding of microbiology of cells starting from the genetic information of the DNA all the way up to large-scale processes of cell populations [4]. Figure 1.2 shows the main functional units of microorganisms. Lately developed and improved methods provide the basis of a global understanding of cellular life. The following methods are extensively used now in order to fill the still remaining gaps in the picture of a functioning cell:

Genomics: The genetic information of many organisms is already available. Analysis is automated (robots). The data, however, only describes the genetic background of a cell, not what genes may be expressed and activated. The genetic background is fixed during the life time of cells and is strain specific.

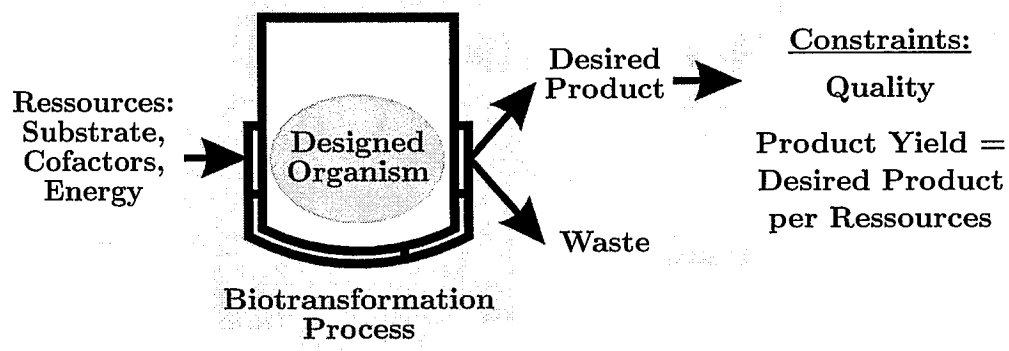


Figure 1.1.: Industrial process and interests.

It is, however, altered through spontaneous mutations [40].

Transcriptomics: Measurement of mRNA with DNA-arrays gives information about expressed genes in a specific physiological situation. Therefore transcriptomics is an important tool for measuring global responses of a cell to changes in environmental conditions. The time scale of such responses is in the range of several minutes to hours.

Proteomics: Gel-electrophoresis allows estimation of enzyme concentrations in cells. Next to transcriptomics, it provides the second important source of data to distinguish between cells growing under different conditions. With proteomics data, the amount of enzymes in a living cell can be analyzed. However, the accuracy is low making the method especially useful for a qualitative comparison of different cells. The exact amount of enzymes plays an important role for the dynamic system behavior. Production of a single enzyme is a process that takes at least 1 to 2 minutes, sometimes significantly longer, therefore measurements during a short time frame are not useful.

Metabolomics: Analysis of intracellular and extracellular metabolite concentrations is termed metabolomics. It provides data for estimation of intracellular dynamics in a time frame of seconds or even milliseconds. Because metabolite concentrations are thought to be the main regulators of intracellular dynamics besides enzyme concentrations, but can change much faster, they are necessary input for simulations of e.g. a response to a substrate pulse [79].

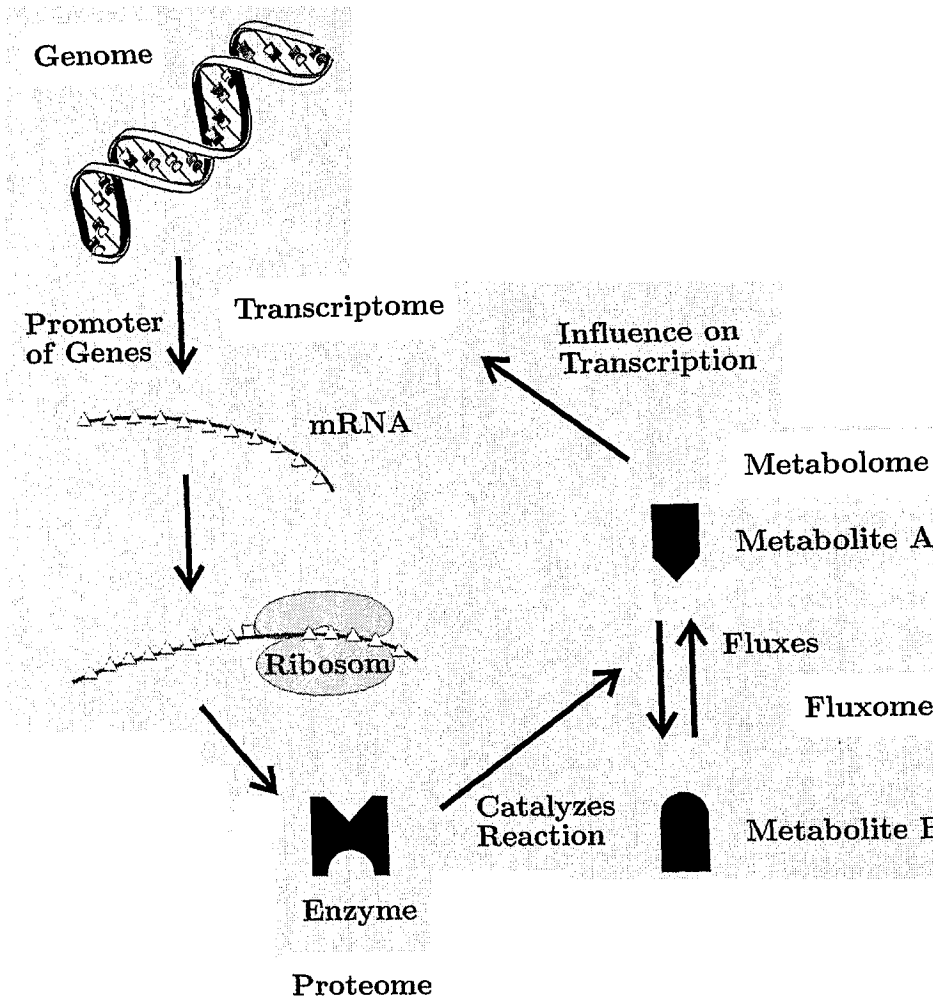


Figure 1.2.: Genome, transcriptome, proteome, fluxome and metabolome: five central parts determining fate of cells.

Fluxeomics: Intracellular fluxes and fluxes through the cell membrane can be calculated from measurements with labeled substrates as e.g. ^{13}C [110], [111]. Such measurements give important information about material fluxes inside cells as well as uptake and excretion fluxes. They are important also for a dynamic model because measurement of metabolite concentrations alone does not give information about fluxes.

Table 1.1.: Methods for detailed *in vivo* analysis of microorganisms.

Method	Measurement tools	Established, ease of use
Genomics	DNA-Sequencing	Well established, commercialized, automated, complete sequencing of an organism is still time intensive.
Transcriptomics	mRNA concentrations using DNA array chip technology	Development phase, chip production is time intensive, analysis difficult due to the lack of experience. Quantifications difficult but comparison between cells or strains is possible.
Proteomics	Protein concentrations with 2-D, 3-D gel electrophoresis, MALDI-TOF	Still in development, quantification and interpretation difficult, interpretation time intensive. Mainly useful for comparison between cells or strains.
Fluxeomics	Relative fluxes with isotopomer analysis, NMR	Established, commercialized, depending on the used substrate the measurements are of high accuracy, limited to certain pathways.
Metabolomics	Metabolite concentrations with enzymatic tests, HPLC, LC-MS, GC-MS	Enzymatic tests, HPLC established, LC-MS, GC-MS in development. All measurements are time intensive. Accuracy can be influenced by other substrates that behave similar then the ones to be analyzed.

Table 1.1 gives an overview over the measurement methods and the current state of the art. Data gained from classical methods considering *in vitro* processes is also useful, but usually their applicability is limited to a qualitative interpretation. Rate equations for enzymatic reactions can be derived from *in vitro* experiments because these equations depend on the structure of the enzyme. Additionally, thermodynamic constants of chemical reactions can be measured *in vitro* in order to gain information about the preferred direction of the reactions. However, kinetic constants of the rate equations (e.g. a maximum reaction rate v_{max}) that are measured *in vitro* should be handled with care. This results from the fact that the differences

between *in vitro* and *in vivo* conditions have been proven too large in general in order to use quantitative information as e.g. kinetic constants from *in vitro* experiments for *in vivo* models [76], [16].

The amount of data gained with the used methods and the complexity of the cellular system make it impossible for a human mind to gain a complete understanding of the functioning of a cell. Figure 1.3 pictures a common problem in computational biology: whereas the amount of available data exploded due to the availability of machinery for generation of accurate data, tools for efficient data analysis have not been developed with the same effort. Increasingly detailed mathematical models describing cellular systems are thought to give hints about intracellular control mechanisms. Due to the complexity of the system, however, a mathematical model for a cellular system probably will in the near future not be better than a five day weather forecast. But if a certain behavior can be predicted with a sufficiently high probability, metabolic engineering will probably achieve its goals faster.

1.2. Scope of this work

The conceptual approach of the pathway modeling project at the Institute of Biotechnology is based on an interaction between modeling and experimental work. The approach of the four basic tasks experiment, analytics, modeling, and experimental design is shown in Figure 1.4. The process was initially started with experiments. This work provides a modeling framework for analysis of such data. Results of models should then be used to design new experiments. This will increase the amount of knowledge that can be gained with a new experiment.

The goal of this dissertation was the identification of pathway models for the main metabolic pathways of *Escherichia coli.* The dissertation focuses on mathematical models based on data from metabolomics. Modeling was guided by experimental data gained in two pulse experiments with *Escherichia coli* K 12. For the time span of the experiments of 45 seconds, models that focus on metabolic regulation have been developed.

The main reactions of the considered *Escherichia coli* pathway can be found at the end of this work in Figure B.1 and on the last page that can be folded out in order to have the pathway in mind while reading this work. Abbreviations used in the pathway maps and throughout this work are summarized in Appendix A.

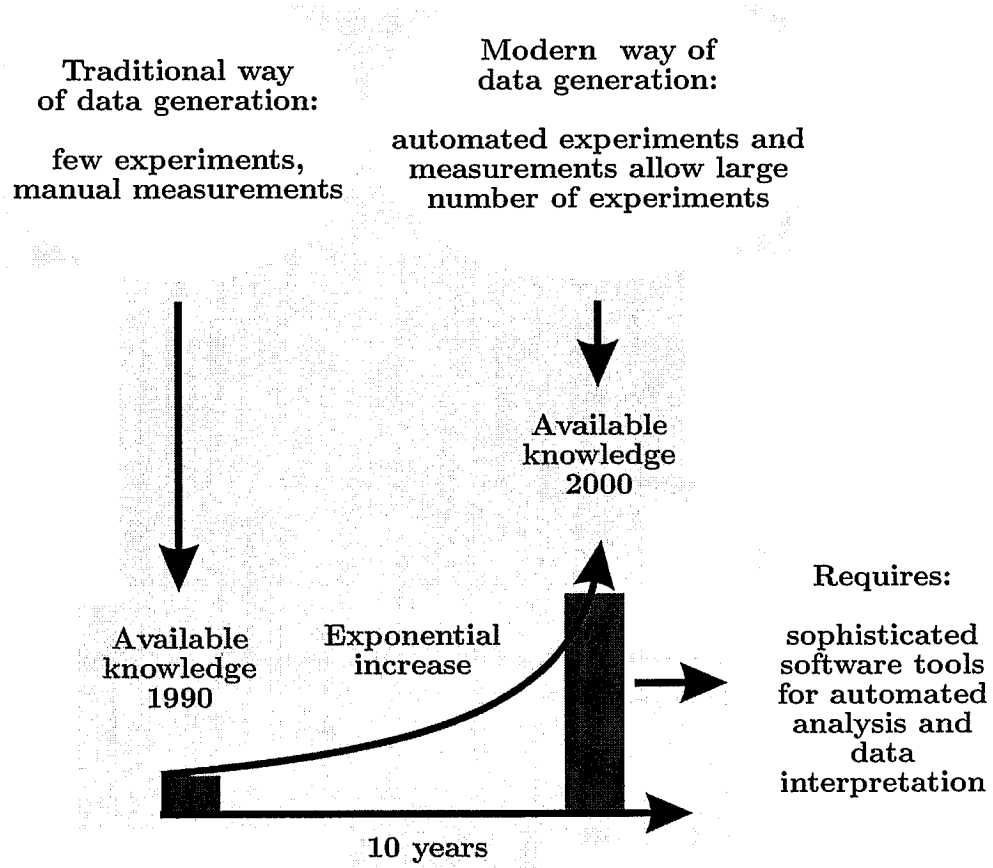


Figure 1.3.: The methods for data generation and collection have changed in recent years, whereas the methods for data analysis could not keep up and are just starting to catch up.

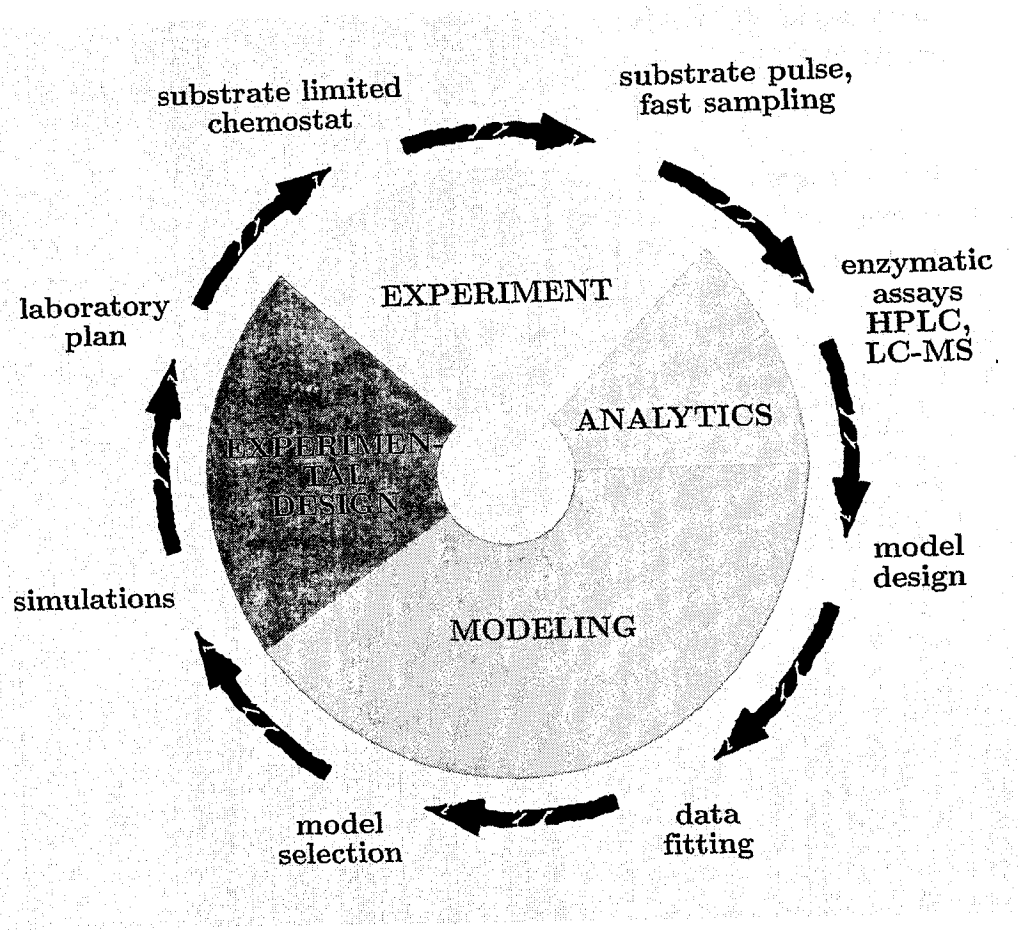


Figure 1.4.: Conceptual approach with experiments, analytics, modeling and experimental design.

In short, the modeling work that is part of this thesis is based on the following main features:

- Dynamic models are used in order to model the dynamic concentration changes of intracellular substances after a substrate pulse.
- Enzyme kinetics are used as a basis for each chemical reaction of the pathway in order to build a mechanistic model of the metabolic network.
- Intracellular metabolite concentration measurements are used for fitting models to data.
- A Metabolic Modeling Tool Software (MMT) is developed that allows model

buildup, simulations and parameter fitting of a large number of alternative models in a short period of time.

- The MMT Software is used to identify a model that is able to describe some dynamic phenomena after a substrate pulse that could not be explained without the help of the model.

Details about underlying experimental data are to be found in Chapter 2. Chapter 3 explains the concept of mathematical models and metabolic engineering and compares this work with other approaches. The developed software is described in Chapter 4, the mathematical solution and parameter fitting methods are part of Chapter 5. First results with so far available experimental data are part of Chapter 6, and a discussion and outlook are presented in Chapter 7.

2. The Model Organism *Escherichia coli*

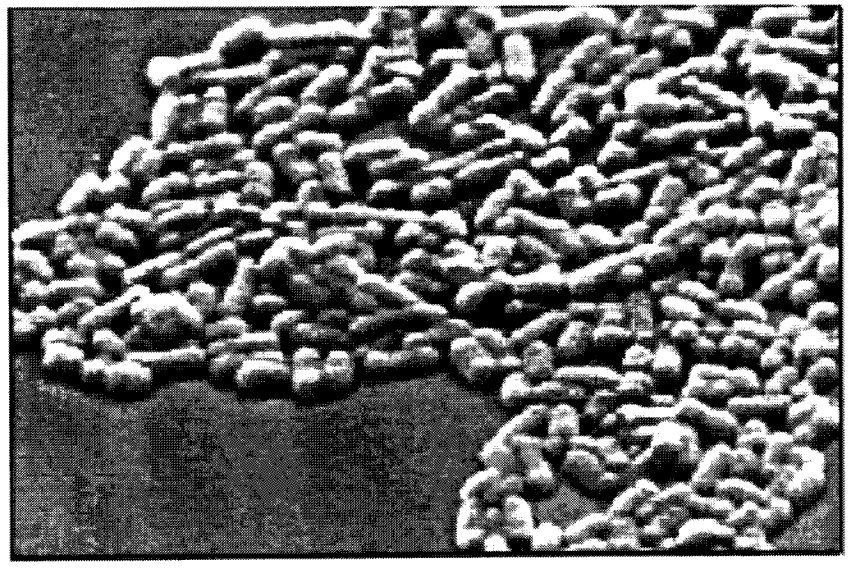


Figure 2.1.: A colony of *Escherichia coli* cells.

Escherichia coli is a well known organism used frequently in industry and research. Figure 2.1 shows a colony of cells. One of the main research focuses at the Institute of Biotechnology at the Research Center Jülich are strains of *Escherichia coli*. This specialization is a results from the fact that *Escherichia coli* cells

- are successfully used in production processes in industry, e.g. for insulin, penicillin acylase [60], tryptophan, indigo, or ethanol (see [65] chapter 108 for an overview).
- have been analyzed and used by many research groups around the world for many decades, making *Escherichia coli* the microorganism with the largest available knowledge. A whole collection of databases can be found on the Internet [28]. Some famous databases also used extensively during this work are [51], [52], [113], [11], [1].

- have a genome that has been sequenced completely [9].
- are considered to be rather simple organisms due to their prokaryotic structure.
- are easy to handle in the laboratory and do not require high safety standards because they are a widespread organism.

Table 1.1 in Chapter 1 summarized the available methods for *in vivo* analysis of microorganisms. Table 2.1 shows the methods and experiments performed at the Institutes of Biotechnology at the Research Center Jülich using strains of *Escherichia coli*. Because many established methods are applied in one research facility for a specific organism, one can hope for quick improvements in metabolic engineering at the Research Center Jülich.

Table 2.1.: Methods and research focus at the Institute of Biotechnology at the Research Center Jülich.

Method	Research at the Institute of Biotechnology
Genomics	Construction of new strains with focus on aromatic acid synthesis pathways
Transcriptomics	DNA Chip Technology to analyze mRNA concentrations under varying growth conditions
Proteomics	2-D gel electrophoresis, using MALDI-TOF for analysis of enzyme concentrations
Fluxeomics	Labeling experiments with NMR analysis to quantify intracellular fluxes, mainly with ^{13}C
Metabolomics	Quantification of intracellular metabolite concentrations with enzymatic tests, HPLC, LC-MS. Steady state and pulse experiments in laboratory scale.
Fermentation	Development of fermentation tools and technologies for defined growth conditions suitable for comparison of strains. Scale-ups.

Based on the experimental data from the applied methods, mathematical analysis is performed with increasing effort. This dissertation, as one of the starting points of extensive mathematical modeling at the Research Center deals with the modeling of data gained from metabolomics only. Therefore, in the following work, this part of the current research will be discussed extensively whereas details from other fields will be omitted. However, modeling of the metabolomics data was used as a sample process and will be extended to more complex models in the future.

2.1. Metabolomics at the Institute of Biotechnology

In order to measure dynamics of intracellular metabolite concentrations a fast sampling device machinery has been developed at the Institute of Biotechnology. The goal of the approach is to develop models describing intracellular dynamics as a response to fast changes in the substrate concentration. For this approach, a fast sampling device is required to provide sufficient data in the first seconds after such a change in the substrate concentration.

The fast sampling technique has some apparative constraints as mixing time of the pulsed substrate, accuracy of the pulse, and accuracy of the sampling volume. Extensive work was necessary for development of a machinery with sufficient sampling frequency and good accuracy in sampling volume [90]. The sampling time finally could be reduced to 220 ms. See Figure 2.2 and 2.3 for a picture of the device. The experimental schedule for data generation is

1. Setup a substrate limited chemostat for 7.5 residence times to achieve a stable steady state.

2. Start the fast sampling device about five seconds before a pulse of substrate is given. With a sampling time of approximately 220 ms, about 20 samples from the steady state are taken.

3. Perform a substrate pulse by injecting substrate with pressure in the fermenter. Pressure is needed in order to minimize mixing time.

4. Continue sampling for about 35 seconds which results in additional 140 to 150 samples of cells after the substrate pulse.

5. Perform analysis of extra- and intracellular compounds with enzymatic assay, HPLC, and LC-MS. Due to the large number of samples and the time consuming methods, this analysis takes about three months of time.

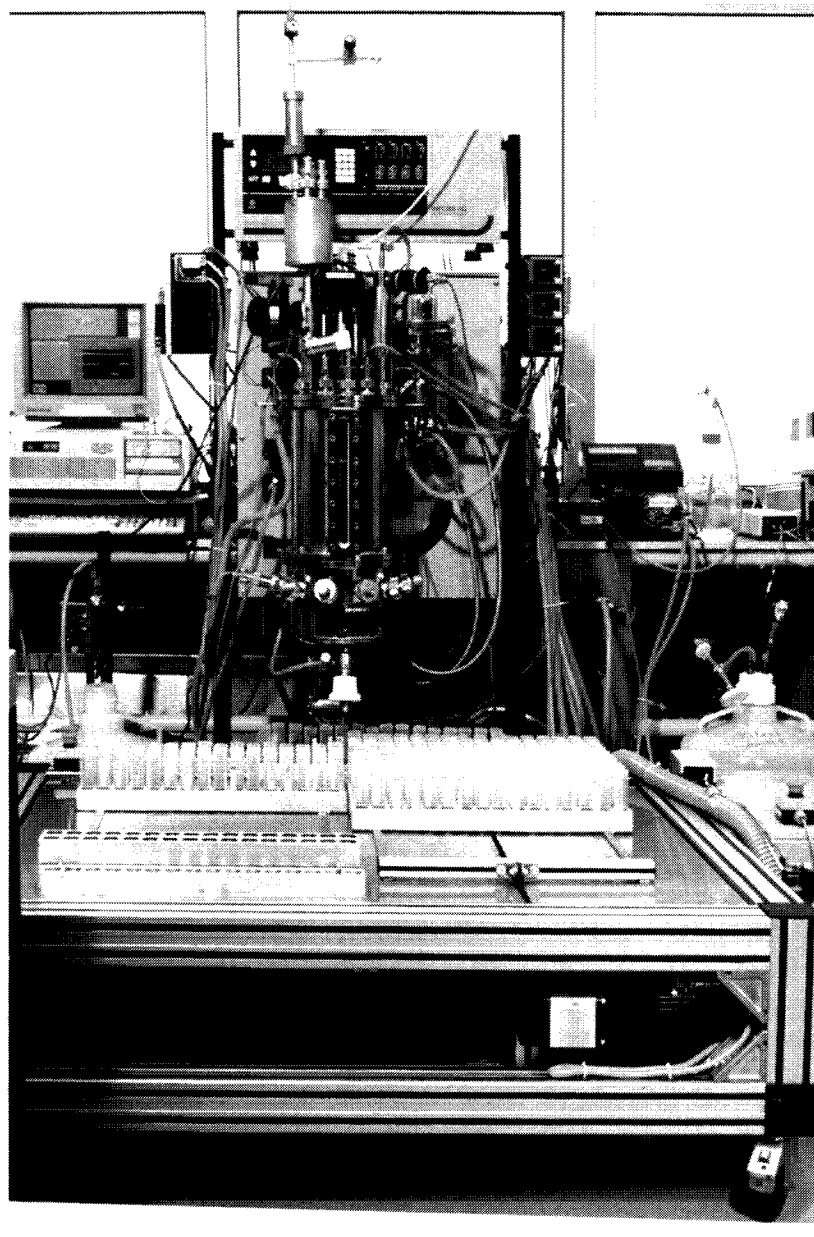


Figure 2.2.: The experimental setup of the fermenter and fast sampling device in the laboratory at the Institute of Biotechnology.

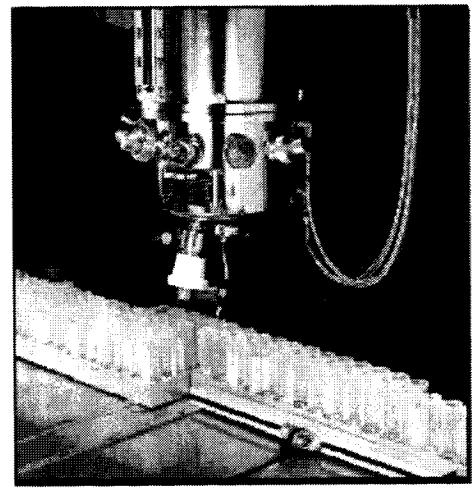


Figure 2.3.: The sampling valve at the fermenter for the fast sampling and the sampling tubes passing below on a conveyor belt.

Table 2.2 summarizes the experiments carried out so far using the fast sampling experimental setup. The table also shows the number of intracellular compounds that have been analyzed.

Table 2.2.: Experiments carried out with the fast sampling device.

No	Limiting Substrate in Chemostat	Pulsed Substrate	Analysis Methods	Number of Compounds analyzed
1	glucose	glucose	enzymatic, HPLC	15
2	glucose	glucose	enzymatic, HPLC	15
3	glycerol	glucose	enzymatic, HPLC	20
4	glycerol	glycerol	enzymatic, HPLC	25
5	glycerol	glycerol	enzymatic, HPLC	25
6	glucose	glucose	enzymatic, HPLC, LC-MS	31

Table 2.3 summarizes intracellular compounds that can be analyzed by various methods and their corresponding limit of detection. The table shows that a mix of available methods is important to maximize the number of metabolites that can be analyzed with good accuracy. See Figure B.1 for location of the measured compounds in the metabolic pathway. For a complete description of the experimental procedure and analysis of compounds with enzymatic assay, HPLC, and LC-MS, see [90], [13], [14].

Experimental data gained with the methods mentioned in this section provides the main input for the mathematical models developed in this thesis. The models presented in Chapter 6 where fitted to data from the glucose-glucose experiment. In these experiments, a sudden change in the substrate concentration from 0.1 to 15 mM GLC was used during the experiments. It is believed that in order to optimize growth conditions, estimation of intracellular reactions under changing conditions is necessary, and this can be simulated with a substrate pulse. Large changes from limited growth conditions to saturated growth conditions are likely to show limitations in the metabolic pathways as well as maximum reaction rates. Additionally, large changes in the substrate concentrations cause large changes in many intracellular metabolites. This makes it easy to distinguish between error of measurements and real concentration change during an experiment.

Table 2.3.: All metabolites quantified in cell extracts of *Escherichia coli* K12 under glucose (GLC) limited conditions at cell densities of 10 g l^{-1}. Intracellular limits of detection in [mM]. PP-Pathway: Pentose-Phosphate-Pathway. TCA-Cycle: Citric-Acid-Cycle. See tables A.2 and A.3 for abbreviations of compounds.

Compound Name	Enzymatic Assay [mM]	HPLC (UV) [mM]	LC-MS [mM]
glycolysis:			
G6P	0.23		0.13
F6P	0.30		0.16
FBP	0.30		0.09
GAP	0.53		
DHAP	0.53		
3PG	0.30		
2PG	0.38		
PEP	0.38		
PYR	0.38		
TCA-cycle:			
CIT	0.15		
ISC	0.17		
FUM			
OXO	0.18		
SUC	0.19		
MAL	0.06		
OAA	0.68		
ACoA	0.30		0.13
glycerol metabolism:			
GLY	0.27		
G3P	0.40		
nucleotides:			
cAMP		0.63	
AMP			0.38
ADP		0.63	0.24

continued on next page

Table 2.3.: (continued)

Compound Name	Enzymatic Assay [mM]	HPLC (UV) [mM]	LC-MS [mM]
ATP		0.63	0.08
NAD		0.63	0.15
NADP		0.63	0.08
GDP		0.63	0.14
GTP		0.63	0.10
FAD			0.02
PP-Pathway:			
6PG			0.38
R5P			0.22
substrate:			
GLC	0.30		
GLY	0.27		

2.2. Biological data interpretation

This section aims at an qualitative understanding of the data used for modeling. The general problem of interpreting phenomena seen in the data is discussed. Ideas presented here are based on common biochemical knowledge and theories. In Chapter 6 mathematical models are presented in order to gain more insight in the phenomena of the data that are discussed here.

Figure 2.4 shows the substances analyzed in one of the experiments. In the experiment a GLC limiting chemostat was combined with a substrate pulse according to the experimental setup described in Section 2.1. Figures 2.5 shows the intracellular concentrations over time of the metabolites involved in the phosphotransferase system (PT-System) of *Escherichia coli* for uptake of GLC. The figures include measured values and error bars. Error estimates include the whole process from sampling to enzymatic assay respectively HPLC. Concentrations of other metabolites are not discussed here but can be found in Figure 6.15.

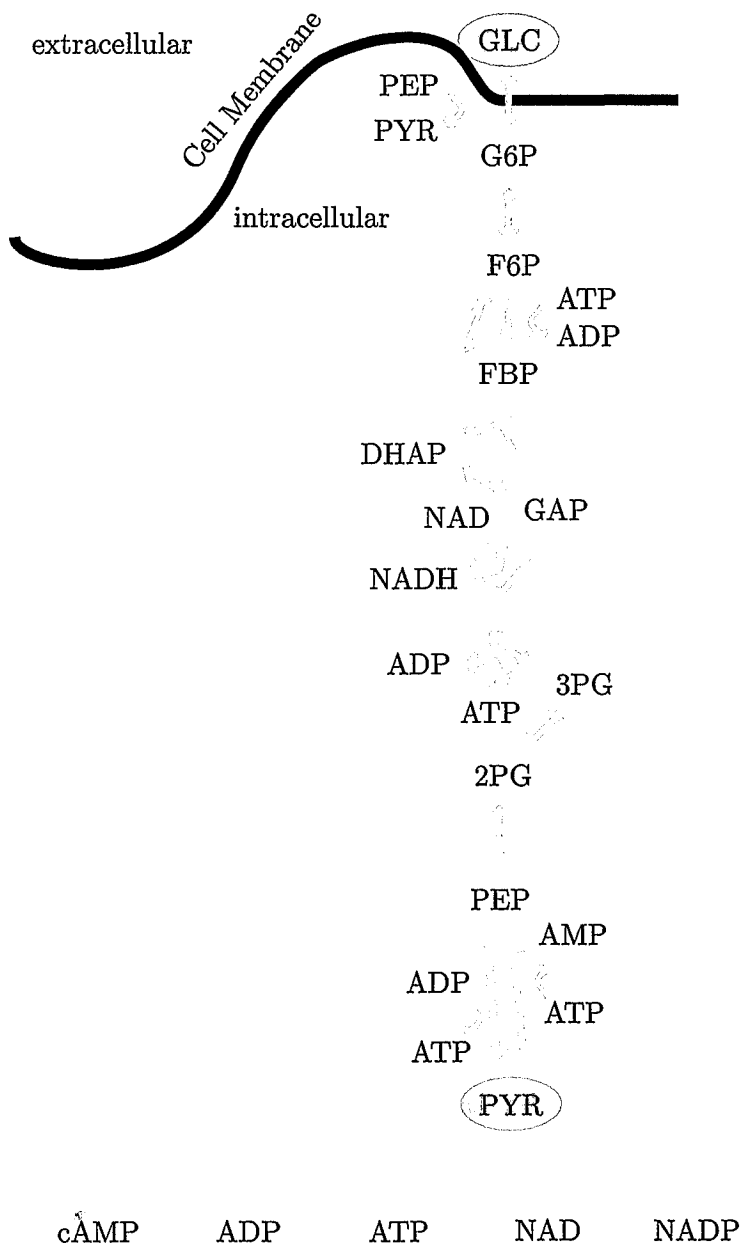


Figure 2.4.: Pathway of *Escherichia coli*. Shown are only compounds measured in the glucose-glucose pulse experiment discussed in Section 2.2.

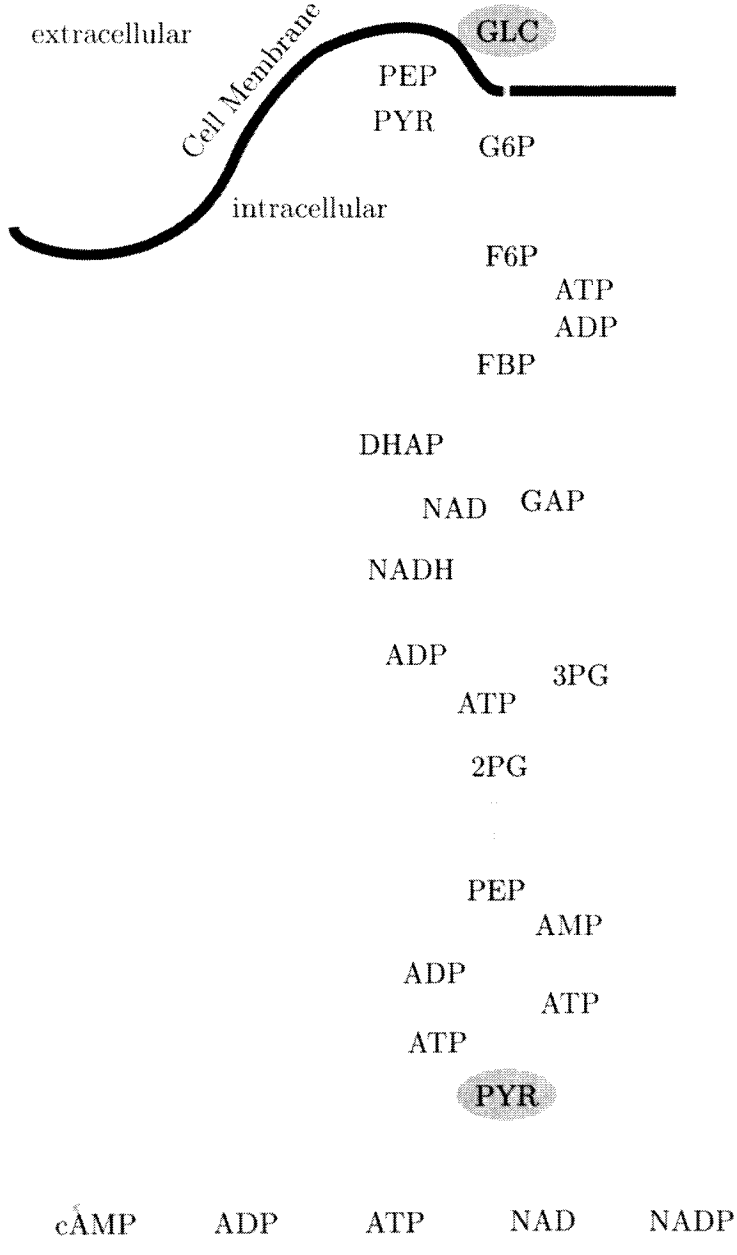


Figure 2.4.: Pathway of *Escherichia coli*. Shown are only compounds measured in the glucose-glucose pulse experiment discussed in Section 2.2.

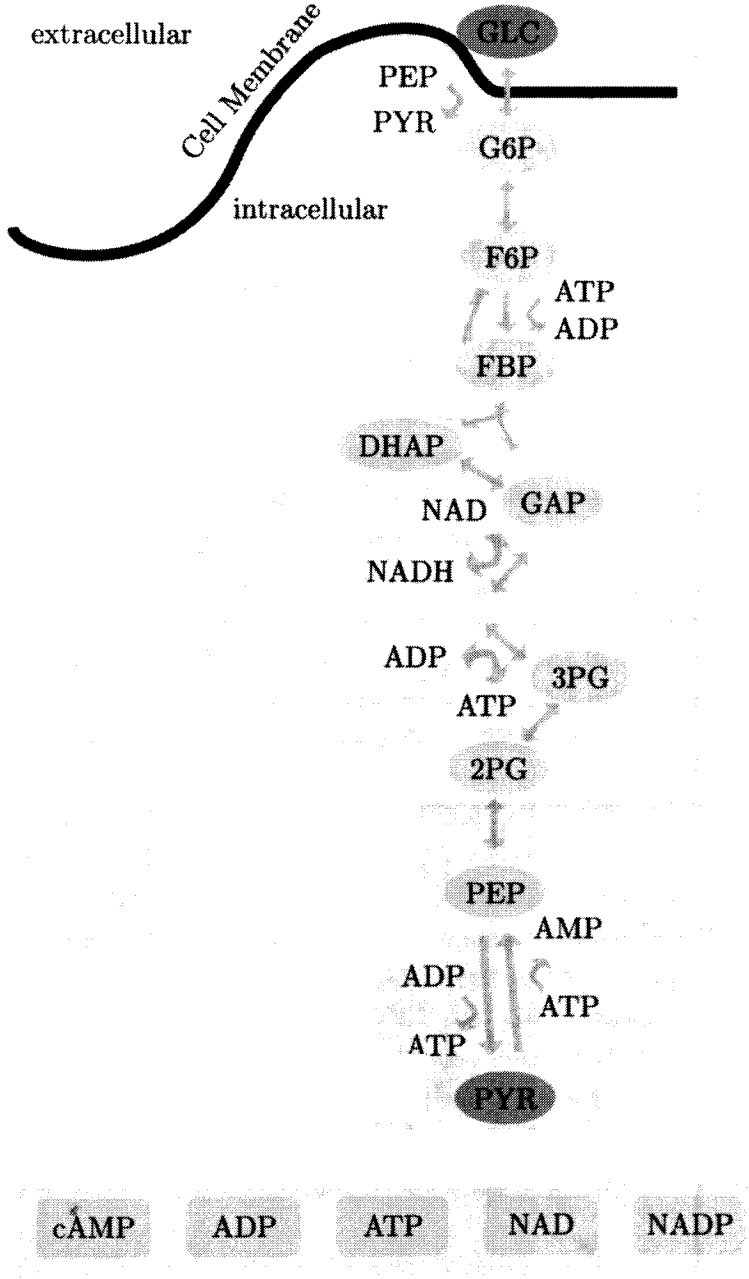


Figure 2.4.: Pathway of *Escherichia coli*. Shown are only compounds measured in the glucose-glucose pulse experiment discussed in Section 2.2.

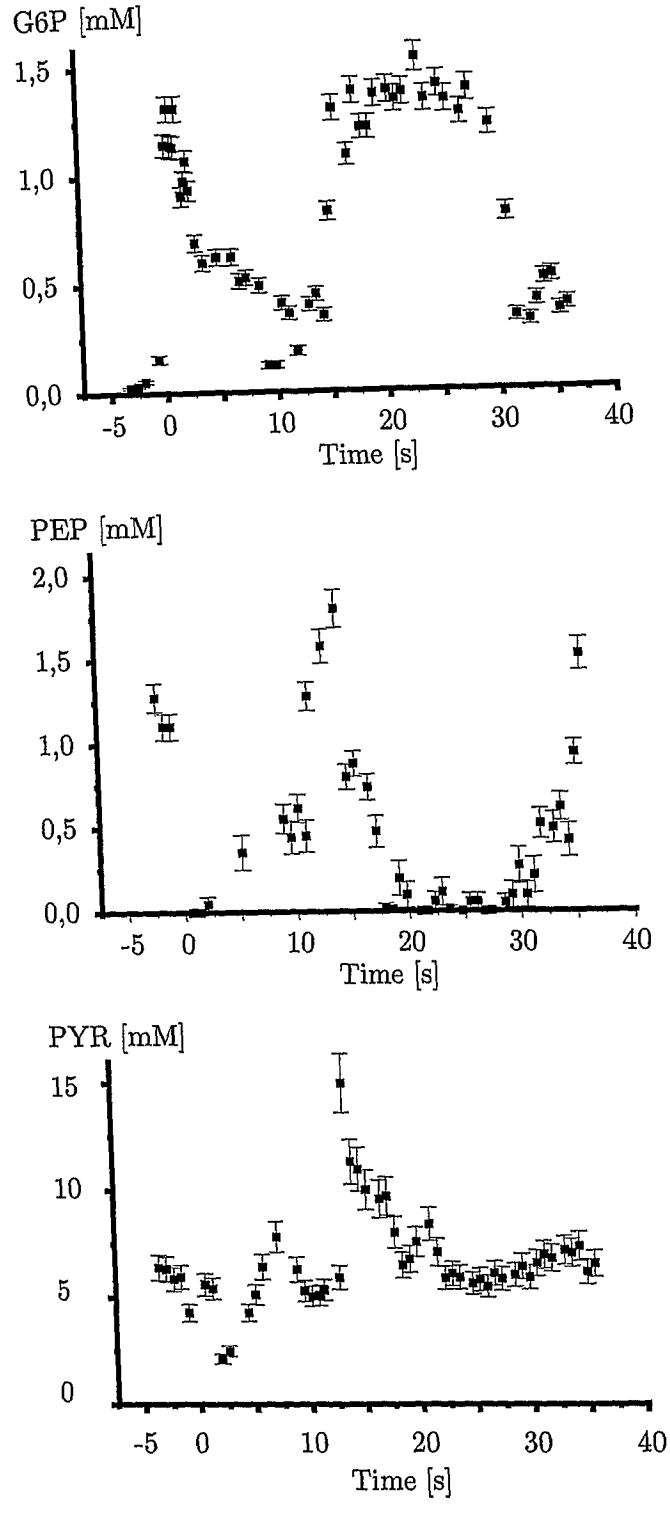


Figure 2.5.: Intracellular G6P, PEP and PYR concentrations in [mM] before and after the GLC pulse. The pulse was given at time $t = 0s$. See text for comments on error bars.

2.2.1. Dynamics of G6P, PEP, and PYR

During growth on GLC, the main substrate uptake of *Escherichia coli* is performed with the PT-System. This uptake system consists of three phosphotransfer reactions with a complicated mechanism [60], [73]. Figure 2.6 describes the main mechanism with the conversion of one molecule of PEP to PYR for each GLC molecule taken up.

Because only GLC was available as a substrate for the cells in the discussed experiment, it can be assumed that all substrate is taken up by the cell through the PT-System and converted to G6P. Under balanced growth on GLC, about 17% of the flux from the total flux exiting the G6P pool enters the PP-Pathway, whereas 80% enter the glycolytic pathway [60]. These rates are estimated from cell composition, substrate consumption, and growth rate of cells. Under such conditions, another 13% of the initial uptake flux exits the glycolytic pathway to form certain biomass components. Still about 70% of the flux end in PEP, the metabolite required for GLC uptake [60]. Because each G6P is split into two PEP molecules (see pathway map, reaction from FBP to DHAP and GAP), under normal growth conditions, PEP is not a limiting metabolite. As will be seen now, the situation changes in the experimental data, where PEP seems to be limiting for GLC uptake and therefore control dynamic system behavior. This question will also be discussed further in Chapter 6 where models for the PT-System and glycolysis are described.

The hypothesis of sufficiently high intracellular PEP concentration is in agreement with the experimentally determined PEP concentration that can be seen in Figure 2.5 for the time -5 to 0 seconds. The cell seems to have sufficient PEP available under GLC limited conditions. Things change dramatically after a sudden pulse of GLC. Because the PT-System is considered to have a large influence on the dynamic response to a substrate pulse, PEP and G6P concentrations as well as PYR, which is produced from PEP, will be further discussed in this section. The intracellular metabolite concentrations from these data sets are also used in Chapter 6 to model the observed short term dynamics. The following section gives some data interpretation from a biochemical viewpoint without use of mathematical analysis tools.

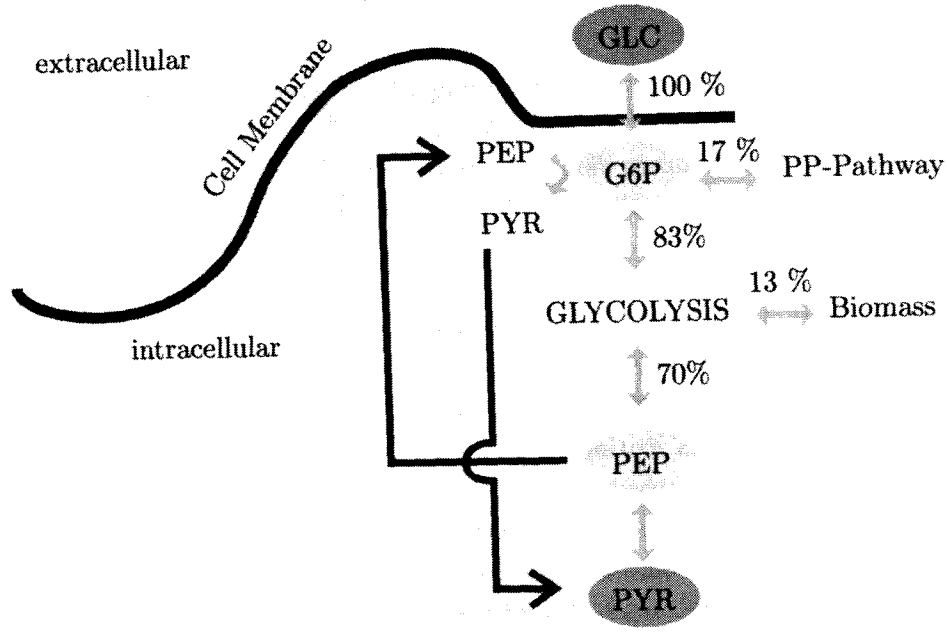


Figure 2.6.: The phosphotransferase system of *Escherichia coli* for uptake of carbosugars. It requires a PEP molecule for each GLC molecule taken up. Shown is also a typical flux distribution for growth on GLC.

2.2.2. Does the PT-System of *Escherichia coli* control intracellular dynamics?

A look at Figure 2.5 for different time periods reveals that some phenomena of the measurements can be explained with the PT-System, whereas others remain questionable.

-5 to 0 seconds: GLC limited chemostat: the G6P concentration remains below detection limit, PEP concentration remains at a relatively high level. Other metabolites show no unusual behavior.

Time 0: The GLC pulse is given in the fermenter, changing the GLC concentration quickly to a value above saturation (15 [mM]). During the remaining time, the GLC concentration remains well above saturation.

0 to 1.5 seconds: An immediate response to the GLC pulse can be seen. The G6P concentration increases, whereas the PEP is used for GLC uptake. An increase of G6P of about 1.2 mmol coincides with a decrease of PEP of about the same amount.

1.5 to 10 seconds: PEP limitation? During this time, the PEP concentration remains below detection limit. At the same time, G6P concentration decreases from its high value present right after the pulse. It is important to keep in mind that a PEP concentration below detection limit does not imply a zero flux to PEP or zero GLC uptake. Instead, it is likely that all new PEP molecules produced are immediately used for the uptake of GLC. Besides, the sum of the fluxes from G6P to F6P and into the PP-Pathway seem to be larger than the GLC uptake, following a negative net-flux for G6P which results in a decrease of G6P.

10 to 15 seconds: During this time a sudden increase in the concentration of PEP can be seen to values higher than before the pulse. With a short delay, the G6P concentration also increases. The reason for the sudden PEP increase is not clear, however, the following increase in G6P seems to have its reason in the availability of PEP for uptake of GLC. It is also worthwhile noticing that the PYR concentration follows the PEP concentration. Because PYR is produced from PEP, an increase in PEP could mean a decreased flux to PYR and should therefore result in a decrease of PYR. Obviously, some other flux distribution seems to be present.

15-25 seconds: A low PEP concentration coincides with a high G6P concentration. Approximately 15 seconds after the GLC pulse a situation occurs that could not be expected a priori. Apparently all PEP is used for additional uptake of GLC, triggering the increase in G6P until 15 seconds after the pulse. Now, however, G6P does not decrease after PEP is used up in contrast to the situation at time 1.5 to 10 seconds. It seems likely that other processes control the fluxes in this time span. A closer look at intermediate metabolite concentrations and their direct influence does not give immediate insight. The effects will be discussed in more detail together with the mathematical models for glycolysis presented in Chapter 6.

25-35 seconds: Repeating situation from 1.5 to 15 seconds. In this time frame the cell behaves like in the time frame 1.5 to 15 seconds. G6P decreases, and shortly thereafter there seems to start another peak in the PEP concentration. Because measurements ended at 35 seconds, it could not be observed how the dynamics continue. It is possible that the system evolves to a steady state or continues to oscillate. Further investigation with extended experiments is helpful to answer these questions.

Other Metabolites:

It is not surprising that considering just two metabolites out of many can not explain the above behavior for the complete time of the experiment. Looking at the

other compounds, however, provides not enough additional insights and does not give any hints what could happen in the time span 15-25 seconds. The example shows the difficulties occurring to interpret a dynamic data set of a highly connected, non-linear dynamic system. The following chapters will show that various mathematical methods used nowadays in order to gain more insight in the system behavior are also difficult to apply and require a large amount of biochemical knowledge.

2.3. Aims of the models for *Escherichia coli*

The background of this work has been presented in this chapter. Additionally, experimental data is shown and phenomena are discussed. The aim of this work is to provide a framework for modeling of microorganisms based on a software tool. The models build with this tool should help to analyze

- intracellular dynamic changes after a substrate pulse as described above
- control mechanisms and limiting steps in complex reaction systems
- importance of futile cycling and anaplerotic reactions
- fluxes in a dynamical system calculated with enzymatic rate equations

Answers to these questions require additional input from flux analysis and steady state experiments. Also, these questions cannot be answered completely because estimation of fluxes in a dynamical system are difficult. However, the chosen concept is thought to be useful for answering some of the above questions.

3. The Modeling Concept

3.1. Modeling approaches

Developing mathematical models in order to describe a system is a common scientific technique. In many areas, e.g. electrical engineering, detailed models based on well established physical laws can be developed. Such models then can be used to predict system behavior a priori, before the systems are actually setup. Situations also occur where a detailed model describes an existing system but not all the parameters are known. This means that all parts of the model are known and can be put in a mathematical form. Then, even if some parameters are not known a priori, the model can be fitted to a set of experimental data - assuming that the quality of the experimental data is good enough.

These situations are almost ideal modeling cases. Models can be used for exact representation of systems, without neglecting significant parts of the system or introducing strong assumptions. The requirements to build such models are a complete qualitative and quantitative knowledge of the system to be simulated. In the biological sciences, this ideal modeling case is far from reality at this time. Many simplifications and assumptions are introduced for all models in this field. This is a result of the complexity of biological systems, the difficulties to gain sufficient and exact data from experiments, and the lack of qualitative and quantitative understanding of the system.

In particular, when dealing with intracellular mechanisms of microorganisms, a situation occurs where many simplifications and assumptions have to be made in order to get models that can still be handled and where enough data is available. This requires a focus on a small part of a "simple" microorganism, and a certain time frame of interest. It requires the formulation of guiding questions that should be answered with help of the models. Then, it is possible to develop models specifically for these questions, neglecting all other parts of the microorganisms. This assumes that these parts are not important for behavior of that part of the system which lies in the focus of interest.

Therefore, models of microorganisms, or models for parts of microorganisms are still limited to specific questions and cannot describe a complete cell as a system. Nevertheless, there are already attempts to build models for a complete microorganism, as e.g. the `E-cell` project [100]. This complete model tries to simulate the dynamics of a single cell by including genetic and metabolic control mechanisms. Because the simulations are not based on experimental data, results have to be interpreted with care.

Keeping in mind the available qualitative and quantitative knowledge for this work, a concept was chosen (Figure 3.1)

- that deals with a reasonably well understood biological system
- that allows measurements with sufficient accuracy for modeling of the system
- that is guided by a number of questions that can be answered with the limited model
- that results in a model that can be handled and results can be interpreted from a biological point of view

This chapter gives an overview of common modeling approaches and describes the details of the presented approach. In general, the following three steps are part of modeling approaches in this field:

1. **Biological assumptions:** At the beginning, biological assumptions about a system (e.g. a part of a microorganism) are made. This determines the type of model useful for description of the system.
2. **Mathematical setup of the models:** The models resulting from the biological assumptions have a certain structure that is important during the process of modeling. This structure determines the application of the model for describing the system.
3. **Approaches for simpler mathematical models:** Because models for describing biological systems are often complex, certain simplifications and model structures are used. Such approaches are e.g. based certain mathematical model formulations of enzymatic rate equations or a reduction of model dimension.

Microorganisms are complex in their structure. A complete, exact model of a colony of cells would have to include many phenomena as e.g. population, structure

or spatial distributions. Biological assumptions are used in order to focus on a part of the system that can be modeled. The main aspects are discussed in the following sections, as well as the kind of models that result from these assumptions. Following are examples of approaches that lead to simpler mathematical structures of the model.

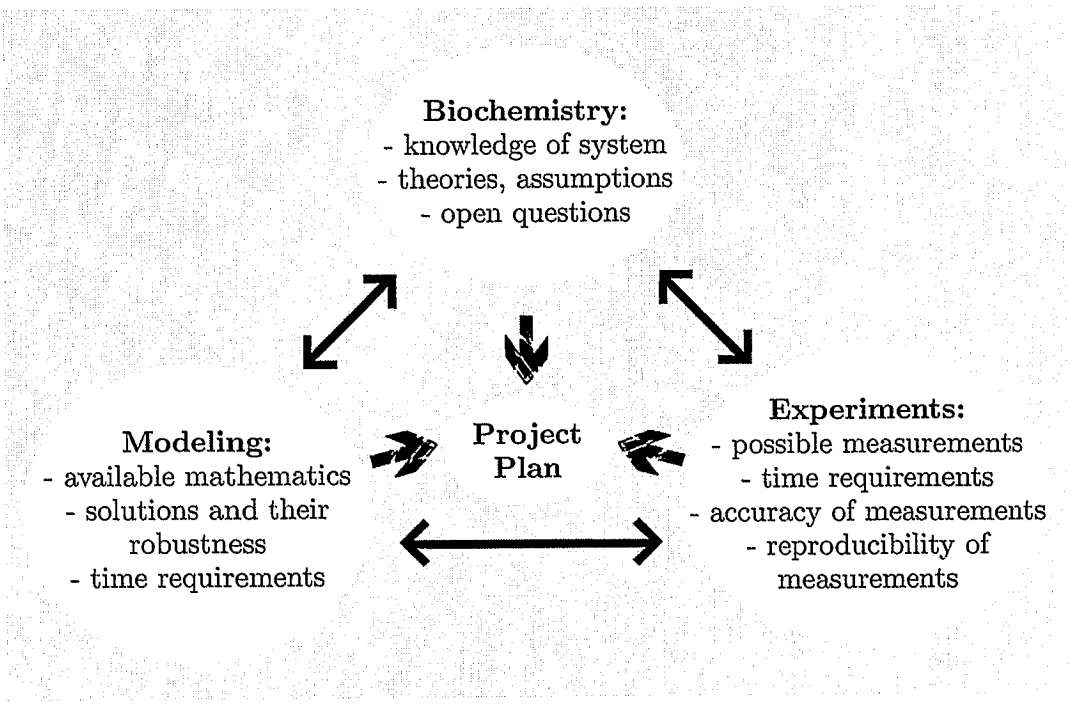


Figure 3.1.: A project plan for modeling parts of a microorganism has to consider availability of technologies, qualitative and quantitative knowledge about the system. Biochemistry provides the knowledge base for models and performing experiments. Experiments are useful for increasing knowledge in biochemistry and provide data for modeling. Modeling helps to understand complex systems in biochemistry and is needed for analyzing data from experiments.

3.2. Biological assumptions

This section describes and justifies the assumptions the dynamic models presented in this work are based on.

3.2.1. Population and single cell models

A population of microorganisms usually consists of cells that are different under many respects. Some main differences are

- size and weight
- age of cells, or the stage they are in, i.e. which part of the cell cycle
- genetic regulation states, i.e. which genes are expressed for the current conditions
- chemical composition

Some average values of these cell properties can be found in standard text books as [60] or [65]. Differences between cells in a population are important because the cells behave differently. E.g. larger cells have a larger surface area for nutrient uptake, or cells produce more structural material in certain parts of the cell cycle. Genetically, mutations occur and a population can consist of certain sub-populations.

Some models try to incorporate population variety. An introduction is given in [19]. Cell-cycles of a population of cells are considered e.g. in [15], a mass structured population model can be found in [61], the penicillin production in combination with a morphological structured model is presented in [114]. For three reasons, differences between cells of a population are often neglected in models of microorganisms:

1. From the above mentioned differences in cells of a population, size and weight are probably the only ones that can be measured. However, differences in size and weight, as all other differences, can hardly be measured with sufficient accuracy for modeling.

2. Including differences between cells of a population would increase the complexity of the models dramatically, because a system of Integro-Differential Equations would result.

3. Cells grown in a small laboratory fermenter as described in Chapter 2 and [90] are considered to differ in rather small ranges.

The underlying experimental data of this work consists of measurements of intracellular compounds of an average cell. No information is available about the various differences in the cells of the population. Also, in the time span of 45 seconds of the experiment, no significant changes in the above mentioned properties as e.g. size occur. Therefore, and in order to reduce complexity of the model, an average cell of the population has been considered and modeled.

3.2.2. Environmental conditions

In a fermenter, gradients in nutrient concentration, temperature, pressure, or pH exist. These differences result in varying non-stationary intracellular dynamics because not all cells see exactly the same environmental conditions. The differences can be especially important in changing environments, where non-linear effects occur. An exact model representation of such a system therefore has to include these differences. Such a model then should also include the influence of such differences on the cells. This, however, requires to model the movement of single cells in e.g. a fermenter. The movement can be modeled by stochastic processes or by dividing the fermenter in a number of compartments and describe movement between these compartments. As a result, either a system of Stochastic Differential Equations or an additional system of Ordinary Differential Equations on the population level results.

Obviously, it is difficult to measure gradients in an environment of cells. It is basically impossible to measure and model the movement of cells inside a fermenter with good accuracy. Recently, a model incorporating different micro environments of a large-scale fermenter and movement between the micro environments has been presented in [44]. Also, models have been developed that simulate distribution of gas and liquid in reactors [93]. For the laboratory fermenter used to generate the data of this work, measurements show a well-mixed environment [90]. Therefore, in this work, effects of gradients in a fermenter are omitted.

3.2.3. Structured and unstructured models

Often models of microorganisms are divided into two classes, structured and unstructured models. In a population, e.g. an age structure or size structure could be

introduced. For single cell models, "structured" means the consideration of intracellular compartments. It is important to distinguish between spatial compartments and conceptual compartments, see Figure 3.2. Spatial compartments are parts of the cell that are surrounded by a membrane, as e.g. mitochondria in eukaryotes. Such compartments are discussed in the following section. A conceptual compartment is the totality of a compound or a group of compounds inside the cell. An example is the biomass in general, a group of compounds of a reaction group (e.g. of glycolysis), or amino acids as a group of compounds. With this definition, a cell can e.g. be structured into reaction groups.

Strictly speaking, a completely unstructured model of a single cell is a model that considers only fluxes into a cell and out of a cell. All intracellular processes are neglected, the cell is viewed as a one-compartment black box. An example can be found in [96].

Defining useful compartments is difficult. If too many compounds are considered as one conceptual compartment, the model might not represent the cell behavior in a correct way. Fluxes between compartments are usually considered as chemical reactions. Therefore, adding more compartments to the models does not change the structure of the model but increases the dimension of the model.

Most single cell models include some structure of cells. The level of detail varies in a wide range. Some models include only a few conceptual compartments that represent a storage of compounds in the cell and therefore e.g. allow estimation of growth after nutrients are depleted [38]. Other models include uptake systems of nutrients and many intracellular compounds and enzymatic reactions [79].

Because most single cell models are structured and just differ in the level of detail, it is not very helpful to divide models just in structured or unstructured. Also, an unstructured model often results from a structured model after simplifications (e.g. with a quasi-stationary assumptions) have removed all the structure. For this work, a structured representation of the single cell is used. The more important questions about the level of structure and the kind of structure are discussed in the following sections.

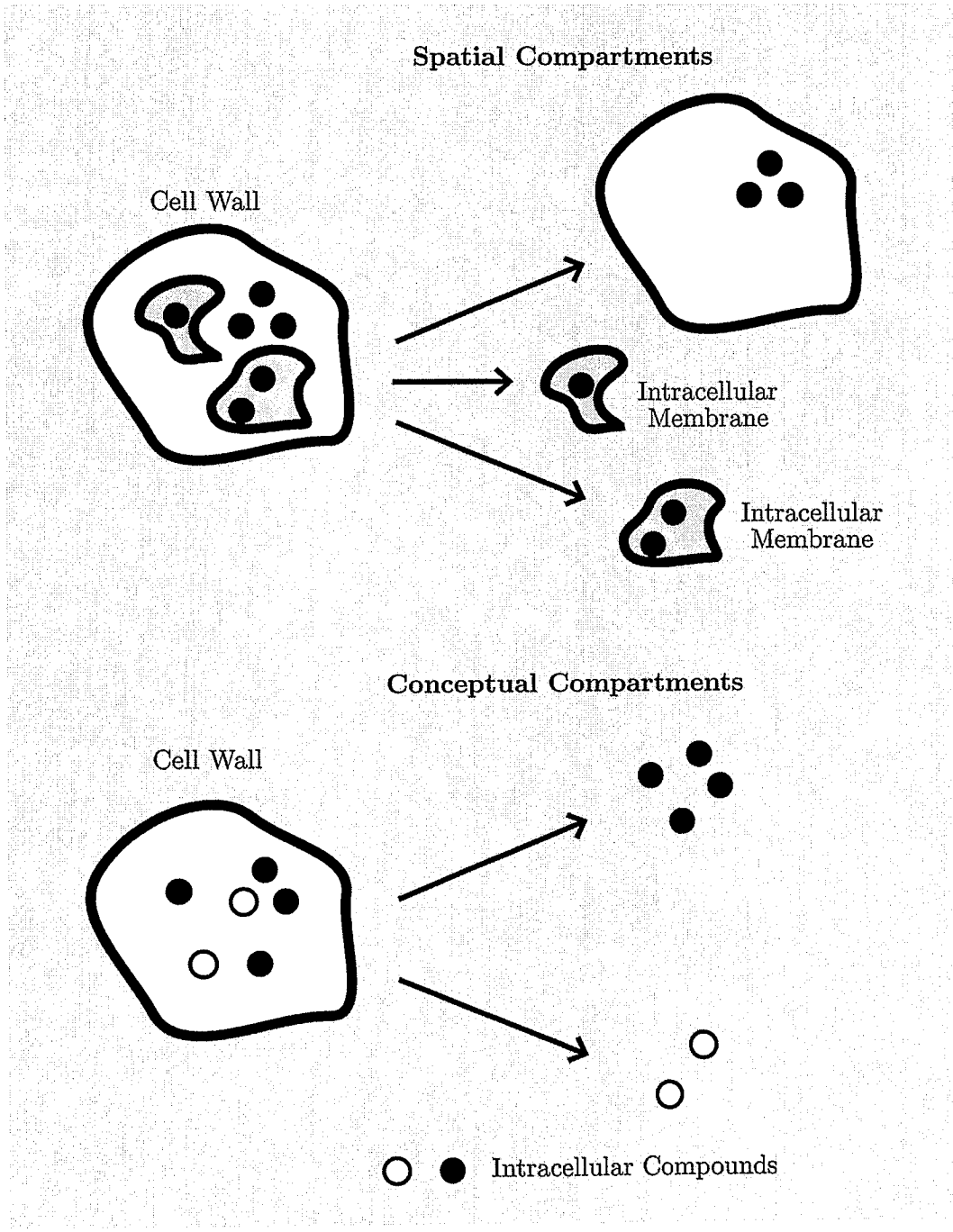


Figure 3.2.: Spatial (top) and conceptual (bottom) compartments in a cell. In the graph on the top, two parts of the cell surrounded by a membrane are shown. Together, three spatial compartments result. In the graph on the bottom, two compounds or groups of compounds are shown that are modeled as two compartments. A combination of conceptual and spatial compartments is also possible.

3.2.4. Spatial compartment models

Spatial compartmentation is important for eukaryotes or higher cells. These cells are complicated in structure and possess various intracellular compartments surrounded by membranes. For some metabolites and nucleotides (e.g. OAA, ATP), two or more separate pools exist in cells. E.g. in yeast, these two separate pools exist in the cytoplasm and in the mitochondria.

Models including spatial compartments have to account for transport systems or incorporate equations to describe the exchange of such metabolites from one pool to the other via the membrane. The approach increases the dimension of the mathematical model but can be formulated in a straightforward way.

The problem, however, is that it is almost impossible to measure intracellular compound concentrations specific to a compartment. Therefore quantitative knowledge about compartment specific concentrations in e.g. eukaryotes is rare. Still, one of the famous single cell models for yeast described in [79] includes such a compartmentation. *Escherichia coli* is a prokaryote and does not contain structural compartments. Therefore, the models of this work could be formulated without structural compartmentation.

3.2.5. Genetic and metabolic effects

Cells adapt to their environment by changing their intracellular levels of enzyme concentrations. This genetic regulation of the cell results from the dependence of enzyme production on the induction of certain genes. Genetic regulation is a relatively slow process. It usually takes at least 30 minutes until enzyme concentrations change significantly.

Genetic regulation can in principle be included in the model as a chemical reaction network. Then the model increases in dimension whereas the mathematical structure remains the same.

Some models are used to model long term effect by inclusion of genetic regulation [58]. These models focus on genetic control and neglect metabolic control through metabolic regulation. Complete models including genetic and metabolic regulation are rare because they are complex and operate on very different time scales. One example that provides a framework for modeling a complete cell including both genetic and metabolic regulation can be found in [55].

This work considers short term effects, in particular data from pulse experiments from a time frame less than one minute. In this time frame, enzyme concentrations can be considered to remain constant. Therefore, genetic regulation is neglected in the models.

3.2.6. Intracellular spatial gradients

As discussed above, populations of cells experience gradients in e.g. nutrient supply or temperature. Inside a cell, concentration gradients also exist. These gradients are likely to exist mainly for certain compounds. After a substrate pulse e.g., substrate will enter the cell. However, it will take some time until the substrate moves from the cell membrane further to the middle of the cell. Therefore, a gradient exists. Gradients also can result from membrane bound enzymes. Reaction products of the reactions catalyzed by these enzymes can be present in significantly larger concentrations at the places produced than elsewhere in the cell. Another point is the high number of large molecules in a cell which makes it difficult for large molecules to move around quickly.

For a correct description of reaction rates the described phenomena need to be included in a model. This requires formulations of movements of molecules inside a cell. Due to the high number of large molecules inside a cell and the complex structure of a cell, such models seem to be impossible to define. Also, measurements of intracellular concentration gradients are almost impossible in microorganisms. Some of the effects of intracellular movement may be examined theoretically. However, so far no modeling approaches describing these effects in small microorganisms like *Escherichia coli* are known. For higher cells, measurements have been taken and a model has been developed. The results obtained indicate that intracellular concentration changes depend strongly on the size of cells [7].

Another effect that has been modeled is the influence of an overall high concentration of large molecules in a cell. This is called macromolecular crowding and is known to increase rates of enzymatic reactions in the cell. The phenomenon can be explained with laws from the theory of entropy [105].

In the presented models, uniform distribution of the present molecules in the cytoplasm of the cell is assumed. Because modeled cell sizes are in the range of $< 10\,\mu m$, an uniform distribution seems realistic. In [91] the diffusion coefficient of water molecules in bacteria was estimated. A mean residence time of water molecules of less than 10 ms was calculated. This residence time means that water molecules move around the cell in a matter of milliseconds. It is therefore reasonable to assume

a uniform distribution of other small molecules (e.g. metabolites) in the cytoplasm of microorganisms.

3.2.7. Stochastic processes

For correct modeling of reaction rates, the total number of compound molecules is of interest. Especially for compounds that are only present in low numbers, a correct model must use the total number of molecules instead of a simple concentration. This results from the fact that binding of compounds to enzymes is a stochastic process. These effects are especially important for genetic regulation, where sometimes only a few molecules inside the cell are sufficient for regulation. Models that describe such effects include stochastic components. If these models also include other chemical reactions, a mixed stochastic and deterministic model results that is difficult to analyze.

Measurements of the intracellular number of molecules is impossible in microorganisms. However, the effects of stochastic if compared to deterministic models can sometimes be examined theoretically. Some models taking into account that only a few molecules of a certain compound are present exist [46], [56]. These models are based on stochastic processes and describe genetic regulation.

This work focuses on metabolic pathways. Metabolites are only a small part of the cell, approximately 3.5% of the dry weight ([65] chapter 3). Nevertheless, because metabolites are small molecules, this small amount of the cell still consists of a large number of metabolite molecules. For an intracellular metabolite concentration of 0.1 mM and an estimated number of $5.0 \cdot 10^{10}$ cells per liter, the number of molecules in each cell is larger than 10^8 (estimated from values found in [65], chapter 3). Therefore, stochastic processes are neglected and a deterministic model is formulated.

3.3. Structure of the models

The models of this work

- are single cell models describing an average cell
- assume environmental conditions to be equal for all cells
- include conceptual but no spatial compartments

- describe metabolic control phenomena
- neglect genetic regulation
- assume uniform distribution of molecules in the cells
- neglect stochastic effects inside the cells

This section describes the general model structure that results from these assumptions. Then, some aspects of model application for various purposes are discussed. An example is introduced that will be used for illustration of concepts throughout this work.

3.3.1. The general metabolic model

All intracellular physiological processes can be described by generalized chemical reaction kinetics. In particular, equations for transport phenomena can be formulated as chemical reactions (e.g. transport of GLC through the membrane with the PT-System is a reaction GLC + PEP $\rightarrow$ G6P + PYR). The complete cell is then described with a biochemical reaction network in common chemical notation [98] including transport phenomena. The chemical reactions describe all four basic fluxes:

Compound fluxes: Fluxes in anabolism and catabolism are processes based on chemical reaction kinetics.

Energy fluxes: Energy fluxes are described by chemical reactions, e.g. usage or production of nucleotides such as ATP or NADH.

Signaling fluxes: Fluxes for signaling (processing of information) are described with reaction kinetics, e.g. reactions for synthesis of macromolecules as DNA or RNA.

Transport fluxes: Fluxes for transport through membranes (intra- or extracellular membranes) are described with chemical reaction kinetics.

Transport phenomena are difficult and many are not yet fully understood. GLY e.g. is taken up by *Escherichia coli* by diffusion, a physical process [65]. GLC, on the other hand, is taken up by the PT-System with a membrane bound enzyme and the consumption of PEP [65]. This is a mix of a physical and a chemical processes. In the models described in Chapter 6 only the PT-system is included and described as a reaction GLC + PEP $\rightarrow$ G6P + PYR. In the following, it is shown how systems of chemical reaction kinetics are formulated in mathematical terms.

3.3.2. Notation used for this work and an illustrative example

As an illustrative example of a chemical reaction network, consider the reactions from F6P to FBP and from FBP to PEP (the intermediate steps of glycolysis are not included here). Figure 3.3 shows the example system. This illustrative example will be used throughout this work. It is always referred to as the "illustrative example".

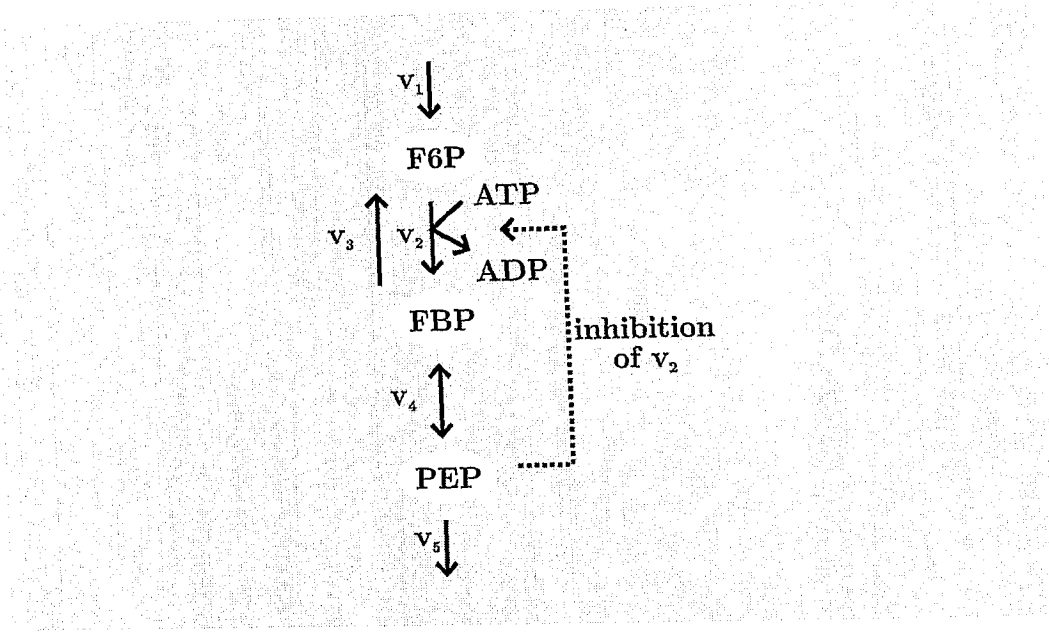


Figure 3.3.: Example of a chemical reaction system with five compounds F6P, FBP, ADP, ATP and PEP and five reaction rates (fluxes) v_1 to v_5. For clarity, fluxes are numbered here in contrast to the text where e.g. $v_3 = v_{FBP \leftrightarrow F6P}$ is used.

The following notation is used:

- By abuse of notation, intracellular compound concentrations will always be written with their abbreviations only, i.e. F6P is the intracellular compound concentration of F6P in [mM]. This is a short but more readable form of the usually used C_{F6P}.
- Kinetic rates are written as $v_{A \to B}$ for an irreversible kinetic rate from compound A to compound B and $v_{A \leftrightarrow B}$ for a reversible kinetic rate. Notice that v in the reversible case can be a net reaction rate (net-flux) that represents the

sum of a forward and a backward reaction that might be catalyzed by different enzymes. Sometimes (for clarity) the fluxes are summarized in a vector and numbered $\mathbf{v}_1$ to $\mathbf{v}_{N_v}$ where N_v is the number of reaction rates included in the model.

Reaction rates $\mathbf{v}$ have a preferred direction. This direction is usually the thermodynamically favorable direction. This direction is defined for each reaction rate. A positive rate means that the flux is in the thermodynamically favorable direction, whereas a negative value means that the flux is in the opposite direction. In this case, the reaction would take place against the thermodynamic equilibrium.

A $\rightarrow$ in the index of $\mathbf{v}$ indicates that only the forward reaction is possible (irreversible reaction), i.e. $\mathbf{v}$ is always positive. A $\leftrightarrow$ means that the reaction is reversible and v can also obtain negative values. In such cases a positive value of v means a flux from left to right. E.g. if $\mathbf{v}_{\mathrm{FBP} \leftrightarrow \mathrm{PEP}}$ is positive, a flux from FBP to PEP is active.

Using this notation, the example network can be written as

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{v}_{\mathrm{F6Pin} \rightarrow \mathrm{F6P}} \\
\mathbf{v}_2 &= \mathbf{v}_{\mathrm{F6P} \rightarrow \mathrm{FBP}} \\
\mathbf{v}_3 &= \mathbf{v}_{\mathrm{FBP} \rightarrow \mathrm{F6P}} \\
\mathbf{v}_4 &= \mathbf{v}_{\mathrm{FBP} \leftrightarrow \mathrm{PEP}} \\
\mathbf{v}_5 &= \mathbf{v}_{\mathrm{PEP} \leftrightarrow \mathrm{PEPOut}}
\end{aligned} \tag{3.1}$$

To distinguish between the forward and backward kinetic rates as for $\mathbf{v}_{\mathrm{FBP} \leftrightarrow \mathrm{PEP}}$ is important e.g. if a different mechanism is active for the reaction rates. Mostly, however, these kinetic rates are summarized in one rate equation for the net-flux $\mathbf{v}_{\mathrm{FBP} \leftrightarrow \mathrm{PEP}}$.

3.3.3. Bipartite graphs

A discrete description of a chemical reaction network leads to a bipartite graph or digraph [2], [109]. Figure 3.4 shows such a graph of the illustrative example. Metabolites are represented by the squared nodes of the graph, reaction rates $\mathbf{v}_i$ are

represented by circle nodes. The nodes are connected to indicate chemical reactions between metabolites. Because reaction v_4 converts one FBP molecule into two molecules of PEP, a double arrow is used in the graph. In this way stoichiometric properties of reactions are incorporated in the graphs. Additionally the bipartite graph or digraph is enhanced by effectors on the connectors (reaction rates) that are indicated by an arrow and a + or - sign. This enhancement violates the definition of a bipartite graph because two nodes of the same kind (two circled nodes) are connected.

Representation of chemical reaction networks using enhanced bipartite graphs or digraphs is useful for model development because it allows structured overviews. Identification of end points (i.e. compounds that appear only once in the system as a substrate or product) becomes easy. From bipartite graphs, mathematical models can be derived easier and automatically. Representation with this approach allows usage of theoretical methods and software from graph theoretical approaches like petri-nets [75].

3.3.4. The stoichiometric matrix

The chemical reaction network can be described in matrix form. This matrix can be extracted from the enhanced bipartite graph of digraph described in the last section. For the graph of the illustrative example in Figure 3.4, the matrix representation is (· represent zero entries)

$$
\begin{array}{l}
 \\
v_{F6Pin \rightarrow F6P} \\
v_{F6P \rightarrow FBP} \\
v_{FBP \rightarrow F6P} \\
v_{FBP \leftrightarrow PEP} \\
v_{PEP \leftrightarrow PEPOut}
\end{array}
\begin{array}{c}
\begin{array}{ccccc} F6P & FBP & ADP & ATP & PEP \end{array} \\
\left(\begin{array}{ccccc}
+1 & \cdot & \cdot & \cdot & \cdot \\
-1 & +1 & +1 & -1 & \cdot \\
+1 & -1 & \cdot & \cdot & \cdot \\
\cdot & -1 & \cdot & \cdot & +2 \\
\cdot & \cdot & \cdot & \cdot & -1
\end{array}\right)
\end{array}
= \mathbf{N}
$$

Numbers indicate the stoichiometric constants. $v_{FBP \leftrightarrow PEP}$ e.g. describes the reaction 1 FBP $\leftrightarrow$ 2 PEP, i.e. the overall PEP production via glycolysis. Notice that the reversible reaction from F6P to FBP is split into a forward and a backward reaction. This is necessary because only in the forward reaction ATP is consumed and ADP produced whereas in the backward reaction these compounds are not part of the chemical reaction. In contrast, the reversible reaction from FBP to PEP can be included as one kinetic rate equation. Only compounds for which balance equations exist in the model are included in the matrix. This is the reason for the fact that external metabolites as e.g. GLC are not part of the matrix.

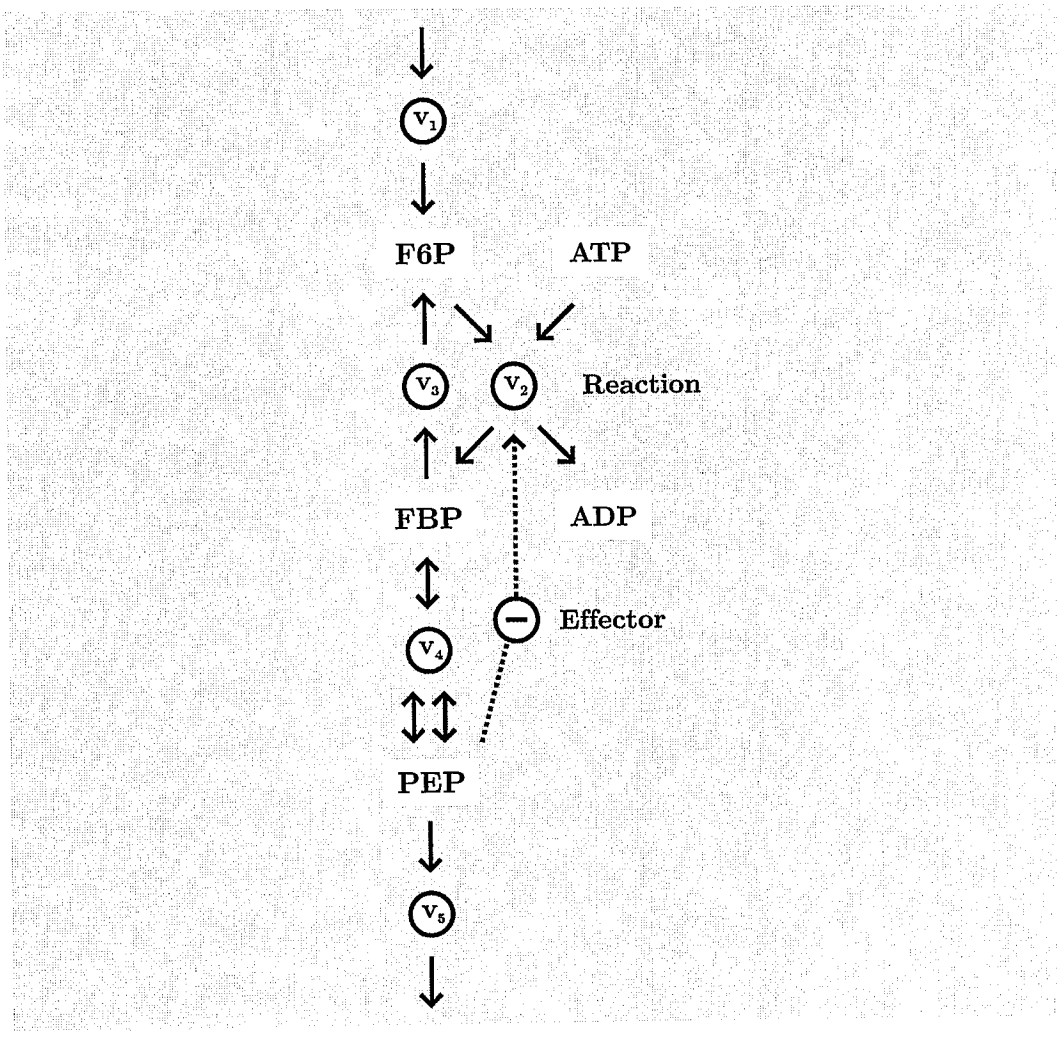


Figure 3.4.: Representation of the example pathway using an enhanced bipartite graph. See text for details.

The matrix that represents the chemical reaction network is called the stoichiometric matrix and labeled $\mathbf{N}$. Each row of the matrix represents one node $\mathbf{v}_i$ of the bipartite graph that itself represents a chemical reaction rate. Thus the rows can directly be derived from the bipartite graph and vice versa. Notice that the rows contain only the stoichiometric coefficients for the chemical reactions without effectors. The matrix therefore contains not all information from the bipartite graph or digraph.

3.3.5. Chemical reaction kinetics

Quantitative models are based on the stoichiometric matrix $\mathbf{N}$ and a flux vector $\mathbf{v}$. In the non-stationary case, the flux vector $\mathbf{v}$ is not constant with time. Instead, the reaction rates that form the flux vector are functions of compound and enzyme concentrations. The rate equations for $\mathbf{v}$ are therefore functions

$$\mathbf{v} = \mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{E}, \mathbf{P}) \tag{3.2}$$

where $\mathbf{C}$ and $\mathbf{S}$ are the vectors of all intracellular respectively extracellular compound concentrations, $\mathbf{E}$ is the vector of enzyme concentrations, and $\mathbf{P}$ is the vector of all kinetic parameters. Because $\mathbf{E}$ is usually not known *in vivo*, most kinetic rates assume $\mathbf{v}$ to be a function of $\mathbf{C}$, $\mathbf{S}$ and $\mathbf{P}$ only. The enzyme concentration $\mathbf{E}$ is then hidden in the kinetic constants in the parameter vector $\mathbf{P}$. In the following $\mathbf{v}$ will therefore be written as a function $\mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{P})$.

The rate equations include effectors such as the negative influence of PEP on $\mathbf{v}_2$ in the illustrative example. Thus the effectors included in the bipartite graph or digraph in Figure 3.4 can be found in the reaction kinetics $\mathbf{v}_i$, whereas the rows of the stoichiometric matrix contain only the stoichiometry of the chemical reactions as explained in the last section.

The description of the reaction rate equations is based on biochemical knowledge and tries to explain cell behavior based on such knowledge, e.g. [80], [27], [22]. The most famous equation is the Michaelis-Menten equations which reads as

$$v = v_{max} \cdot \frac{C}{K_m + C} \tag{3.3}$$

for a reaction rate v in dependence on two parameters v_{max} and K_m and a substrate concentration C. The equation is derived from the mechanism of one substrate binding to one enzyme. v_{max} is the maximal reaction rate, K_m is the concentration where $v = 0.5\ v_{max}$, the half saturation constant. Not for all reactions such easy mechanistic formulations are appropriate, e.g. if effectors are included in the equation the equations get more complicated. In particular, many reaction rates have several parameters that have to be determined for a correct model of the system. These problems will be discussed later in this chapter.

For the example chemical reaction system above, assume a constant influx $\mathbf{v}_1$ to F6P, and Michaelis-Menten equations for all other fluxes. The reaction rate from FBP to PEP is modeled with a reversible Michaelis-Menten equation. This equation

contains two v_{max} values indicated by V_f for the forward and V_r for the backward reaction. Also K_m values exist for both directions of the reaction. The equation reduces to the irreversible Michealis-Menten equation if the substrate or product concentration is zero. Now the non-stationary model has the following form:

$$
\begin{aligned}
\dot{\mathrm{F6P}} &= v_{\mathrm{F6PIn}\rightarrow\mathrm{F6P}} - v_{max,v_2} \cdot \frac{\mathrm{F6P}}{K_{m,v_2} + \mathrm{F6P}} + v_{max,v_3} \cdot \frac{\mathrm{FBP}}{K_{m,v_3} + \mathrm{FBP}} \\
\dot{\mathrm{FBP}} &= v_{max,v_2} \cdot \frac{\mathrm{F6P}}{K_{m,v_2} + \mathrm{F6P}} - v_{max,v_3} \cdot \frac{\mathrm{FBP}}{K_{m,v_3} + \mathrm{FBP}} \\
&\quad - \frac{V_{f,v_4} \cdot \frac{\mathrm{FBP}}{K_{mS,v_4}} - V_{r,v_4} \cdot \frac{\mathrm{PEP}}{K_{mP,v_4}}}{1 + \frac{\mathrm{FBP}}{K_{mS,v_4}} + \frac{\mathrm{PEP}}{K_{mP,v_4}}} \\
\dot{\mathrm{ADP}} &= v_{max,v_2} \cdot \frac{\mathrm{F6P}}{K_{m,v_2} + \mathrm{F6P}} \\
\dot{\mathrm{ATP}} &= -v_{max,v_2} \cdot \frac{\mathrm{F6P}}{K_{m,v_2} + \mathrm{F6P}} \\
\dot{\mathrm{PEP}} &= 2 \frac{V_{f,v_4} \cdot \frac{\mathrm{FBP}}{K_{mS,v_4}} - V_{r,v_4} \cdot \frac{\mathrm{PEP}}{K_{mP,v_4}}}{1 + \frac{\mathrm{FBP}}{K_{mS,v_4}} + \frac{\mathrm{PEP}}{K_{mP,v_4}}} - v_{max,v_5} \cdot \frac{\mathrm{PEP}}{K_{m,v_5} + \mathrm{PEP}}
\end{aligned}
\tag{3.4}
$$

For a realistic description of ADP and ATP, additional terms would have to be included in the model. Otherwise all ATP will be converted to ADP in a short period of time.

3.3.6. Representation of dynamic pathway models

Using the stoichiometric matrix $\mathbf{N}$ and the flux vector $\mathbf{v}$, the following equation describes the dynamic behavior of the chemical reaction system:

$$\dot{\mathbf{C}} = \mathbf{N} \cdot \mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{P}) \tag{3.5}$$

$\dot{\mathbf{C}}$ is the derivative with respect to time of the vector of intracellular compound concentrations. Therefore, a system of Ordinary Differential Equations (ODEs) results. The system is non-linear due to the non-linearity of the rate equations $\mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{P})$.

For the illustrative example network of chemical reactions above the following system of ODEs results (**v** with numbers in short notation used for clarity):

$$\dot{\mathbf{C}} = \begin{pmatrix} \dot{\text{F6P}} \\ \dot{\text{FBP}} \\ \dot{\text{ADP}} \\ \dot{\text{ATP}} \\ \dot{\text{PEP}} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 \\ \mathbf{v}_2 - \mathbf{v}_3 - \mathbf{v}_4 \\ \mathbf{v}_3 \\ -\mathbf{v}_3 \\ 2\,\mathbf{v}_4 - \mathbf{v}_5 \end{pmatrix} \tag{3.6}$$

During this work, estimated reaction rates do not depend explicitly on time, i.e. **v** is only a function of **C**, **S** and **P**. This means that the right hand side of Equation 3.5 is time independent and the system of ODEs is autonomous. Changes in extracellular compound concentrations (e.g. to simulate a substrate pulse) can be modeled by changing the corresponding value in the vector **S** at a specific time during solution of the system of ODEs.

Non-stationary models allow the investigation of dynamic system behavior. Simple systems were used more than 20 years ago to describe e.g. the reaction F6P $\rightarrow$ FBP that is catalyzed by the enzyme phosphofructokinase. It has been shown that oscillations can occur with a simple system of 2 ODEs due to the product inhibition of phosphofructokinase by FBP ([95], [36]).

If a steady-state for the microorganism is assumed all components of the flux vector **v** are constant, i.e. $\dot{\mathbf{C}} = \mathbf{0}$. Additionally, fluxes into and out of each compound pool add up to 0, i.e. the right hand side of Equation 3.5 is 0:

$$\mathbf{0} = \mathbf{N} \cdot \mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{P}) \tag{3.7}$$

The now constant vector of reaction rates v must fulfill the system of linear stoichiometric equations

$$\mathbf{N} \cdot \mathbf{v} = \mathbf{0} \tag{3.8}$$

which is the classical stoichiometric model.

The system of equations of the stoichiometric model usually does not have a unique solution for pathways of microorganisms. This can already be seen in the illustrative example model: Suppose a solution of Equation 3.8 has been found for the model, then multiples of this solution will be a solution again. E.g. $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}_4 = 1$, $\mathbf{v}_5 = 2$ and $\mathbf{v}_3 = 0$ is a solution, and $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}_4 = 2$, $\mathbf{v}_5 = 4$ and $\mathbf{v}_3 = 0$ is also a

solution. Additionally, the exchange flux between F6P and FBP is not unique. E.g. $\mathbf{v}_1 = \mathbf{v}_4 = 1$, $\mathbf{v}_5 = 2$, $\mathbf{v}_2 = 3$ and $\mathbf{v}_3 = 2$ is also a solution.

Stoichiometric models are used e.g. for metabolic flux analysis [102], and calculation of possible intracellular fluxes or optimal fluxes under a certain constraint [106], [20], [10]. Data from flux analysis can be used to estimate model parameters in the steady state situation. Stationary models have not been analyzed during this work.

3.3.7. The standard system of non-linear ODEs

For this work, non-stationary models are developed that are based on the straightforward modeling approach described above that leads to systems of non-linear ODEs. These models are based on a mechanistic description of chemical reaction kinetics. Many models can be found in the literature that are based on this approach. Some include genetic regulation or properties as e.g. size in addition to metabolic control. A classical model that predicts cell composition, cell size and cell shape can be found in [27]. It describes transcriptional effects and structures the cell in different compartments.

In [81] and [79] models describing the central metabolic pathways of *Saccharomyces cerevisiae* can be found. These models are focused only on metabolic control for short term experiments. The model presented in [79] is the more complicated of the two with 22 dynamic mass balances and 23 reaction rates. The system of differential equations depends on more than 80 parameters. These two models are mentioned here because their focus is the same as for the models in this work: describe metabolic regulation of the central metabolic pathways with a mechanistic model.

The examples mentioned are difficult to handle in a mathematical sense because they depend on many parameters. The illustrative example as presented in Equation 3.4 contains a total of 11 kinetic parameters. These parameters are not known a priori but have to be estimated from experimental data. Such data is still rare and is often noisy. The advantage of such straightforward modeling is the biochemical description of the system. This allows interpretation of results from a biological point of view. E.g. if the parameter v_{max} is identified, the capability of an enzyme catalyzing the reaction can be estimated.

The next section presents methods for simplifying such models and different structural approaches that lead to mathematical formulations that are easier to handle.

3.4. Linearization, simplification and structuring techniques

A model that was build to represent a system can be used e.g. for simulation of an experiment, parameter estimation, and prediction of system behavior. As already mentioned above, many modeling approaches for microorganisms result in large mathematical models with many parameters. Therefore, many attempts can be found in the literature for simplifications of such models in order to simplify usage of the model. Some attempts use a different structure of the models in order to simplify usage of models. A few famous examples are given here.

3.4.1. Linearization of enzyme kinetic reactions

Metabolic Control Analysis

Linear representations for enzyme kinetics of metabolic networks have been introduced in order to estimate kinetic parameters important in the control of fluxes. Metabolic Control Analysis (MCA) is based on such linear representations [50], [42]. The approach is based on Flux Control Coefficients (FCCs) $k_j^{\mathbf{v}_i}$ defined by

$$k_j^{\mathbf{v}_i} = \frac{\mathbf{E}_j}{\mathbf{v}_i}\frac{d\mathbf{v}_i}{d\mathbf{E}_j} = \frac{d\ln(\mathbf{v}_i)}{d\ln(\mathbf{E}_j)} \tag{3.9}$$

where $\mathbf{v}_i$ is a flux for reaction i at a defined steady state and $\mathbf{E}_j$ is an enzyme activity of enzyme j. Flux control coefficients therefore indicate the relative change in the steady state flux $\mathbf{v}_i$ in dependence on an infinitesimal change in the enzyme activity of $\mathbf{E}_j$. In a similar way, Concentration Control Coefficients (CCCs) can be defined

$$k_j^{\mathbf{C}_i} = \frac{\mathbf{E}_j}{\mathbf{C}_i}\frac{d\mathbf{C}_i}{d\mathbf{E}_j} = \frac{d\ln(\mathbf{C}_i)}{d\ln(\mathbf{E}_j)} \tag{3.10}$$

for the change in a compound concentration $\mathbf{C}_i$ in dependence on an infinitesimal change in the enzyme activity $\mathbf{E}_i$. Sometimes the enzyme concentrations $\mathbf{E}_i$ are replaced e.g. by compound concentrations that can be analyzed easier. The principle approach is the calculation of flux coefficients in dependence on other concentrations or fluxes of any kind.

If the theory is extended to reaction networks it allows the estimation of overall changes in fluxes and compound concentrations. Because the linearization assumes

a linear relationship between enzyme activity and flux it is only valid locally. The approach is therefore limited to flux distributions and compound concentrations close to the steady state used to calculate the control coefficients.

In the illustrative example five FCCs and five CCCs can be calculated if fluxes and enzyme or compound concentrations from steady states are available. It is obvious that the enzyme with the largest FCC exerts the largest control of the fluxes. If this specific enzyme is changed, it results in a large change of the overall flux.

Biochemical System Theory

Biochemical System Theory (BST, [87], [88]) assumes that reaction rates can be described by general power law representations:

$$\mathbf{v}_i = \alpha_i \prod_j \mathbf{X}_j^{\mathbf{g}_{ij}} \tag{3.11}$$

where $\mathbf{X}_j$ are system variables (metabolite concentrations, enzyme activities, effectors, etc.), α_i are apparent rate constants, and $\mathbf{g}_{ij}$ are apparent kinetic orders. Notice that $\mathbf{v}_i$ represent usually the sum of the influx and outflux of a metabolite. If logarithmic variables are introduced in the rate equations of the form in Equation 3.11, a linear system of equation results. Therefore the BST approach can be considered as a linearization of non-linear kinetics in logarithmic coordinates.

If compared to direct linearizations, linearizations on a logarithmic scale are much more accurate for non-linear reaction kinetics. Notice that the approach lacks a biochemical background of parameters for complex reactions. Therefore interpretation of results is difficult. An advantage of the approach is the easy way to define a model of a large chemical reaction system.

In order to setup models for reaction networks, reaction rates $\mathbf{v}_i$ in power law form have to be formulated for each reaction. As a result, a system of ODEs evolves. The system with power-law representations for reaction rates is called a synergistic system, or S-system [89], [104]. In [40] a model based on the S-system approach is discussed and applied to optimization of regulatory architectures in metabolic networks.

The illustrative example can be written in the S-system notation. E.g.

$$\mathbf{v}_2 = \mathbf{v}_{\mathrm{F6P} \rightarrow \mathrm{FBP}} = \alpha_2 \, \mathbf{C}_1^{g_{2,1}} \cdot \mathbf{C}_1^{g_{2,1}} = \alpha_2 \, \mathrm{F6P}_{\mathrm{F6P}}^{g_{2,\mathrm{F6P}}} \cdot \mathrm{ATP}_{\mathrm{ATP}}^{g_{2,\mathrm{ATP}}} \tag{3.12}$$

for the reaction from F6P to FBP with two educts. In order to incorporate the inhibition of this reaction by PEP the equation can be extended to

$$v_{\mathrm{F6P}\rightarrow\mathrm{FBP}} = \alpha_2 \, \mathrm{F6P}_{\mathrm{F6P}}^{g_{2,\mathrm{F6P}}} \cdot \mathrm{ATP}_{\mathrm{ATP}}^{g_{2,\mathrm{ATP}}} \cdot \mathrm{PEP}_{\mathrm{PEP}}^{g_{2,\mathrm{PEP}}} \tag{3.13}$$

where the coefficient $g_{2,\mathrm{PEP}}$ has a negative value. Equations for the other fluxes of the illustrative example are formulated in the same way.

3.4.2. Simplified enzyme kinetic reactions

Enzyme kinetic reactions are often simplified in order to reduce the number of kinetic parameters and to allow inclusion of qualitative effects in an easy way. A few examples are given in this section. From the discussed examples, the multiplicative equations, fuzzy logic approximations and pool lumping methods are used in the models presented in Chapter 6.

Multiplicative equations

Equation 3.3 shows the Michaelis-Menten kinetics as a simple rate equation. The equation depends only on a single compound concentration. For reactions with more than one educt, the Michaelis-Menten reaction cannot be used. Instead of formulating complex equations with a large number of parameters for such cases it is possible to multiply a number of simple kinetic equations. E.g. in the case of a reaction with two educts two Michaelis-Menten equations can be multiplied:

$$v = v_{max} \cdot \frac{\mathrm{C}_1}{K_{m1} + \mathrm{C}_1} \cdot \frac{\mathrm{C}_2}{K_{m2} + \mathrm{C}_2} \tag{3.14}$$

C_1 and C_2 are the concentrations of the two educts of the reaction. An inhibitor by a third compound concentration C_3 can added by another multiplication:

$$v = v_{max} \cdot \frac{\mathrm{C}_1}{K_{m1} + \mathrm{C}_1} \cdot \frac{\mathrm{C}_2}{K_{m2} + \mathrm{C}_2} \cdot \frac{K_I}{\mathrm{C}_3 + K_I} \tag{3.15}$$

K_I is an inhibitory constant.

The multiplicative approach allows automatic model generation e.g. from a bipartite graph or from databases that contain chemical reactions. In particular, it is possible to build dynamic models directly from public Internet databases. In the

illustrative example, it is possible to formulate a rate equation for the reaction from F6P to FBP as

$$v_{\mathrm{F6P}\rightarrow\mathrm{FBP}} = v_{max} \cdot \frac{\mathrm{F6P}}{K_{m,\mathrm{F6P}} + \mathrm{F6P}} \cdot \frac{\mathrm{ATP}}{K_{m,\mathrm{ATP}} + \mathrm{ATP}} \cdot \frac{K_I}{\mathrm{PEP} + \mathrm{K_I}} \tag{3.16}$$

Fuzzy logic approximations and piecewise linear approximations

Fuzzy logic based modeling tries to simplify equations for enzymatic reaction rates and is a canonical approach [59]. Instead of using equations with many parameters, effectors are included using a fuzzy approach. The approach is often based on algebraic equations such as the Michaelis-Menten kinetic to describe the chemical reaction kinetics. Inhibitors and activators are then introduced as effectors that change e.g. the maximum reaction rate v_{max}. This influence is described with fuzzy-logic rules.

If the reaction from F6P to FBP in the illustrative example is modeled with a single substrate Michaelis-Menten kinetic then

$$v_{\mathrm{F6P}\rightarrow\mathrm{FBP}} = v_{max} \cdot \frac{\mathrm{F6P}}{K_{m,\mathrm{F6P}} + \mathrm{F6P}} \tag{3.17}$$

The inhibition of PEP of this reaction can be included using a fuzzy rule for the parameter v_{max}:

$$v_{max} = \begin{cases} v_{max,1} & \text{for PEP} \leq K_{I,\mathrm{PEP}} \\ v_{max,2} & \text{for PEP} > K_{I,\mathrm{PEP}} \end{cases} \tag{3.18}$$

where $K_{I,\mathrm{PEP}}$ is a minimal concentration of PEP for inhibition of the reaction. Notice that $v_{max,1} > v_{max,2}$ for an inhibitory effect.

The reduction in number of parameters simplifies parameter fitting. On the other hand, some biochemical insight is lost due to the qualitative description of effectors.

It is also possible to use piecewise linear approximations for non-linear rate equations. The Michaelis-Menten equation can be approximated by two functions: one for small compound concentrations with a proportionality assumption and one for

large compound concentrations where the reaction rate remains constant, e.g.

$$v = \begin{cases} a_1 \, \mathbf{C}_1 & \text{for } \mathbf{C}_1 \leq L_c \\ a_2 & \text{for } \mathbf{C}_1 > L_c \end{cases} \tag{3.19}$$

where L_c is a limit concentration. This approach is similar to the definition of fuzzy rules for rate equations. The fuzzy approach is more general because it also allows non-linear functions in the piecewise rate definition.

Power-law formulations

Enzyme kinetic equations can be simplified also by using power-law forms. A power-law formulation of enzyme kinetic rate equations was already introduced in the last section, see Equation 3.11. It is also possible to define functions for reaction rates where compound concentrations appear in the exponent, i.e.

$$v = a^{k_1 \, \mathbf{C}_1} b^{-k_2 \, \mathbf{C}_2} \tag{3.20}$$

with positive constants a, b, k_1 and k_2. The second term represents an inhibition of the reaction by $\mathbf{C}_2$ due to the negative sign for k_2. A problem of such a representation is the fact that reaction rates do not converge to a finite maximal rate for increasing compound concentrations (here $\mathbf{C}_1$) but increase without limit.

In the illustrative example the reaction F6P to FBP can be formulated as

$$\mathbf{v}_{\mathrm{F6P} \rightarrow \mathrm{FBP}} = a_{\mathrm{F6P}}^{(k_1 \cdot \mathrm{F6P})} \cdot a_{\mathrm{FBP}}^{(k_2 \cdot \mathrm{FBP})} \cdot a_{\mathrm{PEP}}^{(-k_3 \cdot \mathrm{PEP})} \tag{3.21}$$

where k_3 has a negative sign to describe the inhibitory effect of PEP on the reaction rate $\mathbf{v}_{\mathrm{F6P} \rightarrow \mathrm{FBP}}$.

Cybernetic modeling

Cybernetic models include genetic and metabolic control and are therefore designed to build complete models of microorganisms, e.g. [33] or [31]. The approach does not require the explicit formulation of any chemical reaction rate and therefore simplifies model setup significantly.

In cybernetic modeling, cells are thought to aim at metabolite homeostasis while

growing and adapting to their environment in an optimal way. The term homeostasis here means almost constant intracellular compound concentrations, a phenomenon often observed in microorganisms. The reaction network is considered as a network with multiple feedforward and feedback control mechanisms. Assuming optimality constraints allows simplifications of the reaction system because details of reaction kinetics are eliminated. In the illustrative example the dynamics can be described by

$$\dot{\mathbf{C}} = \mathbf{N_{int}} \cdot \mathbf{v} + \mathbf{N_{ex}} \cdot \mathbf{v} \tag{3.22}$$

where $\mathbf{N}_{int}$ is the stoichiometric matrix describing network reactions or condensed pathways. $\mathbf{N}_{ex}$ describes exchange fluxes with the medium, i.e. uptake or excretion processes. $\mathbf{v} = (\mathbf{v}_1, ...\mathbf{v}_n)$ is the vector of fluxes.

The cybernetic modeling approach in general does not formulate equations for the fluxes $\mathbf{v}_i$ - even though some fluxes might be explicitly formulated. Instead, in order to solve the model system, constraints are introduced. Such constraints are e.g. a maximal growth rate, maximal production of structural compounds, maintain homeostasis, or maximum GLC uptake rate. Mathematically this is expressed by minimizing or maximizing one or more of the fluxes $\mathbf{v}_i$ in Equation 3.22. The problem is thus reduced to a solution of the system of linear Equations 3.22. If the system contains some explicitely formulated fluxes, these appear as nonlinear constraints of the linear system.

In the illustrative example a constraint would be e.g. a maximal flux through the pathway. Without further constraints, however, there is no limit to the solution and the maximal flux is infinity. This illustrates the need for additional constraints in cybernetic models. An additional constraint could be a fixed value for the flux $\mathbf{v}_{\mathrm{PEP}\rightarrow\mathrm{PEPout}}$. Then the maximal flux through the pathway is limited.

Cybernetic models provide a tool for describing and analyzing cellular networks without explicitly formulating rate equations. However, finding realistic constraints that are necessary for system analysis is the crucial point of cybernetic modeling.

3.4.3. Pool lumping to reduce model dimension

A system describing metabolic regulation can be simplified by eliminating compound concentrations from the model. The approach sets up complex models (possibly also with complicated reaction rate equations) and then reduces the dimension of the model by elimination of compounds.

In [103] a model of the main metabolic pathway of *Saccharomyces cerevisiae* is shown. The model consist of 97 differential equations and 83 rate equations. This system is reduced by excluding pseudo steady state metabolites. These are metabolites that have high turnover rates. The accumulation of such metabolites can be neglected and these metabolites eliminated. Additionally, a time scale separation is performed in order to reduce the system further. Very fast and very slow processes can be eliminated or pools lumped together to a single pool. As a result, a system of 27 differential equations with 15 reaction rates results.

Elimination of compounds to reduce model dimension requires a significant amount of knowledge about intracellular behavior. If this knowledge is available, however, a significant reduction of the problem dimension results.

In the illustrative example, the reactions from FBP to PEP could be eliminated, see Figure 3.5. This results in four ODEs instead of five and reduces the number of rate equations also from five to four. The overall flux through the pathway may then be described with the same accuracy as with the complex model. However, other reaction rates in the system that are influenced by FBP cannot be modeled correctly after elimination of FBP from the system.

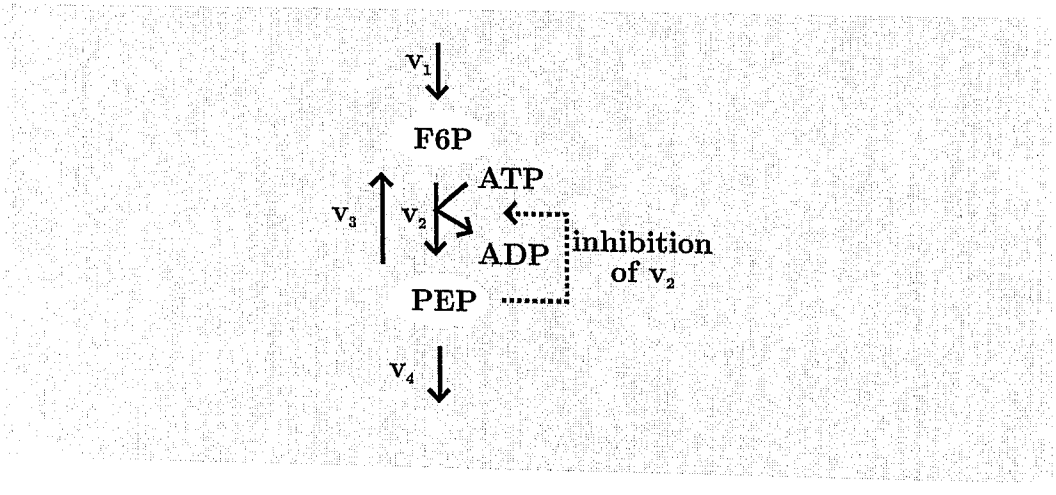


Figure 3.5.: The simplified illustrative example. The FBP concentration is eliminated from the model.

3.4.4. Top Down Approach: conceptual modeling

In order to make modeling of a system easier and more obvious, cells can be divided into functional units. As described in [55], units consist of cellular processes that are functionally linked. These units can be described in detail or by using a black

box model. The approach aims at representing the complete cell including genetic and metabolic regulation. It allows to model parts of the cell in a simple way if no detailed information is available. The idea is combined with a software concept that allows to model functional units with different levels of detail by exchanging mathematical modeling objects.

In the illustrative example a first model could summarize the reactions from F6P to PEP in a functional unit called glycolysis, see Figure 3.6. Notice that the functional representation lacks the ATP consumption and the inhibition by PEP. Instead the flux through glycolysis could be regulated by the available substrate mix: if GLC is available only, the flux through glycolysis is high. If the cells grow mainly on e.g. Acetate (ACE), the flux through glycolysis is lower or even reversed because Gluconeogenesis may be active.

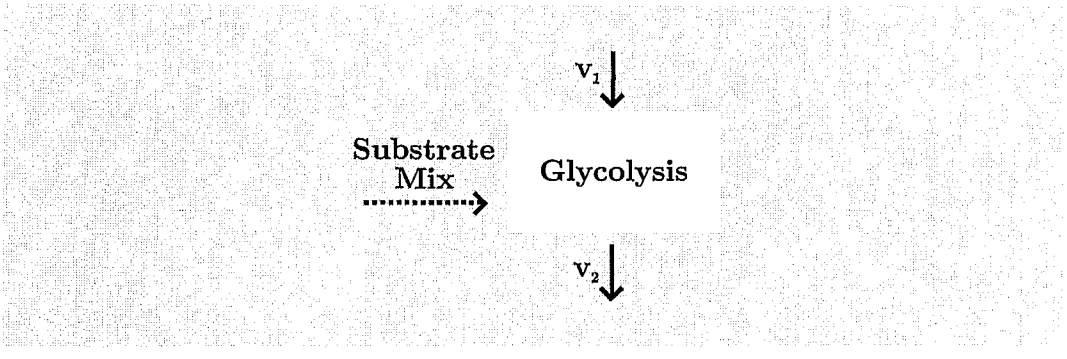


Figure 3.6.: The simplified illustrative example. The reactions from F6P to PEP are summarized as a functional unit "glycolysis".

In a top down approach more details can be added to the model, finally replacing the functional unit "glycolysis" by the individual reaction steps. This approach allows to estimate the importance of detailed representation of functional units by analyzing the differences between models with different levels of detail. If no differences between a simple and a detailed model can be seen, the detailed description can be eliminated from further models.

3.5. Model details and problems

As described in the last section, the models of this work are mechanistic models based on the reaction network. Three important points of such an approach are discussed in more detail in this section due to their importance:

1. **Pathway selection:** It is not possible to include the complete pathways in a model. How to choose and how reliable is the available information from textbooks and in public databases on the Internet?

2. **Kinetic reaction rates:** Metabolic regulation is based on enzyme kinetic reaction rate equations. How to find a reasonable enzyme kinetic equation for this complicated mechanisms?

3. **Additional effectors of kinetic reaction rates:** Many effectors influence the enzyme reaction rate. How can these effectors be included in a kinetic rate equation?

3.5.1. Pathway selection

The knowledge of the underlying pathways of microorganisms is necessary for simulation and biological interpretation. Even though *Escherichia coli* is a well studied organism, identification of pathways is not finished. One reason for this is the fact that many enzymes catalyze more than one reaction and therefore can accelerate bypass pathways. Also, some pathways are active under so far unknown conditions like the methylglyoxylate pathway from DHAP to PYR ([65], chapter 14).

Of greater importance, however, is the fact that many metabolites are relatively unstable, making it difficult to measure them. Therefore the actual *Escherichia coli* metabolic pathway could be much more complicated then the recent understanding suggests. It is important to take these effects into account throughout model development. A model should not just be able to model certain situations with a fixed pathway model but be flexible in the whole setup of its components and structure. Wherever models are presented in this work, a discussion of the currently available knowledge on *Escherichia coli* pathways, enzymatic regulation phenomena and the developed mathematical model will clarify some problems that occurred during analysis.

With the help of MMT Software that allows to build alternative models of a pathway in a matter of minutes a large amount of models could be created for this work. Analysis of these models can give hints which pathway model is the correct one.

3.5.2. Kinetic reaction rates

Enzymatic reactions are simple in principle (Figure 3.7). Substrates bind on enzymes and are released in a different chemical composition as products of the enzymatic reactions. A closer look at the enzymatic structure, process of binding and dissociation, however, reveals a complicated picture (see e.g. [21], [94]) that is summarized in this section. The phenomena mentioned here are used to derive enzyme kinetic rate equations. In the models discussed later, such equations are introduced and used.

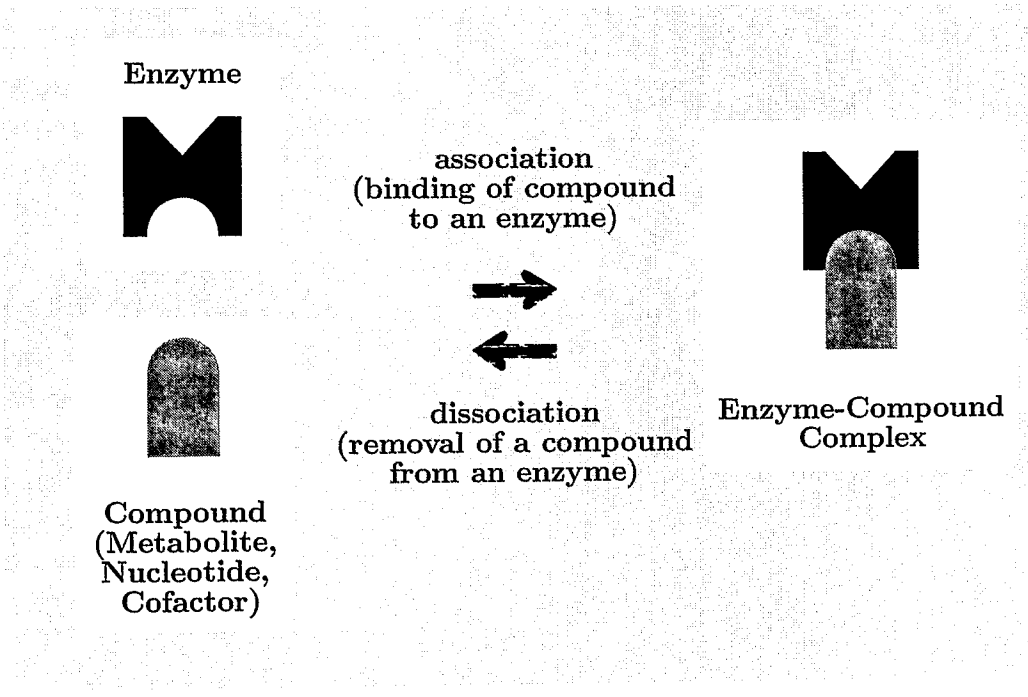


Figure 3.7.: In enzymatically catalyzed reactions, one or more compounds (substrates) bind to an enzyme (association) and one or more compounds (products) are released (dissociation).

Conformation: The three dimensional structure of enzymes cannot be analyzed or calculated easily. Without knowledge of the exact structure it is only possible to estimate what kind of substrates can bind on what site of the enzyme.

Number of binding sites: It is often not known how many binding sites an enzyme possesses, see Figure 3.8 for an example. Occurrence of several binding sites can significantly accelerate the reaction rate per enzyme molecule.

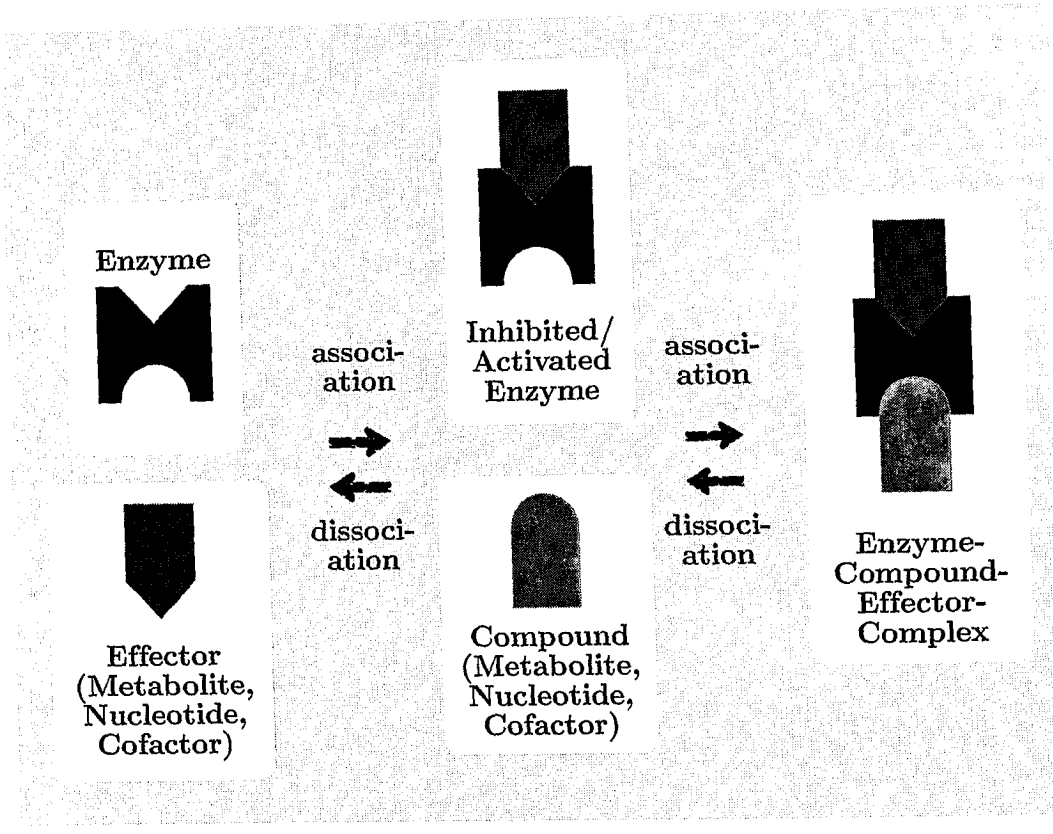


Figure 3.8.: An enzyme that has already a compound attached has a different probability of binding to another compound. Many other effectors not shown as e.g. other compounds present in the cell and temperature or pH influence probability of association and dissociation.

Probability of association (binding) and dissociation: Parameters for the probability of association and dissociation are very difficult to measure. Also, for more complex enzymes, the catalyzing effect on a reaction rate is usually not multiplied by the number of binding sites of the enzyme. Instead, the association and dissociation constants change.

Reversibility: Most processes are reversible, i.e. products can bind to the enzyme and substrate can be released from the enzyme-substrate complex.

Order of binding: In multi substrate or multi-product cases it is sometimes important what substance binds first to the enzyme.

Cofactors: Compounds that bind to an enzyme can change association of dissociation constants for the binding of compounds to the same enzyme.

Macromolecular crowding: It has been shown *in vitro* that a high concentration of certain large compounds can increase reaction rate of enzymatic reactions. This phenomenon is called macromolecular crowding [105], [83].

Metabolite channeling: Under yet unspecified circumstances a large fraction of a specific enzyme can be bound in the enzyme-substrate-product complex leaving only a small fraction of free enzyme, substrate-enzyme, or product-enzyme complexes. It is thought that a large fraction of enzyme bound in the enzyme-substrate-product complex increases the reaction speed. This effect is termed metabolite channeling. Metabolite channeling is connected to conditions of macromolecular crowding [83].

Most of the mentioned parameters and effects have to be estimated from *in vitro* experiments. It is then not clear if the same effects occur also in an *in vivo* environment.

3.5.3. Additional effectors of kinetic reaction rates

Other effects influence the association and dissociation constants of enzyme binding. e.g.

- the concentrations of substrates, products and cofactors
- enzyme concentrations
- compartmentation of cells and distribution of enzymes, substrates and other compounds inside the cell
- temperature, pressure, pH

The number of effects that cannot be quantified makes approximations necessary. Usually only one or two effectors of enzymatic reaction rates are included in the rate equations used in this work. The equations are given in Chapter 6 where some models of the *Escherichia coli* pathway are presented.

4. The MMT Software

In order to build complex models describing the pathways of microorganisms, software tools are required. Such tools allow to construct and analyze models faster as if models are build manually from scratch. They also help to keep an overview of complex pathway models that have been setup and results from simulations and parameter fitting. This chapter describes the advantages of a modeling tool that is based on a database system and introduces the structure and functioning of such a tool: the Metabolic Modeling Tool (MMT) developed for this work.

4.1. The need of efficient modeling software

There is a number of obvious reasons for using software tools for modeling metabolic networks in the way used in this work. Some are:

- The models consist of a large number of chemical reactions that can be defined once and used for several model setups.
- A graphical interface is important to keep an overview of the network of chemical reactions. This interface should be combined with structuring techniques for the network in order to allow users to explore the networks.
- Parameter values can be defined easier if they are organized nicely, e.g. in a table, by the software.
- Results can be viewed directly in a structured way after simulations or parameter fittings have been performed.
- Setup of the mathematical equations in a format useful for numerical solution algorithms can be performed by software tools.

Various approaches have been made to build tools that address the above points. The following sections give an overview of available tools. The list is not complete but captures the main approaches.

4.1.1. General modeling software for dynamical systems

A large number of software is available for simulating dynamical systems. Three examples are mentioned here that show the variety that exists. Each tool has its specific capabilities and lacks others, which makes application for the purpose of this work impossible or inconvenient.

- MODELICA [68] is a tool for models based on Differential Algebraic Equations. It includes algebraic algorithms for an automated model simplification. The tool lacks a graphical user interface and algorithms for parameter fitting. It is free of charge.
- SIMULINK is a MATLAB [99] extension that allows to build dynamic models. The tool allows model setup with a graphical user interface. Simulations and parameter fitting algorithms are available. SIMULINK and MATLAB are commercial software tools.
- MAPLE [107] is a software focused on algebraic methods. It can simplify model equations and perform model analysis with algebraic models. Solution algorithms for ODEs are available but parameter fitting is not possible. MAPLE is commercial software.

4.1.2. Modeling software developed for chemical reaction systems

A number of software tools have been developed that focus on chemical reaction networks. These tools allow definition of chemical reactions and rate equations for the reactions. Some tools include predefined rate equations. A short overview is given here.

- MetaModel [23] is an early development of a metabolic simulator. It contains algorithms for solution of ODEs. It is limited to 20 reactions and 30 metabolites. The project is discontinued.
- METATOOL [72] is focused on structural properties of metabolic networks and is limited to steady-state analysis. It calculates the null space of the steady-state equation $\mathbf{N} \cdot \mathbf{v} = \mathbf{0}$ and performs system reduction.
- MIST [29] is a metabolic simulator with a graphical user interface. It is limited to calculation of control coefficients. The project is discontinued.
- ESSYNS [49] is a tool that focuses on analysis of models in S-system representation. The latest version dates from 1991.

- METASIM [97] is another early development. It bases modeling of cell regulation on an object-oriented approach. The latest version dates from 1992.

- KINSIM and FITSIM [6], [25] are tools for setup of chemical reaction systems and for parameter fitting. The model is setup in a notation similar to a programming language. The tool lacks a graphical user interface and is limited to 100 species and 100 reaction steps. The software is free of charge but no further development is taking place (latest version from 1997).

- E-cell [100] is an integrated modeling approach, i.e. it includes genetic regulation, metabolism and signaling phenomena. It provides a graphical overview of the pathway and numerical algorithms for solving the system of ODEs. It lacks algorithms for parameter fitting. The software is free of charge and an ongoing project.

- SCAMP [85] is a tool developed for chemical reaction networks. The tool is command line driven and is not limited in the number of reactions and metabolites. It provides e.g. solution algorithms for the system of ODEs and algorithms for steady state analysis, model reduction, calculation of control coefficients, and calculation of the Jacobian. It lacks parameter fitting algorithms. The software is free of charge and its development has halted. The tool is succeeded by JARNAC.

- JARNAC [86] is the continuation of the SCAMP project. It provides the same functionality as SCAMP and additionally includes a graphical user interface. JARNAC is an ongoing project.

- DynaFit [57] is a tool for performing least-square regression of chemical kinetic, enzyme kinetic, or ligant-receptor binding data. The tool provides a graphical user interface and contains algorithms for solution of ODEs and parameter fitting. Simulation results (time-courses) are presented in graphical form. The tool does not provide an interface for easy setup of large chemical reaction systems and is therefore more suitable for small systems. No information is available if the project is ongoing.

- DBsolve [37] is a powerful tool for complex chemical reaction networks. It allows solution of non-linear algebraic equations and systems of ODEs. Additionally, bifurcation analysis of ODEs and parameter fitting to experimental data can be performed. The tool has an graphical user interface. It is severely limited in the dimension of the reaction system: a maximum of 50 ODEs and 50 parameters can be used only. It is an ongoing project.

- GEPASI [63] is another powerful tool for complex chemical reaction networks. It includes algorithms for solution of ODEs and parameter fitting. Additionally, metabolic control and stability analysis can be performed. The tool lacks an

overview of the chemical reaction network which makes it difficult to define large models. The latest version dates from 1999.

4.1.3. Criteria for an efficient modeling software

In addition to provide graphical interfaces for model setup and result interpretation, modeling software useful for large chemical reaction systems should contain more features. Efficiency of modeling software increases with the following automated features:

- Simplifications of the mathematical model using algebraic methods.
- Translation of the model equations into a procedure of a programming language in order to get better performance during simulation and parameter fitting.
- Extensions of the model by e.g. explicitly calculating derivatives of model equations with respect to parameters.

Another important aspect of modeling software is the combination with a database system. In the field of genomics, database systems are essential for all analysis due to the amount of available data. In contrast, use of databases for chemical reaction kinetics including explicit model formulations and experimental data is still rare. All above mentioned modeling tools are not based on database systems. However, database systems have some big advantages also for applications in this field:

- an increase in storage efficiency because relations avoid redundant entries
- a systematic approach to data organization and storage
- a data consistency that cannot be maintained without use of database systems
- an easier reuse of information from other models already developed
- an easier storage of results

The need to build software tools on database systems that include chemical reaction kinetics has been generally realized and software systems are already in the development phase [32], [112]. There are also approaches to define a formal modeling language similar to the SQL language for relational databases and combine such a modeling language with a relational database system [74].

4.1.4. What is the difference between available software tools and MMT?

There have to be good reasons to start programming a software that somebody else already made available in a similar way. The MMT software was developed due to limitations in the available software tools mentioned above that have been considered too severe.

General modeling software as `MODELICA` or `SIMULINK` has been rejected because this software is not specifically for chemical reaction networks. This makes it difficult to keep an overview of complex chemical reaction networks that are included in the model. Only four of the above mentioned tools specific for chemical reaction networks allow parameter fitting to experimental data: `KINSIM/FITSIM`, `DynaFit`, `DBsolve` and `GEPASI`. `KINSIM/FITSIM` does not have a graphical user interface and is limited to 100 compounds and reactions. `DynaFit` and `GEPASI` are not convenient for large chemical reaction networks because they do not provide pathway overviews. `DBsolve` provides such an overview, but it is severely limited in the model dimension: 50 ODEs and 50 parameters are not enough for complex pathway models of microorganisms.

The most important limitation common to all of these points that lead to the decision to develop the MMT software is the database connection. `DBsolve` allows to import information from the WIT [113] database. This information is limited to the reaction network, i.e. the stoichiometric matrix. None of the above mentioned tools are connected to a database system that stores complete model information, parameters, and results.

4.2. Global structure of the MMT software

Simulation software nowadays subdivides modeling into pre-processing, numerical algorithms, and post-processing [17]. Pre-processing is concerned with the model setup, post-processing means analysis and interpretation of results. The MMT software is based on the same approach. A general overview of the software architecture can be seen in Figure 4.1. Figure 4.2 shows the user interface for representation of the chemical reaction network and Figure 4.3 the window for setup of a kinetic rate equation for a chemical reaction.

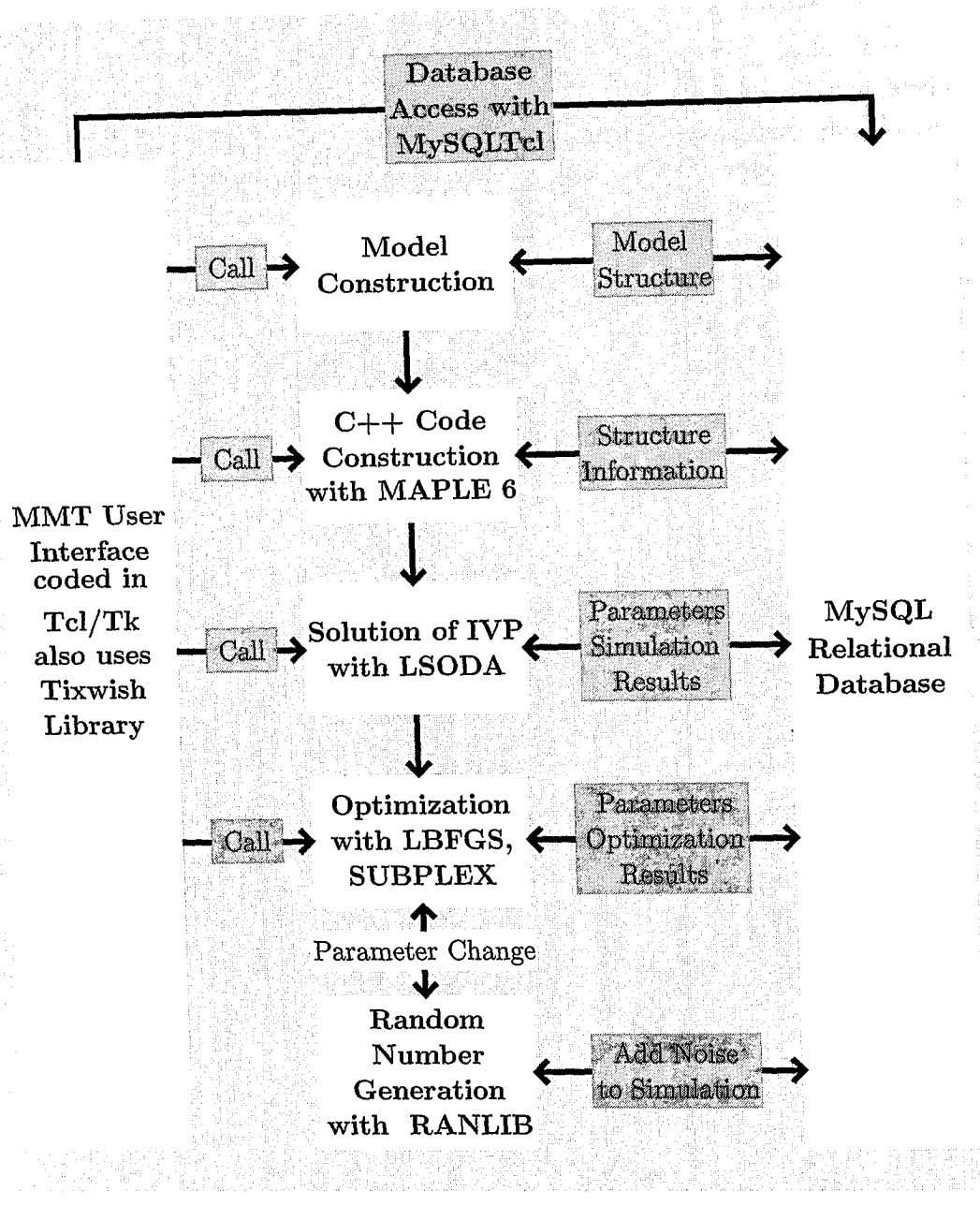


Figure 4.1.: Global structure of the MMT Software. Red/italic text indicates a software library, external software, or external algorithm used during a specific task. Vertical arrows between the processes indicate the direction when building new models. The random number generator is called during parameter fitting and to add random noise to simulation output.

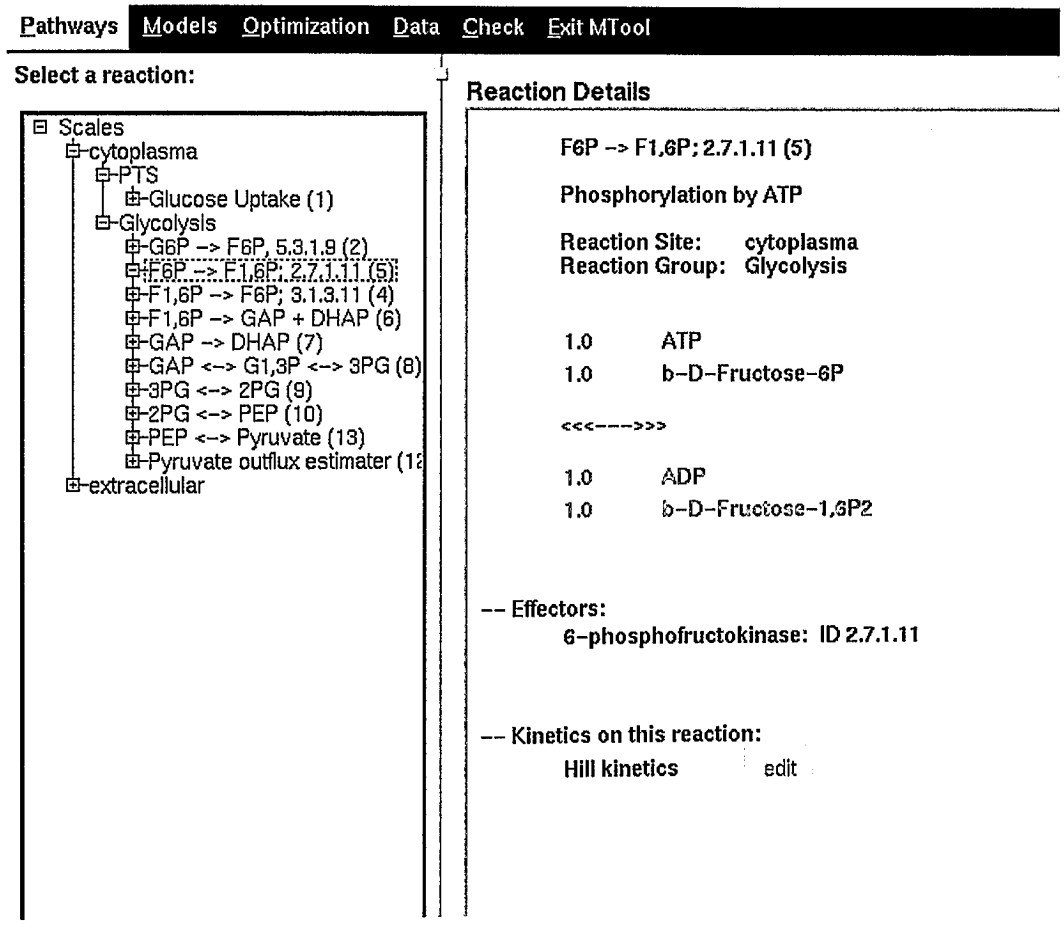


Figure 4.2.: The main window of the MMT user interface provides an overview of the reaction network. The left part of the window shows the individual reactions ordered by reaction groups (Glycolysis, PTS) and compartments (cytoplasm, extracellular). The right part of the window shows the details of the reaction currently highlighted in the reaction tree. "Effectors" are all compounds that influence the reaction (enzymes, inhibitors etc.) but are not included stoichiometrically in the reaction. A click on the "edit" button in the "Kinetics" section opens the window for the kinetic rate equation defined. This window is shown in the following Figure 4.3.

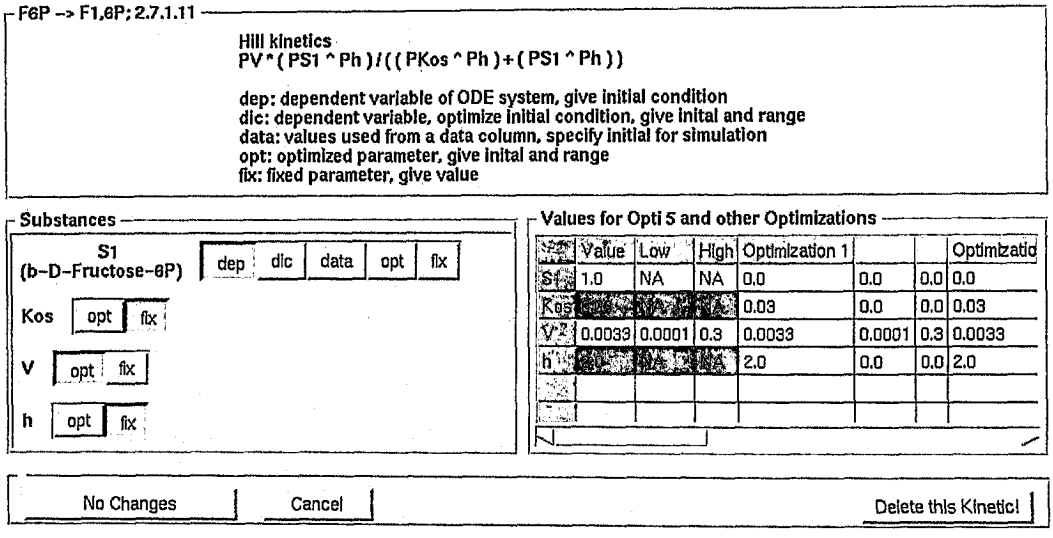


Figure 4.3.: The window where details of equations for chemical reaction kinetics are defined. Shown is a Hill kinetic for the reaction from F6P to F1,6P. The upper part of the window displays the equation and explanations for the possible definitions of variables and parameters. These definitions can be seen in the lower left part of the window. The lower right shows initial conditions and ranges for variables and parameters and corresponding values from other parameter fittings ("Optimizations") with the same model.

The figure shows that the user interface on the left hand side is used to control all processes from model construction to simulations and parameter fittings. Some of the subprocesses are performed with external software and numerical algorithms. A main point of the whole system is that it is based on a relational database that contains all information about the model and all results. Thus, there is a unique point where all relevant information is saved in an efficient way.

All external software and algorithms used by the MMT software except for MAPLE are free of charge for research purposes. Additionally, all parts of the software are available at least for the two important operating systems LINUX and MICROSOFT WINDOWS.

4.3. The MMT database

The concept of the MMT software is based on a relational database where complete model information, simulation and parameter fitting values, and results are stored. All other parts of the software are build around this database. They have the purpose to make use of the stored information for model definition, solution of ODEs or the parameter fitting problem, as well as representation of results.

The database system was setup for the specific problem of modeling chemical reaction systems. It stores such networks in a set of tables. The part of the database that contains the mathematical models is focused on the specific model formulation introduced in Chapter 3.

4.3.1. Structure of the database

For the given problem, the relational database turned out to provide a good structural background for efficient storage of the required information. The complete database was designed with an entity relationship diagram [62]. The diagram is too complex to be shown here. Setup of the database was then performed according to the *Third Normal Form* for relational databases [62]. Therefore, an efficient and consistent storage is guaranteed already from the architecture of the database. The structure of the MMT database is divided into three levels that will be explained in detail in this chapter. The three levels are (substantially described in more detail in the next sections):

1. **Groups of Models:** The database contains various groups of model structures, usually referred to as the *modeling groups level* of the database.

2. **Models:** Each group of models can contain different models. This is the *model level* of the database.

3. **parameter fittings:** For each model, various simulations and parameter fittings (parameter fittings) can be performed. This is usually referred to as the *parameter fitting level* of the database.

Figures 4.4 and 4.5 show the three levels for the illustrative example. Division of the database into these three hierarchical structures increases efficiency and allows easy creation of new models.

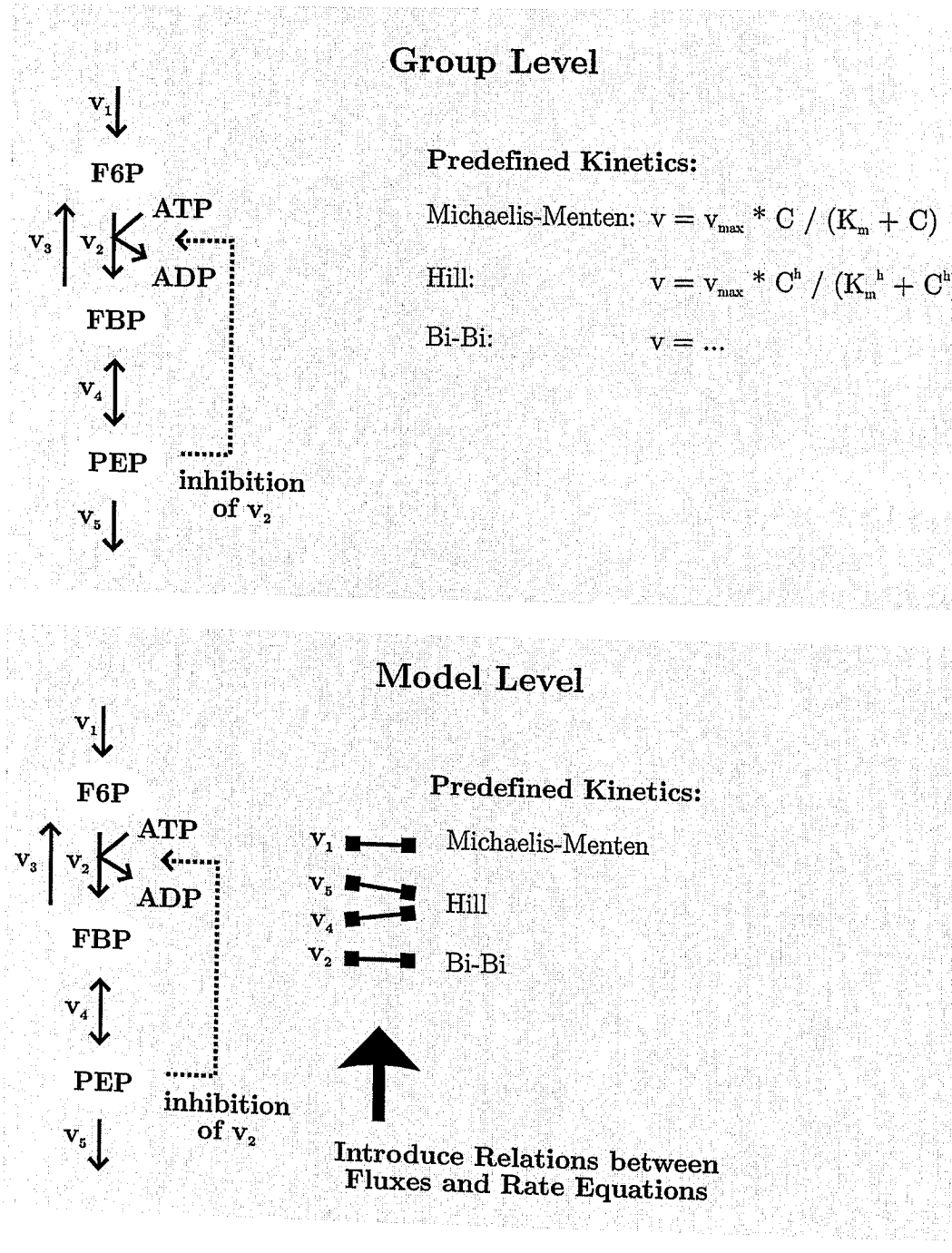


Figure 4.4.: The group and model levels of the database for the illustrative example.

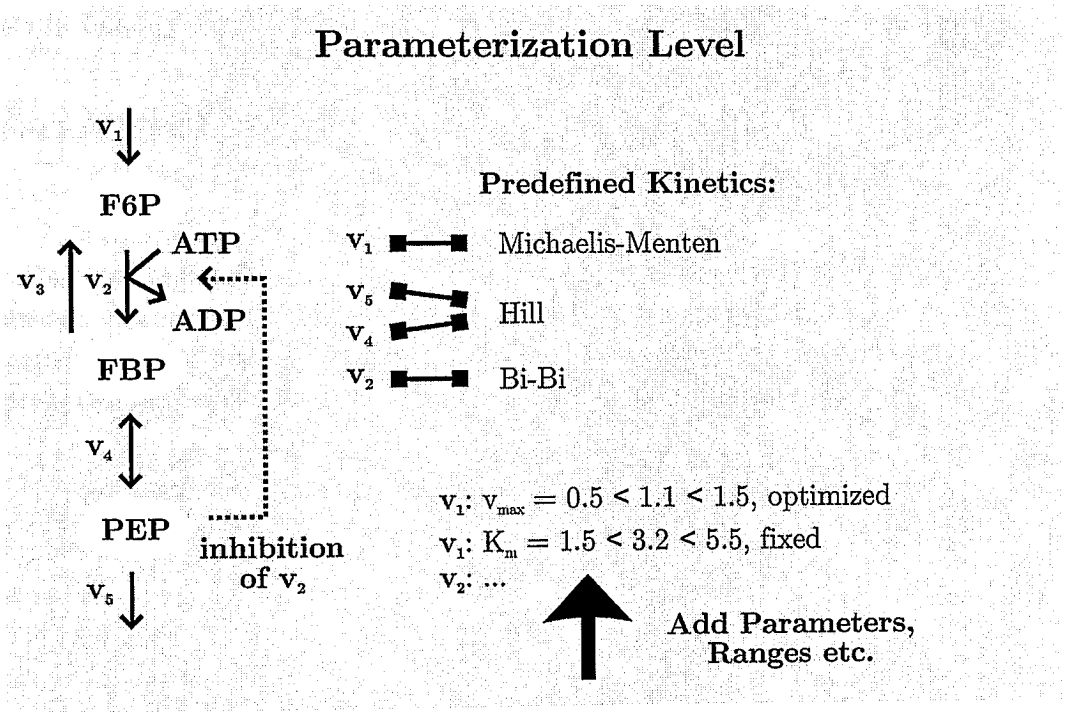


Figure 4.5.: The parameter fitting level of the database for the illustrative example.

The modeling groups level

The modeling groups level is the highest level of the database system. It is called modeling groups level because it stores information that is used by a collection of models in the next lower level, the model level (explained in the next paragraph). The group level contains all information about a chemical reaction network. In detail, the tables belonging to the level store information about

- all compounds of the network.
- the chemical reaction network, i.e. the chemical reactions with educts, products, enzymes, and effectors in a structured way. This means that the reactions are grouped by location which could be e.g. cytoplasm, extracellular or mitochondria.
- reaction groups of the network which are used to divide a large network in smaller parts. E.g. a network can be divided into glycolysis, PP-Pathway, etc.
- spatial compartments of the model, these are e.g. the cytoplasm, mitochondria, but also the extracellular environment (fermenter).

- data files related to compounds, i.e. a relation between a compound of the network and a column in a file containing experimental data.
- a collection of mathematical formulations for chemical reaction kinetics that can be related to a chemical reaction if a specific model is defined.

Figure 4.6 gives an overview of the modeling groups level. Notice that the mathematical formulations for chemical reactions are not yet related to a chemical reaction but are only possible formulations. Figure 4.7 shows how two reactions of the example model (influx to F6P and reaction from F6P to FBP) are stored in the group tables.

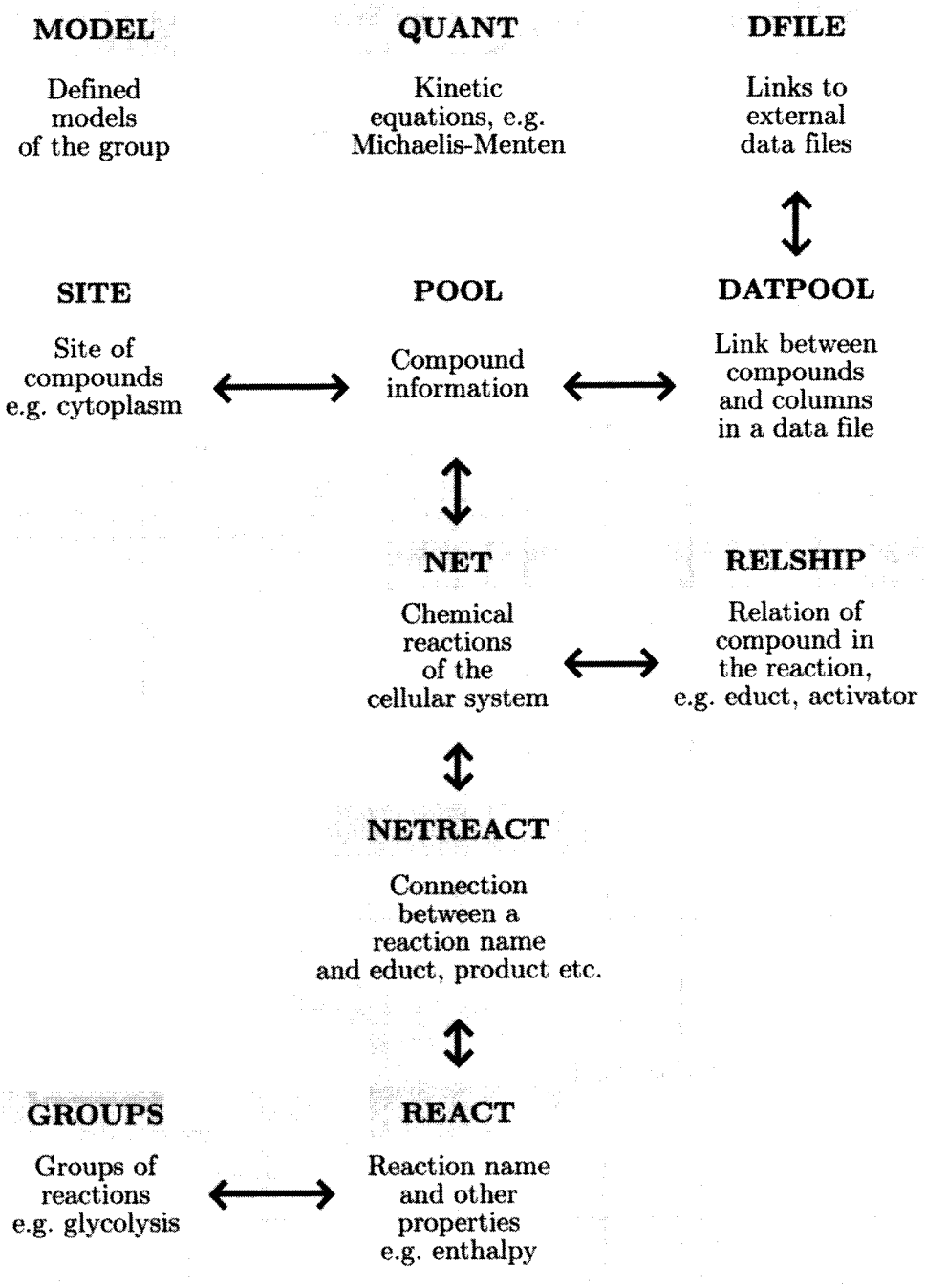


Figure 4.6.: The tables of the modeling groups level in the MMT database. Arrows indicate relations between the tables of the database.

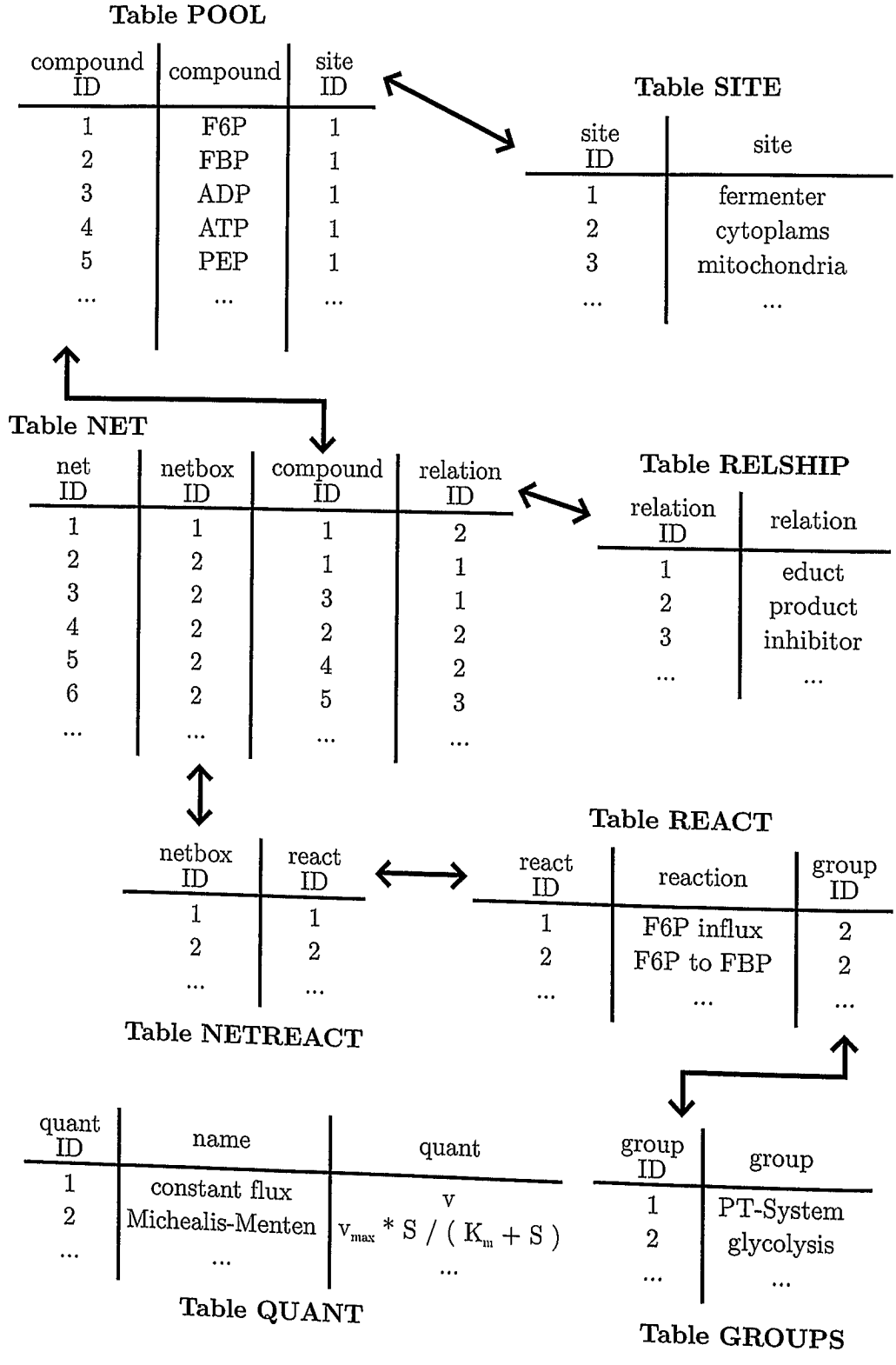


Figure 4.7.: Entries of tables of the modeling groups level for the illustrative example. Not all tables, columns, and entries are shown. Arrows indicate relations between the tables.

The model level

See Figure 4.8 for an overview of the model level. The model level inherits all information stored in the tables of the modeling groups level. The model level adds to the information stored on the modeling groups level merely relations between mathematical formulations of the table QUANT and chemical reactions stored in the table NET. Mathematical formulations for the rate equations can be added to each chemical reaction step. However, not every chemical reaction needs to be part of a specific model.

If a reaction network is defined in the modeling group tables, a new model of this complete network or parts of this network is defined by two tables that contain only relations between existing tables. For the purpose of building a large number of different models for comparison, this setup therefore allows efficient storage and quick definition of new models. It also guarantees consistency between the models because all information is stored in common tables on the modeling group level and therefore is consistent between all models of a group. Figure 4.9 shows parts of the entries for the illustrative example model. It can be seen that the table KINETIC relates each variable of a kinetic equation to a compound or parameter.

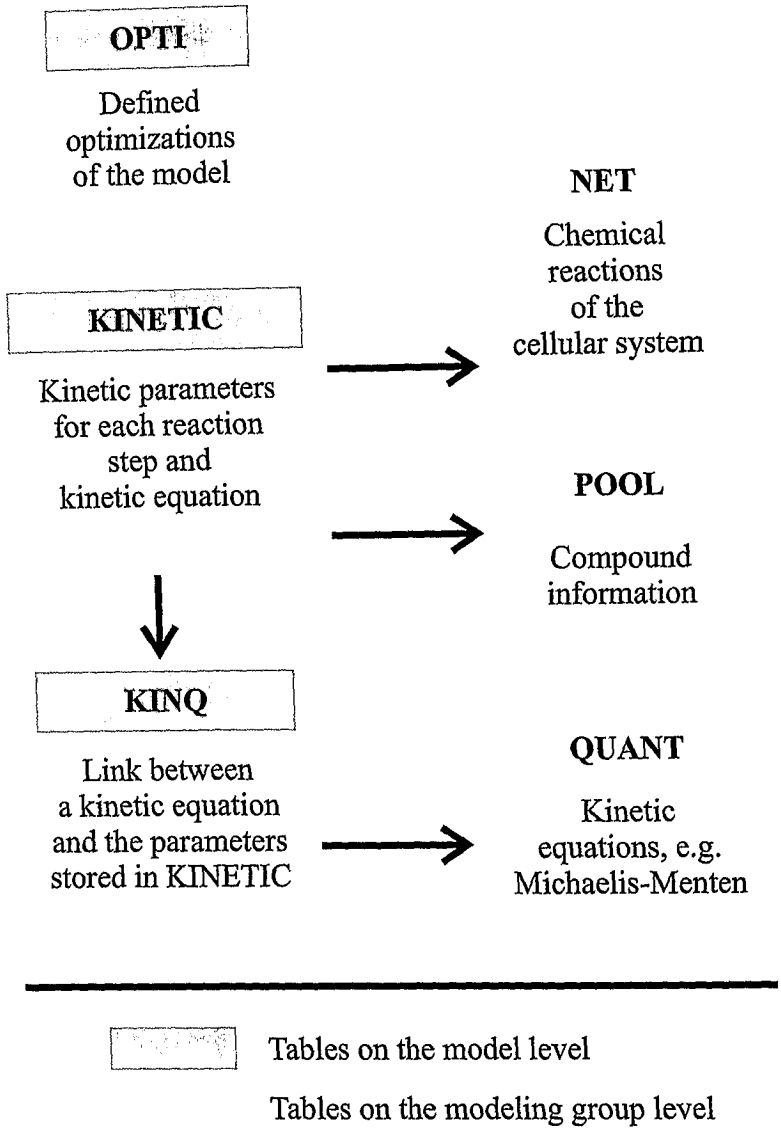


Figure 4.8.: The tables of the model level and the relations to the tables of the modeling groups level.

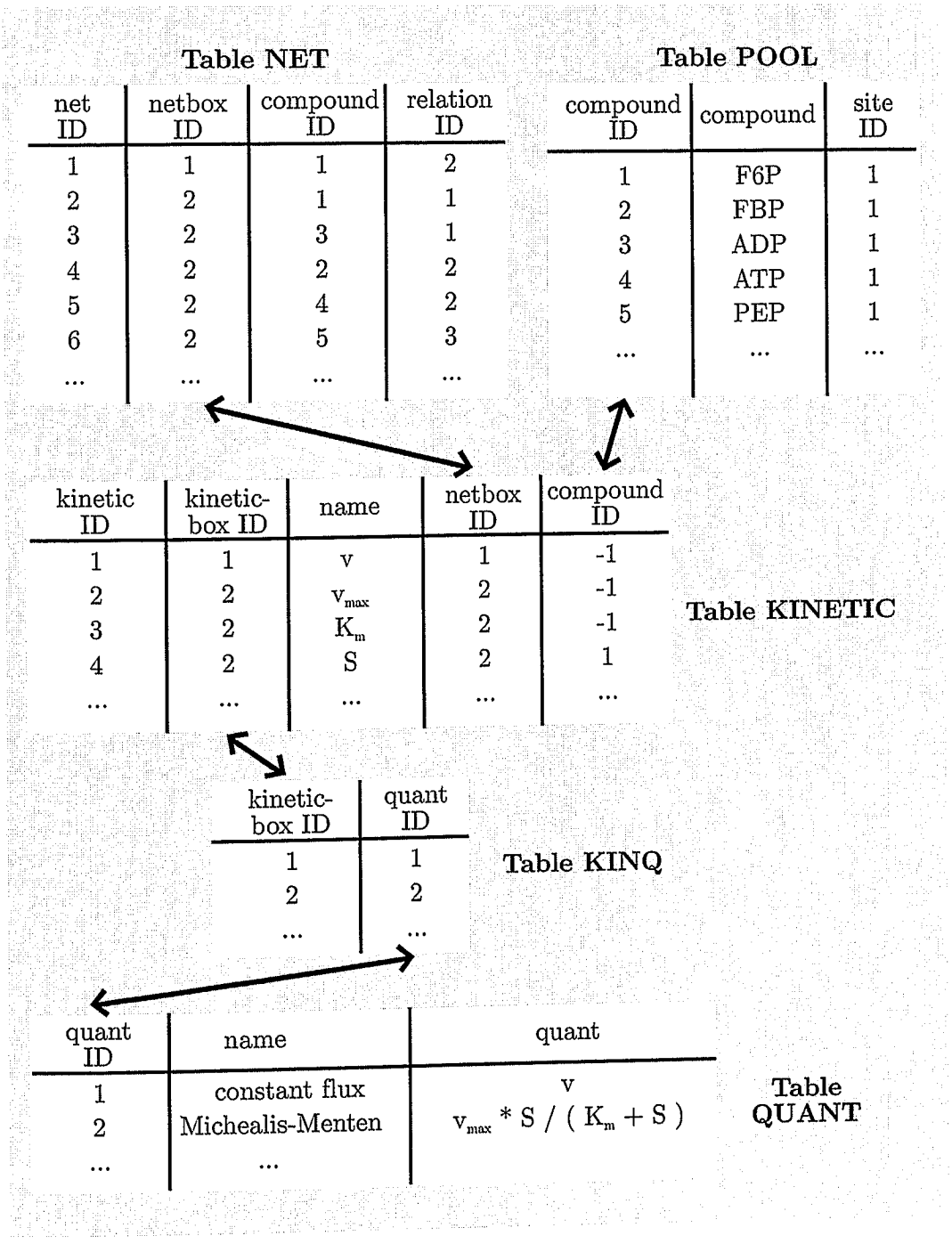


Figure 4.9.: Entries of tables of the model level for the illustrative example. Not all tables, columns, and entries are shown. Arrows indicate relations between the tables. In table KINETIC compound IDs with value -1 represent parameters whereas values > 0 represent compounds out of the table POOL.

The parameter fitting level

The third level of the MMT database deals with the parameter fitting of the defined models. The purpose of the database development was the usage for simulation and parameter fitting. These two tasks require some additional information that is stored in the tables of the parameter fitting level (Figure 4.10):

- information about variables. Options are dynamic, fixed or use values from experimental data.
- information about which initial conditions and parameters are fixed and which are subject to parameter fitting.
- definition of ranges for optimized initial conditions and fitted parameters.
- definition of weights for dynamical variables subject to parameter fitting. Weights will be explained in Section 5.3.
- definition of the data file used for a specific parameter fitting as a relation to the table DFILE on the modeling groups level.
- definition of parameters for the simulation and parameter fitting tasks (which algorithms, tolerance parameters, etc.).

Storage of all these variables and parameters including all starting parameters and algorithm specific parameters allows the reproduction of results. In addition to these information, results are stored in the four results tables RESVAR, RESERR, RESSIM and RESFLUX. Notice that this information is redundant because, as just argued, reproduction of results is always possible. However, reproducing results requires computation time that is usually more limiting than the disk space required for storage of results.

It is advantageous to allow definition of several parameter fittings for one model. For a defined parameter fitting, an unlimited amount of simulations and parameter fittings can be performed. This means that an parameter fitting algorithm can be started several times with varying initial conditions, parameter values and ranges. Additionally, different parameter fitting algorithms can be chosen and parameter values specific for such an algorithm can be changed. All initialization parameters of the algorithms are stored in RESINIT, results are stored in the four result tables RESVAR, RESERR, RESSIM and RESFLUX. The user defines how many results of a single parameter fitting are stored in the result tables. Most of the times it is sufficient to store the final result of a parameter fitting and not all intermediate results.

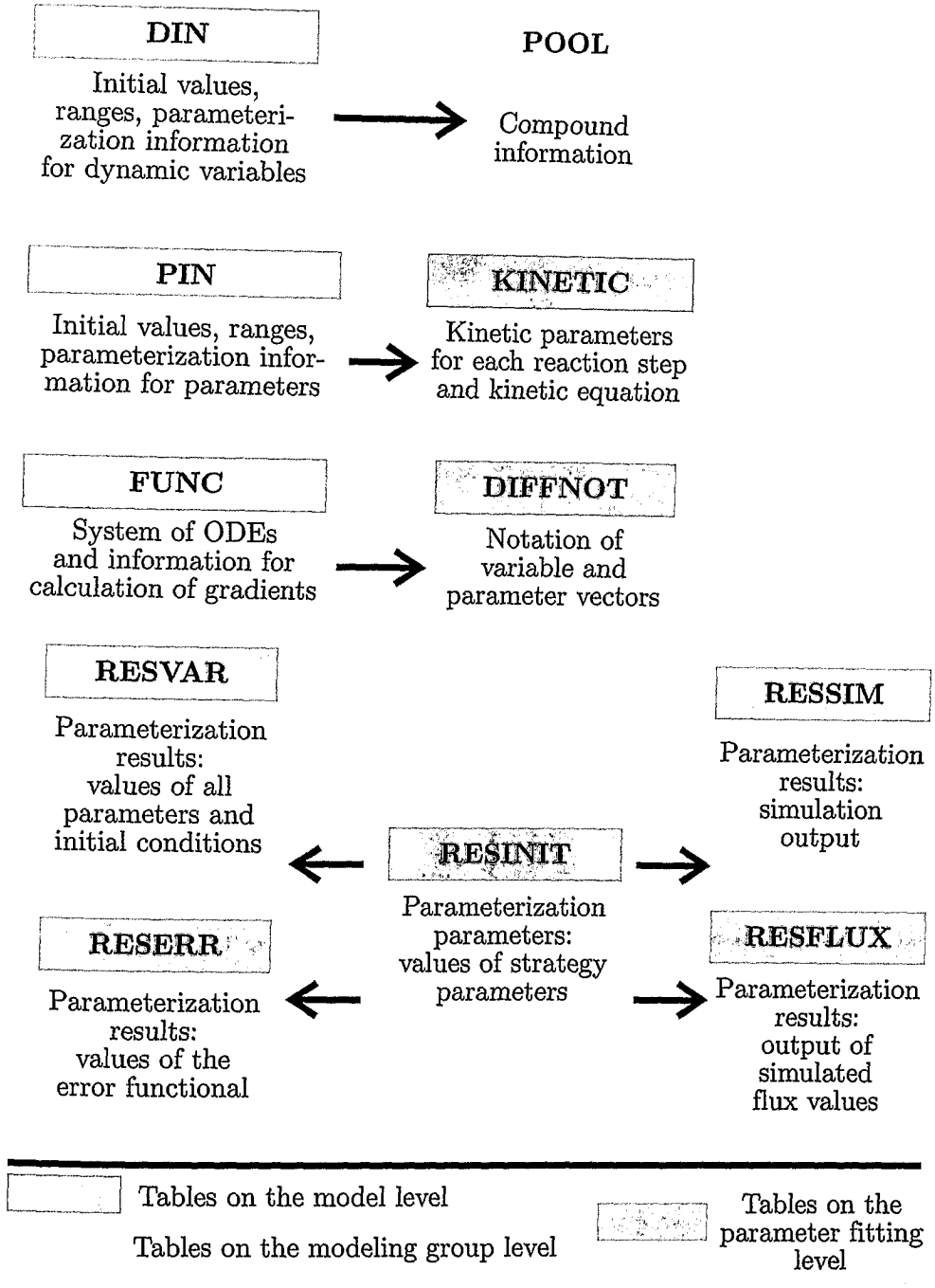


Figure 4.10.: The tables on the parameter fitting level and their relations to tables from the other two levels of the MMT database. Tables FUNC and DIFFNOT are automatically created for construction and export of the model as a C-code.

However, a new parameter fitting has to be defined if the structure of the parameter fitting problem is changed. This can result from changing either the number of dependent variables or the number of optimized initial conditions or parameters. Assume that the properties of a parameter v_{max} are changed from "optimized" to "fixed". This changes the total number of fitted and fixed parameters of the model. It removes an entry in the vector of fitted parameters and adds an entry in the vector of fixed parameters. Because initial values of parameters or parameter fitting results are stored in such vectors, changing the dimension of these vectors would result in inconsistencies in the result tables.

Storage of results was setup in such a way because it is more efficient. Also a clearer structure of the database results if new tables on the parameter fitting level are used for a new mathematical structure of the model. Especially the tables RESFLUX and RESSIM would otherwise be very large and it would be time intensive to extract a specific result from these tables.

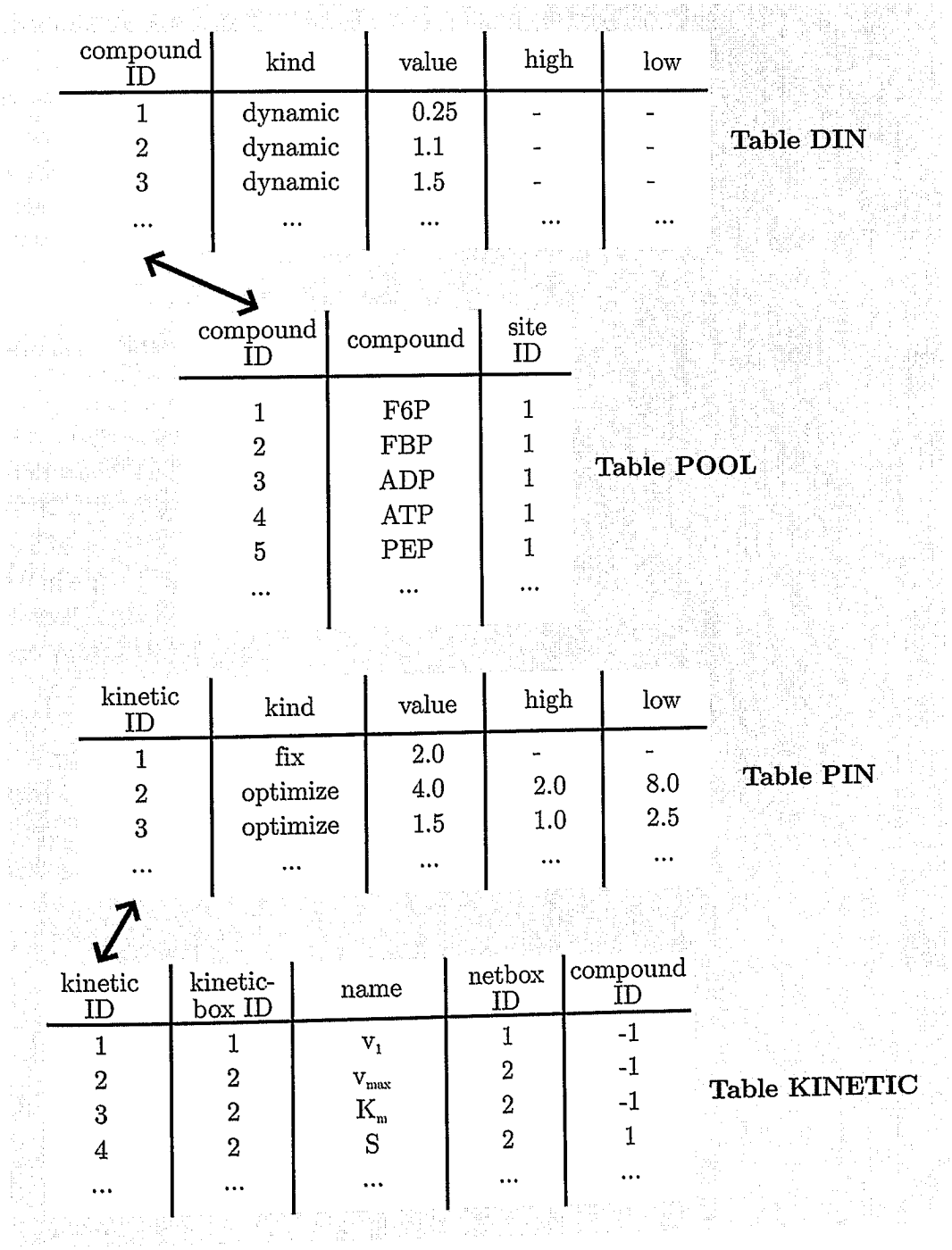
Figure 4.11 shows the tables DIN and PIN on the parameter fitting level for the illustrative example. The tables with the results are not shown. They consist of simple lists either of parameter values or of compound concentrations for each time step.

The relational database system `MySQL`

The relational database system `MySQL` was used for implementation of the database structure. `MySQL` is a commercial software system [101]. It is, however, free for use in science and education. The database system `MySQL`

- is a relational database that supports the standard query language SQL
- is easy to install
- has no limitations in number and sizes of tables
- is a full multi-user database that can be installed on a server and used by many users simultaneously
- can be accessed from other programming languages as e.g. `C` and `Tcl/Tk`

The database is currently installed on a system with an Intel Pentium III processor with 800 Mhz. The current size of the database content is 0.78 Gigabyte. About 0.75 Gigabyte or 96% of the total are results. The defined 50 Models that are stored in the database take up the remaining 4% (30 MB) of the storage.



compound ID	kind	value	high	low
1	dynamic	0.25	-	-
2	dynamic	1.1	-	-
3	dynamic	1.5	-	-
...	...	...	...	...

Table DIN

compound ID	compound	site ID
1	F6P	1
2	FBP	1
3	ADP	1
4	ATP	1
5	PEP	1
...	...	...

Table POOL

kinetic ID	kind	value	high	low
1	fix	2.0	-	-
2	optimize	4.0	2.0	8.0
3	optimize	1.5	1.0	2.5
...	...	...	...	...

Table PIN

kinetic ID	kinetic-box ID	name	netbox ID	compound ID
1	1	v_1	1	-1
2	2	v_{max}	2	-1
3	2	K_m	2	-1
4	2	S	2	1
...	...	...	...	...

Table KINETIC

Figure 4.11.: Entries of tables of the parameter fitting level for the illustrative example. Not all tables, columns, and entries are shown. Arrows indicate relations between the tables.

Because an increase in the overall size of the database results in bad performance of the MMT, a change of the result storage is anticipated in the next version of MMT. It has been argued above that storage of all results in the database saves computation time. However, it has become clear now that this computation time is lost again during database access. Therefore it is more efficient to use the available parameters in RESINIT and RESVAR and perform a simulation to obtain the time courses stored in RESSIM and RESFLUX.

4.4. The MMT user interface

A view at the structure of the MMT database presented in the last section shows that a user interface is essential for modeling. This interface has to provide an overview of all the defined models, mathematical details, simulation and parameter fitting settings, and results. It also has to help the user to quickly define new models and parameter fittings and show results in a way that allows an interpretation. The user interface is also needed to guarantee consistency of the database entries.

The MMT user interface is coded in the script language `Tcl/Tk`. With use of the publicly available library `MySQLTcl` [78] SQL commands can be send directly from `Tcl/Tk` to the database and results can be stored in `Tcl/Tk` variables. Widgets from the publicly available extension library `Tixwish` were used. This library provides easy to use widgets that could be included in the `Tcl/Tk` code of the MMT user interface.

Some advantages and disadvantages of the script language of `Tcl/Tk` are:

- − Future development of `Tcl/Tk` has become questionable
- − No object oriented language
- − Cryptic commands require detailed documentation
- − Low performance
- \+ Widely distributed and free of charge
- \+ Extension for access of SQL databases available
- \+ Supports `LINUX` and `MICROSOFT WINDOWS`
- \+ Easy to program, especially for graphical user interfaces
- \+ Many extensions as `Tixwish` available that save programming time

The MMT user interface consists of a main procedure and 40 subroutines for the individual tasks (overall number of lines approximately 24000).

The following sections describe the options and tasks that are handled by the MMT code. The tool only stores information after performance of a specific task. All information is checked for consistency and immediately written to the MMT database.

4.4.1. Setup of a modeling group

To setup a modeling group, compounds and reactions have to be written into the database. This is mainly done by a feature "Add Reactions". This feature shows in a table all available compounds of the model. The user can click on a compound to include it in the chemical reaction. New compounds can be defined also. If a new compound is defined, the software asks for other details of this compounds in order to get all required information for consistency of the MMT database. Double entries are rejected, but still a compound can exist several times - e.g. a system will have to include OAA in the cytoplasm and the mitochondria.

After all educts, products, and effectors of a reaction have been typed, the global properties of a reaction have to be defined. Required fields are a name and a reaction group (e.g. glycolysis). Figure 4.2 above shows a screen shot from the overview of the chemical reaction network as provided by the MMT user interface.

Mathematical equations for the table QUANT can be defined separately. Equations are stored in MAPLE notation which allows to copy equations from or to a MAPLE file. As mentioned in the last section, the database stores relations between compounds and columns in external data files. These data files contain the experimental data. The MMT software analyzes the data file and asks the user to assign a compound to each column of the data file. The data file is then sorted with respect to time and a new external data file is produced. This data file can be used from all models of the modeling group, the information which column belongs to which compound is stored in the table DATPOOL.

After completion of the definitions on the group level, various models can be created and analyzed. Notice that certain restrictions apply to changes in the modeling group tables after models have been defined:

- no reactions that are part of a model can be deleted.
- stoichiometric coefficients can be changed only if no results are stored that

rely on these coefficients. Change of stoichiometric coefficients is sometimes useful, e.g. for energy costs of groups of reactions. For most reactions the stoichiometric coefficients are fixed.

- compounds and effectors can be added to reactions but they might make results unusable.

Two strategies should be used in order to avoid problems that could result from these restrictions:

1. Include all reactions needed from the beginning in the modeling group.
2. If changes or additions have to be made (e.g. inclusion of a new pathway), copy the complete modeling group information to a new modeling group (this is automatically done) and make the changes in the new group.

4.4.2. Define models

In the MMT database structure models are represented by relations between reactions and mathematical formulations. Therefore, defining models for a modeling group can be done by choosing a mathematical equation and a chemical reaction from tables produced by the user interface. After an equation and a reaction have been related, the software asks the user to specify which parameter in the equation corresponds to a compound in the reaction and which parameter is really a parameter. E.g., if the Michaelis-Menten equation

$$v = v_{max} \cdot \frac{C}{K_m + C} \tag{4.1}$$

is used, the user could define the C to be G6P and will define v_{max} and K_m as parameters.

Often new models differ from existing models by only a few equations. The MMT software allows to copy models, and then make these simple changes without a complete redefinition of the system. A complete modeling group including all defined models can be copied or a single model can be copied. If a single model is copied, the copy remains in the same modeling group as the original one. This is a necessary limitation because the tables on the model level are related to the models on the group level and a different modeling group possesses different group tables.

4.4.3. Defining and performing simulations and parameter fittings

Analyzing and using models is a time consuming process. The MMT software helps to efficiently perform simulations and parameter fittings. For most tasks mentioned, default procedures and parameter values exist that can be redefined by the user. In particular, parameter values from any previous simulation or parameter fitting can be reused. Either all parameter values from a previous result are copied or single values can be copied by the user. Figure 4.3 above shows the window for the input of a kinetic reaction rate for the reaction from F6P to F1,6P. In the lower part of the window, the user defines the properties of the variables and parameters. Simulations and parameter fittings require the following definitions and selections from the user:

1. Define which compounds are not constant and should be included in the system of ODEs. In the illustrative example, F6P, FBP, ADP, ATP, and PEP have been included as dynamic variables. Dynamic variables can also be defined with optimized initial conditions which is important if no initial condition is available from experiments or the available data is inaccurate.

2. Define variables for which values from an experimental data file are used during solution of the Initial Value Problem (IVP). This is useful for compounds where good measurements are available but that cannot or should not be included in the model as dependent variables. This could be e.g. the GLC concentration in a fermenter or the ATP concentration in a cell (the ATP concentration is important but is difficult to model due to the large number of reactions producing and consuming ATP). Such input data can also be from interpolation of experimental data. This results in smooth data representations.

3. Define parameters to be fixed or to be subject to parameter fitting.

4. Define initial values for all variables and parameters. Define ranges for optimized initial conditions and fitted parameters.

5. Choose a data file from the table DFILE that has been linked to the modeling group. Experimental data is not stored in the database because the data is constant for all models, simulations, and parameter fittings. Because the numerical algorithms for simulation and parameter fitting require a text file as input it is more efficient to keep this text file always available. Otherwise this text file would have to be produced for each simulation and parameter fitting.

6. Choose an error functional and weights for each compound where data values are available.

7. Define tolerances for solution of the IVP.

8. Choose an algorithm for parameter fitting and choose all settings specific for this parameter fitting algorithm.

9. Choose a method for sensitivity calculation that also allows to reduce the parameter space to the most sensitive parameters.

10. Define the number of parameter fitting runs that should be performed. One run ends if the maximal number of function evaluations or the tolerance of the fit is reached. The next run starts with randomly changed parameters. The user needs to define the amount of random change of these parameter changes.

11. Choose frequency of output: after every parameter fitting, only if an improvement was made, or only at the end of all parameter fitting. These options are available separately for the fitted variables and the error functional, the dependent variables versus time, and the fluxes versus time.

4.5. Automated model generation by the MMT software

Before the first simulation or parameter fitting with a model, the system of ODEs for the IVP has to be setup. This setup is performed by the MMT software. Several steps are necessary that are described here.

4.5.1. Extract information from the database to setup the IVP

The structure of the database system is described at the beginning of this chapter. In order to setup an IVP the MMT software performs several tasks:

1. It analyzes each compound that is part of the chemical reaction network. If a compound is defined as a dynamic variable, an ODE has to be defined. This ODE consists of several terms. In the illustrative example, the following terms have to be included in the ODE for the F6P concentration:

$$\dot{\text{F6P}} = v_{\text{influx, F6P}} - v_{\text{F6P}\leftrightarrow\text{FBP}}$$

2. After the ODEs for all compounds are defined, the fluxes v_i have to be replaced by kinetic rate equations, e.g. Michaelis-Menten equations. For the

F6P concentration in the illustrative example, this will result in the following equation:

$$\dot{\text{F6P}} = v_{\text{influx F6P}} - v_{max,v_2} \cdot \frac{\text{F6P}}{K_{m,v_2} + \text{F6P}} + v_{max,v_3} \cdot \frac{\text{FBP}}{K_{m,v_3} + \text{FBP}}$$

3. After the system of ODEs is setup, all variable and parameter names are renumbered. Compound concentrations are replaced by the components of a vector $\mathbf{y}$, parameters by the components of a vector $\mathbf{P}$. Then the above equation for F6P could read (vector indices chosen arbitrarily):

$$\dot{\mathbf{y}}_1 = \mathbf{P}_1 - \frac{\mathbf{P}_2 \cdot \mathbf{y}_1}{\mathbf{P}_3 + \mathbf{y}_1} + \frac{\mathbf{P}_4 \cdot \mathbf{y}_2}{\mathbf{P}_5 + \mathbf{y}_2}$$

   This step is necessary to avoid duplicate names in the systems of ODEs that belong to different kinetic rate equations. Because the equations stored in the table QUANT are already in MAPLE 6 format, the replacement is the only necessary step.

4. Create output files in text format: the next step is to write the IVP to a text file. This text file is kept in MAPLE notation. Therefore, MAPLE can be used to convert the IVP into a C-code. Additionally, several text files are produced with initial values and parameters for the simulation and parameter fitting algorithms.

5. The created C-code is linked into a simulator and optimizer that is described in the next section. A program function results that can be used for simulations and parameter fittings.

For the calculation of sensitivities or to use the gradient method for parameter fitting, each differential equation has to be derived with respect to the initial conditions and parameters it depends on (see Section 5.2.3). This process is much more complicated than setting up the IVP. To create the derivatives, a temporary table is setup that includes all dependencies of the ODEs on parameters and initial conditions. Then the derivatives required can be written in MAPLE format to a text file. MAPLE is then used to explicitly calculate these derivatives. Figure 4.12 and Figure 4.13 show examples of the files created by the MMT and passed to MAPLE.

The derivatives of ODEs are converted by MAPLE into a C-code. For the illustrative example the ODE for $\mathbf{y}_1$ has to be derived with respect to $\mathbf{P}_1$ to $\mathbf{P}_5$ and with respect to the initial conditions of $\mathbf{y}_1$ and $\mathbf{y}_2$. MAPLE will have to calculate 7 derivatives just for the first ODE. This shows that the C-code becomes very large if derivatives are calculated explicitly. More details about the setup of sensitivity equations are discussed in Section 5.2.3.

```
with(codegen,C):
f := [                                        v1
  - ( 1.00000000000000 * ( ( p[184] * y[1]/ p[186] - p[187] * y[3]/
p[189] ) / ( 1 + y[1]/ p[186] + y[3]/ p[189] ) ) ) - (
1.00000000000000 * ( p[556] * ( y[1] * y[21] - y[25] * y[22]/ p[561]
) / ( y[1] * y[21]                                  * ( 1 + y[25]/ p[562] )
+ p[563] * ( y[1] + p[564] ) + p[565] * y[21] + p[556] / ( p[566] *
p[561] ) * ( p[567] * y[25] * ( 1 + y[1]/ p[564] ) + y[22] * ( p[568]
* ( 1 + p[565] * y[21]/ ( p[564] * p[563] ) ) + y[25] * ( 1 + y[21]/
p[569] ) ) ) ) ) ) + ( 1.00000000000000 * ( p[846] / ( 1 + y[7]/
p[848] )^ p[849] * ( p[850] * z[27] * y[2] - p[853] * y[7] * y[1] ) /
( 1 + p[855] * z[27] + p[856] * y[2] + p[857] * y[2] * z[27] + p[858]
* y[1] + p[859] * y[7] + p[860] * y[1] * z[27] ) ) )
,
                                              v2
  + ( 1.00000000000000 * ( ( p[49] * y[11]/ p[51] - p[52] * y[2]/
p[54] ) / ( 1 + y[11]/ p[51] + y[2]/ p[54] ) ) ) - ( 1.00000000000000
* ( p[55] * ( y[2] * y[9] - y[7] * y[8]/ p[60] ) / ( y[2] * y[9]
* ( 1 + y[7]/ p[61] ) + p[62] * ( y[2] + p[63] ) + p[64] * y[9] +
p[55] / ( p[65] * p[60] ) * ( p[66] * y[7] * ( 1 + y[2]/ p[63] ) +
y[8] * ( p[67] * ( 1 + p[64] * y[9]/ ( p[63] * p[62] ) ) + y[7] * ( 1
+ y[9]/ p[68] ) ) ) ) ) ) + ( 1.00000000000000 * ( p[444] * ( y[8] *
y[7] - y[2] * y[20]/ p[449] ) / ( y[8] * y[7]
* ( 1 + y[2]/ p[450] ) + p[451] * ( y[8] + p[452] ) + p[453] * y[7] +
p[444] / ( p[454] * p[449] ) * ( p[455] * y[2] * ( 1 + y[8]/ p[452] )
+ y[20] * ( p[456] * ( 1 + p[453] * y[7]/ ( p[452] * p[451] ) ) +
y[2] * ( 1 + y[7]/ p[457] ) ) ) ) ) ) + ( 0.00000000000000 * ( p[693]
* y[8]/ ( y[8] + p[695] * ( 1 + y[9]/ p[697] ) ) * y[3]/ ( y[3] +
p[699] * ( 1 + y[8]/ p[700] + y[9]/ p[701] + y[20]/ p[703] ) / ( 1 +
y[9]/ p[704] + y[20]/ p[705] ) ) * 1 / ( 1 + ( 1 + y[2]/ p[707] )^4 *
p[708] / ( 1 + y[3]/ ( p[699] * ( 1 + y[8]/ p[700] + y[9]/ p[701] +
y[20]/ p[703] ) / ( 1 + y[9]/ p[704] + y[20]/ p[705] ) ) )^8 ) ) ) -
( 1.00000000000000 * ( ( p[752] + p[753] * y[12] + p[755] * y[4] +
p[757] * y[12] * y[4] ) / ( 1 + p[758] * y[12] + p[759] * y[4] ) *
p[760] * y[2]/ ( p[762] + y[2] ) ) ) + ( 1.00000000000000 * ( ( 1 - (
y[8] - p[764] ) / p[765] * p[766] * y[8] ) * ( p[767] * y[8] * y[13]
- p[769] * y[9] * y[2] ) / ( 1 + p[772] * y[8] + p[773] * y[13] +
p[774] * y[8] * y[13] + p[775] * y[2] + p[776] * y[9] + p[777] * y[2]
* y[9] ) ) ) - ( 1.00000000000000 * ( p[846] / ( 1 + y[7]/ p[848] )^
p[849] * ( p[850] * z[27] * y[2] - p[853] * y[7] * y[1] ) / ( 1 +
p[855] * z[27] + p[856] * y[2] + p[857] * y[2] * z[27] + p[858] *
y[1] + p[859] * y[7] + p[860] * y[1] * z[27] ) ) )
,
...
];
```

Figure 4.12.: Example of a MAPLE file with the fluxes created by the MMT. The functions $\mathbf{v}_1$ and $\mathbf{v}_2$ are shown. The total model dimension of the example was 26.

```
ydot[27] := + diff(f[1],y[1]) * y[28];
ydot[29] := + diff(f[1],y[1]) * y[29] + diff(f[1],p[184]) ;

...

ydot[283] := + diff(f[7],y[1]) * y[53] + diff(f[7],p[860]) ;
```

```
C([dy[1]=ynew[1]],
filename="/home/hurlebau/ModellingOptimize/modgroup20/model19/opti2/odePl
ainExt.c");
C([dy[2]=ynew[2]],
filename="/home/hurlebau/ModellingOptimize/modgroup20/model19/opti2/odePl
ainExt.c");

...

C([dy[283]=ynew[283]],
filename="/home/hurlebau/ModellingOptimize/modgroup20/model19/opti2/odePl
ainExt.c");
```

Figure 4.13.: Example of a MAPLE file with command lines for calculation of sensitivities (top). All rates $\mathbf{v}_i$ (called $\mathbf{f}_i$ in the graph) need to be derived with respect to initial conditions and parameters ($\mathbf{y}_0$ to $\mathbf{y}_{25}$ are dynamic variables, $\mathbf{y}_{26}$ to $\mathbf{y}_{314}$ are parameters). $\mathbf{ydot}_{27}$ to $\mathbf{ydot}_{283}$ are the additional ODEs for the sensitivity functions. The bottom of the figure shows examples of MAPLE commands that are used to write the system of ODEs to files.

At the end of simulations and parameter fittings, the MMT software outputs statistics about parameter changes during parameter fitting. Additionally, the solutions of the IVP can be viewed. The MMT software produces graphs showing compound concentrations versus time or kinetic reaction rates versus time. In the illustrative example five graphs are produced, one for each compound F6P, FBP, ADP, ATP, and PEP. The graphs will show the concentrations of these compounds versus time. Additionally, five graphs are produced that show the values of $\mathbf{v}_1$ to $\mathbf{v}_5$ versus time. Thus, it is possible to examine how the fluxes between compounds change over time. The produced graphs are in GIF or POSTSCRIPT format and can be used in other programs e.g. to write reports or summarize results.

Performane issues

The process of setting up the system of ODEs from the information stored in the database and the compiling, linking, and creation of the initial files can take several minutes for a system of 30 ODEs and 300 parameters. The most time consuming process is the setup of the ODE system, i.e. the extraction of the information from the tables of the relational database. However, if the number of SQL statements is reduced by using additional temporary tables, this process can be accelerated significantly.

For simulations or parameter fittings with the same model, the time consuming setup and creation of the C-code is not required anymore. Thus, the initializing time is reduced to a few seconds. Only after choosing a different model for simulation or parameter fitting, the C-code has to be recreated. Future versions of MMT will keep the C-codes belonging to a specific parameter identification in place in order to directly compile and link it to the simulation and parameter identification code. This requires to take care that the C-code on disk is really in correct form, i.e. if changes in the model have been made, the C-code has to be updated. If the model did not change, however, usage of an already available C-code can save several minutes computation time.

If the C-codes are kept in place, it will also be easier to reproduce simulation results with parameters saved in the database. As argued before, this would reduce the amount of data stored in the database by approximately 90%.

4.6. The MMT simulator and optimizer

The complete simulation and parameter fitting procedure of the MMT software is performed by a C procedure. Figure 4.14 shows the structure of the MMT simulator and optimizer. The main procedure is started from the MMT user interface. Once started, the software runs independently of the user interface. Input is provided from text files and the output is written to such. The MMT user interface, however, keeps track of the progress of the simulation or parameter fitting. After recognizing the termination of a simulation or parameter fitting, it reads the data files that contain the results and fits them into the MMT database. The simulator and optimizer maintains the following features:

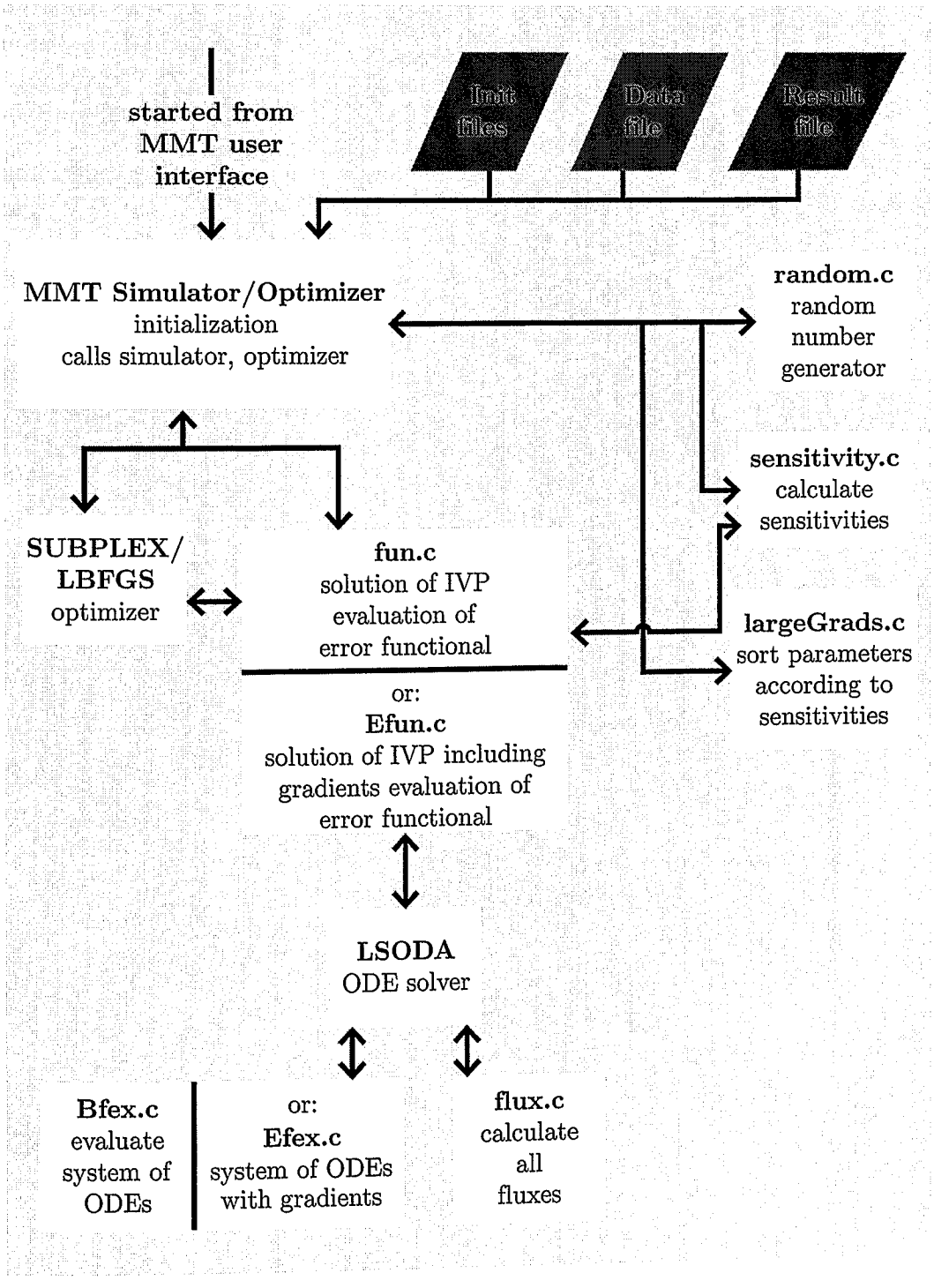


Figure 4.14.: The structure of the MMT simulator and optimizer. Arrows indicate data flow from the external files or function calls and returns. If explicit gradients are needed, `Efex.c` replaces `Bfex.c` and `Efun.c` replaces `fun.c`.

1. **Initialization of parameters:** All initial parameters and the data file are read and stored in vectors and arrays. Additionally, a significant number of parameters for the simulation and parameter fitting algorithm have to be read from external files. Global variables are setup that can be used in the subroutines.

2. **Do an initial simulation:** An initial simulation is done by calling the subroutine `fun`. `fun` calls the Fortran algorithm `LSODA` for solution of the IVP. Also, fun calculates the error functional if a data file was specified for the simulation. If only a simulation should be performed, the results are written to the result file and the program terminates.

3. **Initialize the parameter fitting:** Depending on the user choice, sensitivities of all parameters are calculated and parameters with low sensitivities are excluded from parameter fitting. The sensitivity calculation is performed by the C procedure `sensitivity`. `sensitivity` calls `fun` for solution of the IVP, calculation of fluxes and the error functional. Sorting parameters according to their sensitivities is done by the procedure `largeGrads`.

4. **Start parameter fitting loop:** After all preliminary tasks are finished, the Fortran parameter fitting routine `SUBPLEX` or `LBFGS` is called. These routines call `fun` for calculation of the error functional which they are trying to minimize. Figure 4.15 shows a flowchart for the function `fun`.

5. **Result and termination check:** After the parameter fitting algorithm ends it returns the identified parameter values and the minimal value of the error functional to the main procedure. The main procedure compares results with previously identified results and writes them to a temporary result file. The result file is analyzed be the MMT software and the information stored in the database. The main procedure also checks if the maximum number of restarts is reached. If yes, final results are written and the program exits.

6. **Restart of the parameter fitting algorithm:** The parameter fitting algorithm will be called repeatedly until the maximal number of parameter fittings is reached. According to strategy parameters set by the user, parameters will be changed randomly with a normal distribution over a given range before the parameter fitting restarts. This procedure is important to identify the global minimum of the error function. The algorithm now continues at point 4 with the start of the parameter fitting algorithm.

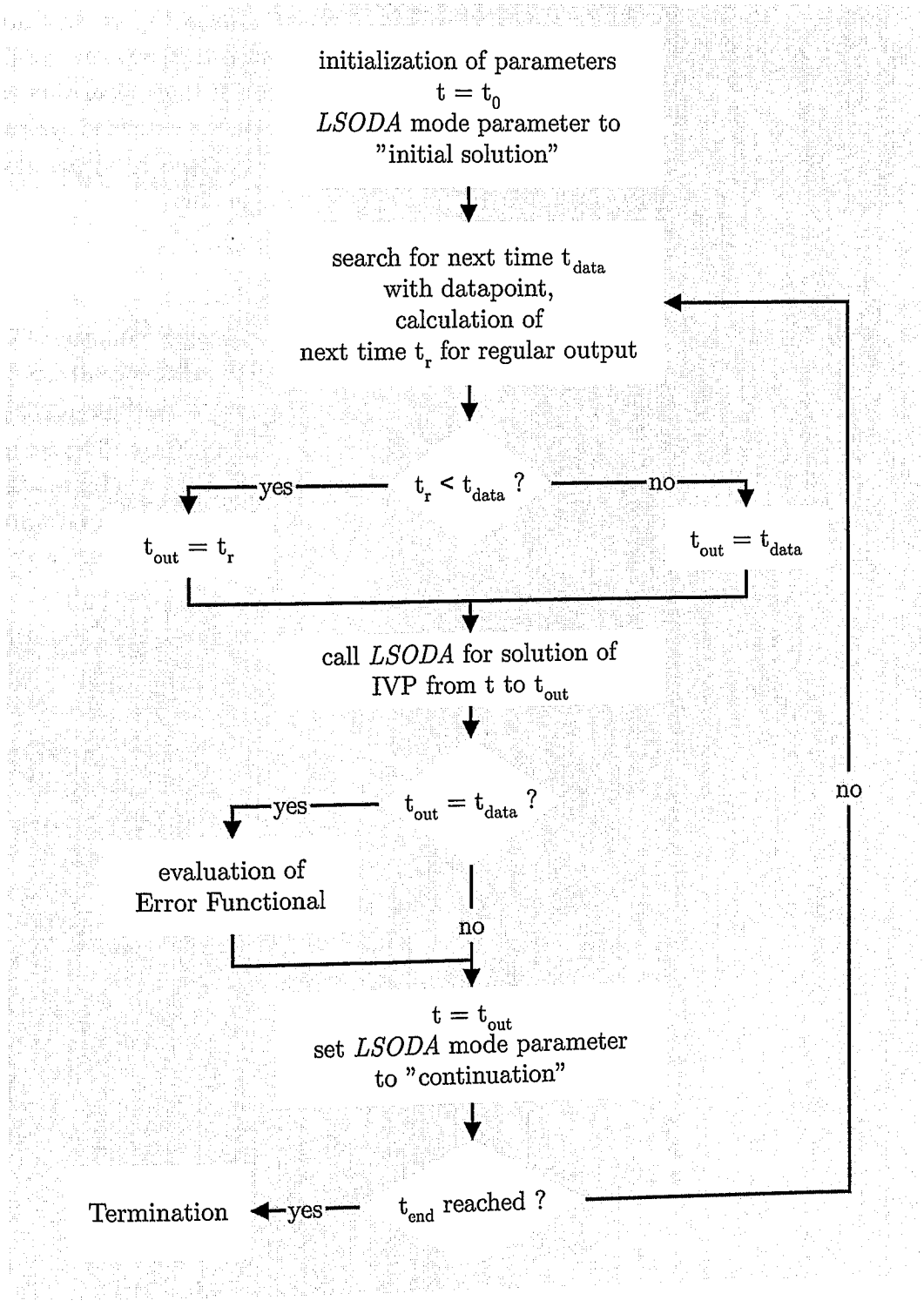


Figure 4.15.: Flowchart for the C-routine fun for solving the IVP. Regular output times are set by the user. Solution is always performed to the next output time (which is either regular output time or a time with a data point). After the first call, LSODA is called with the continuation mode in order to increase performance and accuracy.

Notice that the routines `Efex` and `Efun` that are shown also in Figure 4.14 are replacements for `fex` and `fun`. These replacements are needed if explicit gradient information is needed for an parameter fitting (parameter fitting) algorithm as `LBFGS`, or if explicit sensitivities should be calculated by solving the extended system of ODEs (variation of parameters). As mentioned above, `Efex` and `Efun` are also created from the MMT user interface and coexist with `fex` and `fun`.

4.7. Summary of the MMT software

In this chapter, the MMT software has been described. The tool can be used for model setup and model analysis with simulations and parameter fittings. The setup was created in the framework of a relational database system that combines efficiency and consistency. Thus, models can be build and analyzed quickly, results are reproducible, and prior results can be reused in new models.

5. Numerical Algorithms

This chapter describes the numerical treatment of the system of ODEs that results from the modeling approach and model generation presented in Chapters 3 and 4. Then, methods for applying the developed models to

- simulate chemical reaction networks
- reproduce experimental data sets from experiments described in Section 2.1.

are shown. An easy example consisting of a single ODE shows the problems that occur during fitting of a model to experimental data.

Simulations require the solution of an IVP by numerical algorithms. Reproduction of experimental data sets is based on finding parameters for the model such that the model fits to the data. This is a parameter fitting or optimization task that has to be solved. Figure 5.1 summarizes the general approach of the numerical treatment.

Recall of the Model Properties

The structure of the models developed in this work have been described in Chapter 3. Recall Equation 3.5, which formulates the general non-stationary problem:

$$\dot{\mathbf{C}} = \mathbf{N} \cdot \mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{P})$$

As mentioned, the reaction rates $\mathbf{v}_i$ are usually of non-linear form. The system of ODEs is high dimensional and, because the chemical reactions operate on different time scales, the system is often stiff.

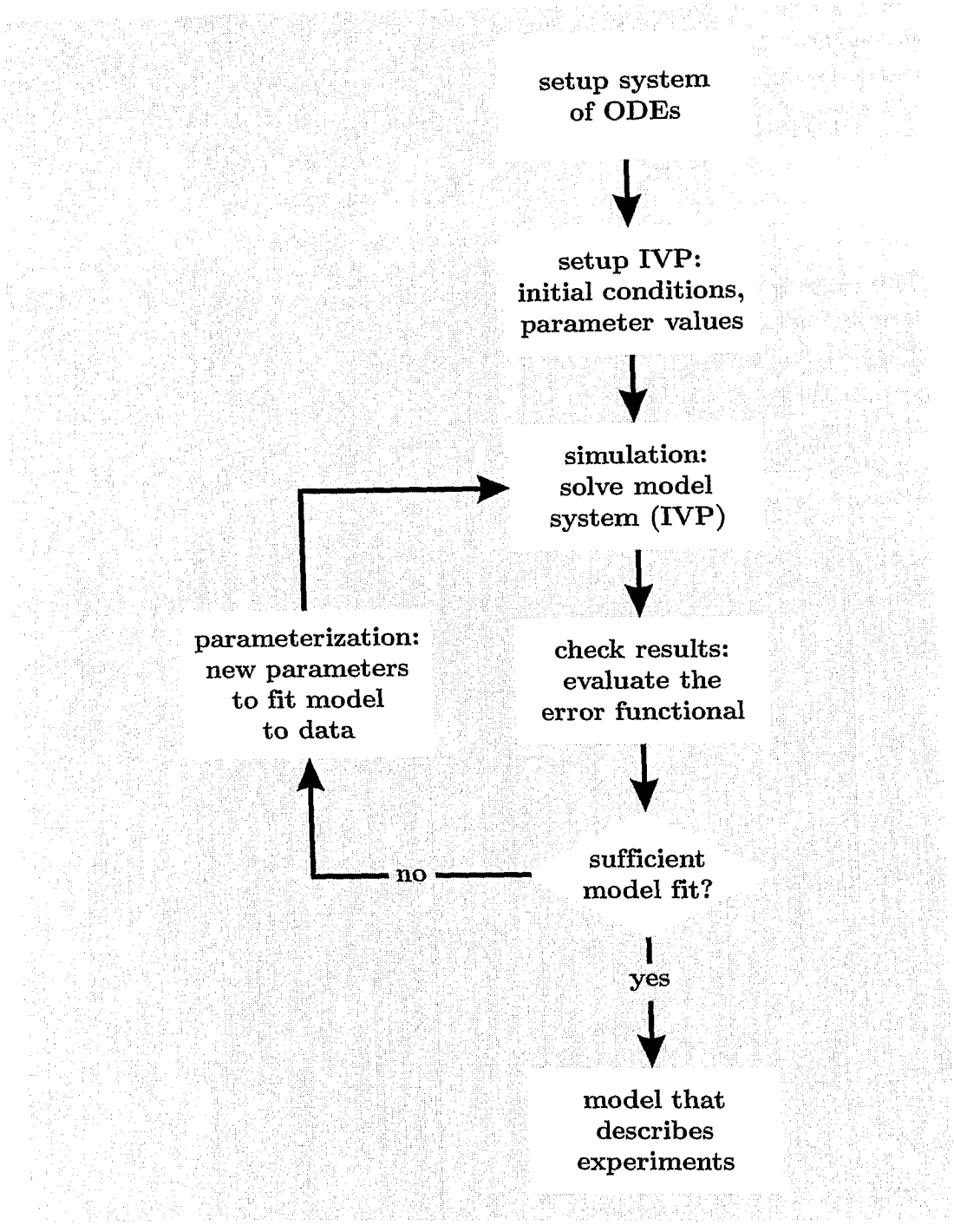


Figure 5.1.: General procedure for setup of the mathematical model, simulation, and fitting to data using numerical algorithms. Evaluation of the error functional means calculating a difference between model output and experimental data set as described in the text.

Additional Notations for this Chapter

In Equation 3.5 the vector $\mathbf{v}$ of reaction rates is written in general form and depends on intracellular and extracellular compound concentrations, enzyme concentrations, and parameters.

From now on $\mathbf{S}$ and $\mathbf{E}$ will be omitted because the enzyme concentrations are not used in the presented models and the only extracellular concentration that is part of the models is the GLC concentration. This concentration is fixed and changes its value only at the time when the substrate pulse is given. Therefore the GLC concentration will be considered as an entry in the parameter vector $\mathbf{P}$. At time $t = 0$, the initial conditions are written as $\mathbf{C}_0 := \mathbf{C}(0)$ and $\mathbf{C} = \mathbf{C}(\mathbf{C}_0, \mathbf{P}, t)$ is the vector of compound concentrations at time t that depends on the initial conditions and the parameters.

The total differential operator of a function f with respect to x will be denoted in the following by

$$D_{\mathbf{x}} f := \frac{\partial}{\partial \mathbf{x}} f \tag{5.1}$$

to distinguish it from the partial derivative $\frac{\partial}{\partial x}$.

To illustrate the difference between the total differential operator and the partial derivative, the following functions are defined:

$$\begin{aligned} f(x,y,t) &= x^2 + y^2 + t \\ x(t) &= sin\, t \\ y(t) &= cos\, t \end{aligned} \tag{5.2}$$

If the partial derivative of $f(x,y,t)$ with respect to x is calculated, all other variables are considered as constants. The same principle is used for the partial derivatives with respect to y and t. Therefore

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x \\ \frac{\partial f}{\partial y} &= 2y \end{aligned}$$

$$\frac{\partial f}{\partial t} = 1 \tag{5.3}$$

In contrast to the partial derivatives calculation of the total differential operator D_x does not assume y and t to be constants. Instead, y and t are treated as functions of t. Calculation of D_x and D_y yields the same result as the partial derivatives. Calculation of D_t yields a different result because x and y are functions of t:

$$\begin{aligned} D_x &= \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial x} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial x} + \frac{\partial f}{\partial t}\cdot\frac{\partial t}{\partial x} = 2x + 0 + 0 = 2x \\ D_y &= \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial y} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial y} + \frac{\partial f}{\partial t}\cdot\frac{\partial t}{\partial y} = 0 + 2y + 0 = 2y \\ D_t &= \frac{\partial f}{\partial x}\cdot\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial t} + \frac{\partial f}{\partial t}\cdot\frac{\partial t}{\partial t} = 2cost - 2sint + 1 \end{aligned} \tag{5.4}$$

The experimental data at time t_i will be modeled by a measure equation

$$\mathbf{y}_i = \mathbf{M} \cdot \mathbf{C}(t_i) \tag{5.5}$$

where $\mathbf{C}(t_i)$ are the compound concentrations at time t_i calculated by solving the IVP and $\mathbf{M}$ is a measure matrix. This matrix is necessary because usually experimental data is not available for all compound concentrations $\mathbf{C}_i$. $\mathbf{M}$ is setup such that $\mathbf{y}_i$ contains only non-zero values for compound concentrations with available experimental data. Also, if the vector of measurements at time t_i is denoted by $\tilde{\mathbf{y}}_i$ let

$$\tilde{\mathbf{y}}_i = \mathbf{M} \cdot \mathbf{C}(t_i) + \varepsilon_\mathbf{i} \tag{5.6}$$

where $\varepsilon_\mathbf{i}$ is the vector of differences between the calculated values $\mathbf{C}(t_i)$ and the measurements $\tilde{\mathbf{y}}_i$. The matrix of all measurements is denoted by $\tilde{\mathbf{Y}}$.

In the illustrative example, suppose measurements exist only for F6P, FBP and PEP. Then

$$\mathbf{y}_i = \mathbf{M} \cdot \mathbf{C}(t_i) \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \begin{pmatrix} F6P(t_i) \\ FBP(t_i) \\ ADP(t_i) \\ ATP(t_i) \\ PEP(t_i) \end{pmatrix} = \begin{pmatrix} F6P(t_i) \\ FBP(t_i) \\ \cdot \\ \cdot \\ PEP(t_i) \end{pmatrix}$$

Suppose that for a time t_1 the following measurement vector $\tilde{\mathbf{y}}_1$ exists:

$$\tilde{\mathbf{y}}_1 = \begin{pmatrix} 1.11 \\ 0.24 \\ \cdot \\ \cdot \\ 3.26 \end{pmatrix}$$

then

$$\tilde{\mathbf{y}}_1 = \mathbf{M} \cdot \mathbf{C}(t_1) + \varepsilon_1 = \begin{pmatrix} F6P(t_1) \\ FBP(t_1) \\ \cdot \\ \cdot \\ PEP(t_1) \end{pmatrix} + \varepsilon_1 = \begin{pmatrix} 1.11 \\ 0.24 \\ \cdot \\ \cdot \\ 3.26 \end{pmatrix}$$

or for the difference between calculated values and experimental data

$$\varepsilon_1 = \begin{pmatrix} 1.11 \\ 0.24 \\ \cdot \\ \cdot \\ 3.26 \end{pmatrix} - \begin{pmatrix} F6P(t_1) \\ FBP(t_1) \\ \cdot \\ \cdot \\ PEP(t_1) \end{pmatrix}$$

5.1. Solutions of the system of ODEs

The non-linear structure of the system of ODEs of the model makes an analytical solution impossible. Therefore, numerical solution techniques have to be used.

All important points for successful and accurate numerical solution of the model equations are discussed in this section.

5.1.1. Properties of rate equations

All rate equations that have been used are smooth functions. If parameters are set in a "reasonable" range, no singularities are possible. "Reasonable" means that the parameter is in a range that can be justified from biological interpretation. It is still possible that during a simulation a floating point overflow occurs because a substrate concentration increases unlimited. This indicates, however, that the parameter values are setup incorrectly or the model is incorrect. Such unrealistic results can therefore be excluded. In addition to these properties, all functions used to describe reaction rates are at least twice differentiable.

For the simulation of the metabolic network, initial compound concentrations are given. Therefore an IVP has to be solved. Existence and uniqueness of a solution with the above mentioned properties of the rate equations follows from the *Fundamental Existence-Uniqueness Theorem*, see e.g. page 74 in [70].

5.1.2. Accuracy of solutions

An important point when solving IVPs with a numerical algorithm on a computer is the accuracy of the solution. Chemical reaction systems often consist of several reactions that operate on a different time scale. If the time scales differ by many orders of magnitude, the system is called a *system of stiff ODEs.* There are, however, no exact mathematical methods on how to classify a system of ODEs as stiff. The historically first opinion is from Curtiss & Hirschfelder in [24]: "Stiff equations are equations where certain implicit methods perform better, usually tremendously better, than explicit ones." Quantities that give information about the stiffness of a system are the Jacobian $\partial f/\partial y$, the dimension of the system, smoothness of the solution, or the integration interval. A detailed discussion about these measures and solutions of stiff systems can be found in [39]. In [39] a comparison of numerical solution algorithms for stiff problems on a set of test problems can also be found.

For solution of the system of ODEs, a publicly available solver `LSODA` [71] was chosen. This algorithm is a variant of the `LSODE` solver (*Livermore Solver* in [45]) that includes also a solution algorithm for non-stiff problems and switches automatically between the two methods. This results in a better performance of the algorithm. The `LSODE` algorithm has been examined and tested during the above mentioned

comparison of solution algorithms in [39]. Results of these tests indicate that LSODE is a very fast code that works accurately for tolerances of $< 10^{-6}$ but might have some accuracy problems for tolerances $> 10^{-6}$.

During usage of LSODA, some accuracy problems occurred. If the error tolerance was chosen larger than 10^{-6}, sometimes negative values resulted for some of the compound concentrations. However, reaction rates $\mathbf{v}_i$ are always chosen to decrease to zero if the compound concentration approaches zero, i.e.

$$\lim_{C \to 0} \mathbf{v}_i(C) = 0$$

for a compound concentration C. Thus negative values result from numerical inaccuracy. These numerical problems could usually be avoided by decreasing the tolerance to values between 10^{-8} and 10^{-12}.

5.1.3. Intermediate time steps

A simulation with an ODE solver from an initial t_0 to a final time t_1 is straightforward if only initial conditions are used as input variables. However, sometimes the solution has to be divided into several time periods. This happens if

- intermediate results are needed, e.g. for comparison with experimental data.
- parameters are not constant with time. This problem occurs for the extracellular GLC concentration that changes in the experiments when the pulse is given. The same problem occurs if a compound is not included in the IVP as a dynamic variable but experimental data is used. Then for all times where an experimental data value is available, this value has to be passed to the ODE solver.

With the algorithm LSODA intermediate time steps do not cause any problems. The algorithm can be stopped at intermediate times. The algorithm then returns to the routine it was called from. Intermediate results can now be stored and parameters can be changed for the next time period. For reasons of performance and accuracy it is important to call LSODA with the 'continuation' option. In this case, the algorithm continues from the time it was stopped using results from the prior time step. This is possible because the results are stored in global variables and are therefore still stored in the memory of the machine.

5.1.4. Smoothing of the experimental data

During simulations experimental data values can be used for some compound concentrations. This is a useful approach especially for compounds with inaccurate measurements or with unknown reaction kinetics. E.g. a correct description of the ATP concentration would have to include several hundred reactions where ATP is produced or consumed. Instead, experimental data values can be used as an input to the simulation algorithm.

In order to remove noise in the experimental data, a polynomial approximation was performed to obtain a function that represents the data in a smooth way. Because polynomial approximations are likely to show oscillations it was important to manually inspect the approximations. See Figure 5.2 for an example of the approximation of the G6P and PEP concentrations.

The problems resulting from polynomial approximations of higher order, mainly the oscillating behavior are well known [108]. A much better approach is the approximation with splines, see also [108]. Notice, however, that even a spline approximation is difficult for the given data due to the dynamics present in some measurements. A closer look at Figure 5.2 shows the problem: in the left figure for G6P concentration at time 10s a few measurements are far from the other measurements. A smoothing spline with high accuray would include these data points which would result in an oscillatory fashioned graph. A lower accuracy of the spline would not pass through these points and provide an approximation similar to the one shown. Because estimation of the measurement error is difficult, it is not clear which would be the more realistic approximation.

The obtained approximation was also used during parameter fitting. Because a larger number of values results for fitting the parameters, it can be hoped that an optimal solution can be identified faster.

5.1.5. Summary of the solution algorithm for the IVP

With the `LSODA` algorithm, the given IVP is solvable fast and with good accuracy, taking into account the stiffness and non-linearity of the problem. Computation time for a system of 26 ODEs with 302 parameters, a time period of 40 seconds with an accuracy of 10^{-10} is approximately 1 second. If sensitivities are calculated, computation time increases significantly. The number of sensitivity functions to be evaluted depends on the number of parameters and initial conditions and on the function dependencies (a sensitivity has to be calculated for each parameter

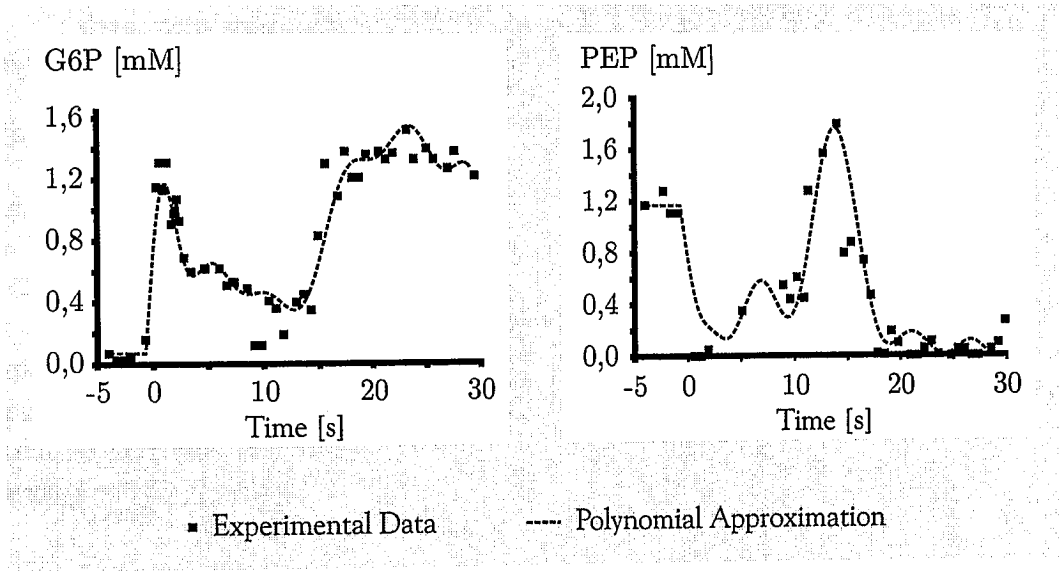


Figure 5.2.: Example of approximation function for the experimental data of G6P and PEP.

and initial condition for each function they are part of). In the above mentioned example, solution of the system of ODEs inclduing all sensitivity functions requires approximately 10 seconds time. The computation time increases with the number of intermediate times t_i for which results are to be stored. The user defines these intermediate time steps.

Results obtained with `LSODA` were compared with solutions from the commercial software tool MAPLE in order to check correct setup of the code. No differences could be seen. The solver was used for all solutions of the IVP. Choosing the tolerance parameter correctly is important. A large tolerance significantly increases efficiency of the algorithm, however, final results should be checked with a small tolerance to identify possible inaccuracy problems.

5.2. Sensitivity analysis

In the last section it has been argued why the solution of the given IVP with specific numerical algorithms is not a problem. This assumes, however, that the following requirements are fulfilled:

1. Requirement: A set of initial conditions (i.e. compound concentrations at an

initial time t_0) is given

2. **Requirement:** A set of initial parameters (i.e. all parameters occurring in the enzymatic reaction kinetics, e.g. Michaelis-Menten 3.3) is given

Under these conditions, a solution can be calculated for a user defined time frame from initial t_0 to final time t_1 (see also Figure 5.1). This also allows simulation of system behavior under different conditions.

Requirement 1 is partly fulfilled with the available experimental data. Other compounds that have not been measured and are still included as dynamical variables in the model need to be estimated. Requirement 2 usually can only rely on data from *in vitro* measurements. These values vary in a broad range. Therefore, three sources of uncertainty result for a given model structure:

1. Initial concentrations of compounds that have not been analyzed.
2. Parameter values for the enzyme kinetic equations.
3. Initial concentrations of compounds that have been measured but with low accuracy.

The relevance of these uncertainties depends on the sensitivities of an initial condition or a parameter value on the simulation result. It is therefore important to calculate sensitivities for each compound and parameter. Then the relevance of errors in initial compound and parameter guesses and relevance of errors in measurements can be estimated.

5.2.1. Interpretation and usage of sensitivities

Sensitivity calculations can be useful and important in many respects. Parameters with low sensitivities are difficult or impossible to be identified. Sensitivity information can be used for

Model Setup: Parameters with low sensitivities can be eliminated from the system.

Experimental Design: Experiments can be planned that allow better identifiability of parameters.

Data Fitting, Parameter Fitting: During fitting of a model to experimental data, sensitivity information can be used to reduce computation time. This aspect will be discussed in more detail in the following section.

Analysis: Sensitivities are needed for statistical analysis of the fitting results. E.g. calculation of the covariance matrix or control coefficients depends on the sensitivities [67].

Conclusions: Interpretation of results has to be carried out keeping in mind sensitivities of each parameter. Parameters with low sensitivities allow less strong conclusions.

5.2.2. Sensitivities of reaction rates at a fixed time t

Sensitivities can be calculated from the reaction rates $\mathbf{v}_i$ itself. Only changes of these rates are analyzed in dependence on parameters and initial conditions. Evaluation is carried out at a fixed time t.

For calculation,

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{P}} \quad \mathbf{v}(\mathbf{C}, \mathbf{P}) \\ &\frac{\partial}{\partial \mathbf{C}_0} \quad \mathbf{v}(\mathbf{C}_0, \mathbf{P}) \end{aligned} \tag{5.7}$$

have to be calculated and evaluated at time t_0.

Sensitivities calculated in this way show the influence of parameters and initial conditions on the kinetic reaction rates. The calculation is fast, but the sensitivities can change and be completely different at other times.

5.2.3. Variation of parameters

Because the used enzyme kinetic rate equations are differentiable with respect to the initial conditions and parameters, it is possible to setup a variation equation system of ODEs that explicitly contains sensitivities of the system (Variation of parameters). This approach is based on the calculation of the sensitivities

$$D_{\mathbf{P}} \quad \mathbf{C}(\mathbf{C}_0, \mathbf{P}, t)$$

$$D_{\mathbf{C}_0} \quad \mathbf{C}(\mathbf{C}_0, \mathbf{P}, t) \tag{5.8}$$

Calculation of these sensitivities cannot be done directly because the function vector $\mathbf{C}(\mathbf{C}_0.\mathbf{P}, t)$ is not explicitly known. However, it is possible to calculate the sensitivities indirectly. If the functions in Equation 5.8 are derived with respect to t and Schwarz's Rule for exchanging derivatives is used, then

$$\begin{aligned} \frac{d}{dt} D_{\mathbf{P}} \mathbf{C}(\mathbf{C}_0, \mathbf{P}, t) &= D_{\mathbf{P}} \left[\frac{d}{dt} \mathbf{C}(\mathbf{C}, \mathbf{P}, t) \right] \\ &= D_{\mathbf{P}} \left[\mathbf{N} \cdot \mathbf{v}(\mathbf{C}, \mathbf{P}) \right] = \mathbf{N} \cdot \left(\frac{\partial \mathbf{v}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{P}} + \frac{\partial \mathbf{v}}{\partial \mathbf{P}} \right) \end{aligned} \tag{5.9}$$

$$\begin{aligned} \frac{d}{dt} D_{\mathbf{C}_0} \mathbf{C}(\mathbf{C}_0, \mathbf{P}, t) &= D_{\mathbf{C}_0} \left[\frac{d}{dt} \mathbf{C}(\mathbf{C}, \mathbf{P}, t) \right] \\ &= D_{\mathbf{C}_0} \left[\mathbf{N} \cdot \mathbf{v}(\mathbf{C}, \mathbf{P}) \right] = \mathbf{N} \cdot \frac{\partial \mathbf{v}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{C}_0} \end{aligned} \tag{5.10}$$

The total differential operator D is calculated as illustrated in the example above (Equation 5.2). Therefore a system of ODEs for the sensitivities results ($\mathbf{1}$ is the unitary matrix)

$$\frac{d}{dt} D_{\mathbf{P}} \mathbf{C} = \mathbf{N} \cdot \left(\frac{\partial \mathbf{v}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{P}} + \frac{\partial \mathbf{v}}{\partial \mathbf{P}} \right) \quad , \quad \frac{\partial \mathbf{C}}{\partial \mathbf{P}}(\mathbf{0}) = \mathbf{0} \tag{5.11}$$

$$\frac{d}{dt} D_{\mathbf{C}_0} \mathbf{C} = \mathbf{N} \cdot \frac{\partial \mathbf{v}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{C}_0} \quad , \quad \frac{\partial \mathbf{C}}{\partial \mathbf{C}_0}(\mathbf{0}) = \mathbf{1} \tag{5.12}$$

It is now possible to integrate the two additional systems of ODEs that result together with the original system of ODEs to obtain the sensitivity functions. This integration is done numerically due to the complexity of the derived system of ODEs.

To illustrate the concept, consider a single ODE

$$\dot{x} = a \cdot x \qquad x(0) = x_0$$

where a is a constant. The explicit solution of this equations is

$$x(t) = x_0 \cdot e^{at}$$

The sensitivity functions are obtained from

$$\frac{d}{dt}\frac{\partial x}{\partial a} = \left(\frac{\partial \dot{x}}{\partial x}\frac{\partial x}{\partial a} + \frac{\partial \dot{x}}{\partial a}\right) = a \cdot \frac{\partial x}{\partial a} + x \quad , \quad \frac{\partial x}{\partial a}(0) = 0 \tag{5.13}$$

$$\frac{d}{dt}\frac{\partial x}{\partial x_0} = \frac{\partial \dot{x}}{\partial x}\frac{\partial x}{\partial x_0} = a\frac{\partial x}{\partial x_0} \quad , \quad \frac{\partial x}{\partial x_0}(0) = 1 \tag{5.14}$$

Define $y_a := \frac{\partial x}{\partial a}$ and $y_x := \frac{\partial x}{\partial a}$ as the sensitivity functions with respect to a respectively x. Then

$$y_a(x,t) = e^{at} + axt + xt - 1$$

$$y_x(x_0,t) = e^{at}$$

Notice that y_x depends only on a and t. This means that the sensitivity of the solution is independent of x_0. The sensitivity function y_a, on the other hand, depends on a and x_0. This sensitivity increases with increasing x_0 and a.

Calculation of the sensitivity functions is time consuming. If the number of parameters is N_P, an IVP of dimension $N^2 + N\,N_P$ has to be solved. It provides, however, a clearer picture of parameter sensitivities. Sensitivity functions show the influence of a certain parameter at a certain time: The solution of the IVP gives $\frac{\partial \mathbf{C}}{\partial \mathbf{P_0}}(t)$ and $\frac{\partial \mathbf{C}}{\partial \mathbf{C_0}}(t)$. This is of great interest in a dynamic situation, because it can be seen if the sensitivity functions of several parameters or initial conditions are linearly dependent. This allows conclusions about the identifiability of certain parameters with the available data set. In [48] a discussion of sensitivity functions for parameters of a Michaelis-Menten based microbial growth model is presented.

5.3. Error functionals

An error functional F for the difference between experimental data and model output is needed for fitting the model to the experimental data (Parameter fitting). Different error functionals were used for parameter fitting that are summarized in this section.

Recall that $\varepsilon_\mathbf{i} = \mathbf{y}_i - \tilde{\mathbf{y}}_i$ was defined as the difference between the vector of calculated values $\mathbf{y}_i$ and $\tilde{\mathbf{y}}_i$ at time t_i. Then $\varepsilon_{\mathbf{i,j}}$ is the difference between calculated value and measurement of compound j at time t_i and $\mathbf{y}_{i,j}$, $\tilde{\mathbf{y}}_{i,j}$ is the calculated respectively measured value of compound j at time t_i. This notation will be used in this section. Additionally, it is assumed that for each compound a weighting factor

$\mathbf{w}_j$ is defined. It is usually small if a compound can be measured with low accuracy only.

Many error functionals depend on weighted, relative squared differences between simulated values and experimental data values. Several modifications have been used in order to improve performance of the parameter fitting algorithm or to influence the fit of specific compounds. In the following, measures $d_k(\varepsilon_\mathrm{i})$ are defined that calculate differences between calculations and measurements for a specific time t_i. In order to get the overall goodness of fit, the sum of $d_k(\varepsilon_\mathrm{i})$ for all times t_i has to be calculated:

$$F(\mathbf{C}_0, \mathbf{P}) = \sum_{i=1}^{N_t} d_k(\varepsilon_\mathrm{i}) \tag{5.15}$$

where N_t is the number of time steps t_i. Notice that the error functional depends on the vectors of initial conditions $\mathbf{C}_0$ and parameters $\mathbf{P}$. Five different formulations of $d_k(\varepsilon_\mathrm{i})$ are implemented in the MMT software:

1. **Standard powered differences:** The standard measure for differences between calculated values and measurements is a weighted sum of least squares

$$d_1(\varepsilon_\mathrm{i}) := \sum_{j=1}^{N} \mathbf{w}_j \frac{|\varepsilon_{\mathrm{i,j}}|^p}{\tilde{\mathbf{y}}_{i,j}} \tag{5.16}$$

where $p = 2$ and $\mathbf{w}_j$ is a weighting factor for compound $\mathbf{y}_j$ at time t_i. In some parameter fitting runs, P was set to 4 in order to increase the influence of small values. This sometimes yielded slightly better performance of the parameter fitting algorithm. However, statistical theory is limited to $p = 2$.

2. **Relative Powered differences:** Instead of taking the deviation of the simulated value relative to the data value, various different methods have been used:

   - a deviation relative to the absolute value of the model output,

$$d_2(\varepsilon_\mathrm{i}) := \sum_{j=1}^{N} \mathbf{w}_j \frac{|\varepsilon_{\mathrm{i,j}}|^p}{\mathbf{y}_{i,j}} \tag{5.17}$$

     This increases the influence of times t_j where small calculated values coincide with large data values.

   - a deviation relative to the minimum of the absolute value of the model output and the data value, i.e.

$$d_3(\varepsilon_\mathrm{i}) := \sum_{j=1}^{N} \mathbf{w}_j \frac{|\varepsilon_{\mathrm{i,j}}|^p}{min(\mathbf{y}_{i,j}, \tilde{\mathbf{y}}_{i,j})} \tag{5.18}$$

This calculation increases the influence of times t_j where a small model output or a small experimental value are present.

- a deviation relative to the sum of the absolute value of the model output and the data value, i.e.

$$d_4(\varepsilon_i) := \sum_{j=1}^{N} \mathbf{w}_j \frac{|\varepsilon_{i,j}|^p}{\mathbf{y}_{i,j} + \tilde{\mathbf{y}}_{i,j}} \tag{5.19}$$

Calculation of the error functional is according to equation 5.15 with the appropriate definition of d.

3. **Relative difference in Derivatives:** The rate of change of a metabolite concentration $\mathbf{C}_i$ at a time t is given by $\dot{\mathbf{C}}_i(\mathbf{C}, \mathbf{P}, t)$. From the experimental data, a rate of change has been estimated using three measurement points at times t_{i-1}, t_i, and t_{i+1}:

$$\dot{\tilde{\mathbf{y}}}_{i-1,j} = \frac{1}{2}\left[\frac{\tilde{\mathbf{y}}_{i,j} - \tilde{\mathbf{y}}_{i-1,j}}{t_i - t_{i-1}} + \frac{\tilde{\mathbf{y}}_{i+1,j} - \tilde{\mathbf{y}}_{i,j}}{t_{i+1} - t_i}\right] \tag{5.20}$$

The value of this approximated derivative was compared with the value of $\dot{\mathbf{y}}_{i,j} = \dot{\mathbf{C}}_j(\mathbf{C}_0, \mathbf{P}, t_i)$. If no data point for t_{i-1} or t_{i+1} was available, the next points (t_{i-2} or t_{i+2}) have been checked until a data point was found. Now

$$d_5(\dot{\varepsilon}_i) := \sum_{j=1}^{N} \mathbf{w}_j \frac{|\dot{\varepsilon}_{i,j}|^p}{L + |\tilde{\mathbf{y}}_{i,j}|} \tag{5.21}$$

where $\dot{\varepsilon}_i$ is now the difference between the calculated $\dot{\mathbf{y}}_{i,j}$ and the estimated $\tilde{\dot{y}}_{i,j}$ derivative. L is a number introduced to avoid division by zero in the fractional part of the equation. Calculation of this differential operator is problematic for noisy data, even if three points are included in the calculation. The functional d_4 should therefore only be used for smoothed data.

In all cases deviations have been calculated only for times where a data point was available. In the numerical algorithm this was accomplished by setting the weights $\mathbf{w}_j = 0$ for times and metabolites where no data point was available. Notice that d_1 to d_4 can attain large values if one or more $\tilde{\mathbf{y}}_{i,j}$ attain small values. This problem can be overcome by adjusting the weights $\mathbf{w}$ according to the measurement size.

During parameter fitting usually errors of the second type (d_2) have been used, where the powered distance is taken relative to the minimum of calculated value and data value. The third kind of error functional including derivatives (d_5) was only used in combination with one of the other functionals. This was accomplished by adding the error functionals (e.g. $\lambda d_k + (1-\lambda)d_3$ for $1 \leq k \leq 4$ and $0 \leq \lambda \leq 1$).

5.3.1. Sensitivities with an error functional

Sensitivity estimations can be done by checking the influence of a parameter on the value of the error functional. Instead of calculating only the change of one rate equation at the initial time t_0 as described, an overall influence can be calculated. For calculation $D_{\mathbf{P}}F$ and $D_{\mathbf{C}_0}F$ have to be calculated. Such sensitivities depend on the error functional, the data set used, and the time frame considered. Therefore they give more insight in the global influence of parameters than sensitivities that are calculated at time $t = 0$ only. Particularly the influence on the quality of the model to fit experimental data is visible.

5.4. Parameter fitting

The parameter fitting problem of this work is described as follows:

Find a parameter vector $\mathbf{P}$ *such that the distance between the vector function of the compound concentrations* $\mathbf{C}(\mathbf{C}_0, \mathbf{P}, t)$ *and the matrix of experimental data points* $\tilde{\mathbf{Y}}$ *is minimal, i.e. find* $\min_{\mathbf{C}_0,\mathbf{P}}(F(\mathbf{C}_0, \mathbf{P}))$ *where* $\mathbf{C}_0$ *and* $\mathbf{P}$ *are variable in a certain range.*

Because some compound concentrations may be unknown or measured with a large error only, some initial conditions of the vector $\mathbf{C}_0$ are subject to parameter fitting also. These will be referred to as optimized initial conditions in the following.

Minimizing F for the developed pathway models is difficult because

- lots of parameters (and initial conditions) are subject to parameter fitting
- the number of available data with high accuracy is small
- no good initial values for starting a parameter fitting are available
- due to the non-linearity of the model, local minima are likely to occur that may be identified during parameter fitting

To solve the given problem, two different parameter fitting algorithms have been used that are described in the next section. Additionally, the following strategies have been applied in order to find a good minimum for F:

1. **Define starting points:** Starting points for parameters and optimized initial conditions have been estimated from literature values. Mostly, however, these values resulted from *in vitro* measurements. Additionally, values of these parameters vary in a wide range in literature. Still, literature values are important at least to define a reasonable value for parameters and initial conditions.

2. **Introduce constraints:** For each parameter and optimized initial condition, a range has been defined. It is required that $\mathbf{P}_{i,min} \leq \mathbf{P}_i \leq \mathbf{P}_{i,max}$ for $1 \leq i \leq N_P$ where $\mathbf{P}_{i,min}$ is the minimal value $\mathbf{P}_i$ can obtain and $P_{i,max}$ is the maximum value, N_P is the number of parameters. Also, $\mathbf{C}_{i,min} \leq \mathbf{C}_{0,i} \leq \mathbf{C}_{i,max}$ for all initial conditions subject to parameter fitting. At the beginning, these ranges keep the parameter value close to the literature values, but during longer parameter fittings, constraints can be relaxed to increase the probability of finding the global minimum.

The following sections describe the parameter fitting algorithms used.

5.4.1. Parameter fitting algorithms

The given problem is non-linear and allows no explicit solution of the parameter fitting problem. A variety of numerical methods exists that can be successful for solving such a problem. However, one cannot be sure a priori which method will work best, or which method will fail. Also, computational costs cannot be estimated a priori. Two methods were used for solution of the problem:

1. **Gradient method:** As described in this chapter, explicit derivatives with respect to initial conditions and parameters can be calculated from the system. Therefore, a gradient method can be used for parameter identification [69]. However, gradient methods do not show good performance in finding global minimum if applied to dynamical systems. In particular, the gradient method requires good starting points and is likely to end up in local minima. Results gained from the gradient method were not very promising.

2. **Nelder-Mead simplex method:** A popular method for parameter fitting is the Nelder-Mead simplex algorithm [66]. It is based on a simplex in the n-dimensional parameter space. A simplex is then an $n+1$-dimensional convex hull in this space. The Nelder-Mead algorithm moves this simplex around in the parameter space, changing its size and shape, always calculating and comparing the error functional F. When a minimum is approached, the simplex is reduced in size, and smaller movements of the simplex in the parameter space are performed.

3. **Tom-Rowans SUBPLEX method:** Tom Rowan developed the SUBPLEX parameter fitting method [84] which is based on the Nelder-Mead algorithm. The algorithm uses the Nelder-Mead simplex algorithm on subspaces of the complete parameter space. The purpose of dividing the parameter space into several subspaces is an increased performance of the Nelder-Mead simplex method for lower dimensional subspaces. Subspaces are chosen in the direction of improvement, i.e. initially the algorithm evaluates which parameters are sensitive and are promising for improvement of the fit. After a tolerance is reached during parameter fitting in a subspace another subspace is chosen.

Other parameter fitting algorithms that might be helpful for the problem but have not been tested so far are

1. **Stepwise problem solving:** Dennis et al [26] developed a method for partial parameter identification of a dynamical parameter fitting problem of non-linear ODEs. During the process, the problem is solved for the time period t_0 to t_1, then from t_1 to t_2 and so on. The algorithm is available publicly but has not been tested yet. Therefore it cannot be estimated if the higher computational cost is overcome by a much better performance.

2. **Levenberg-Marquardt algorithm:** This popular method [64] is a well performing algorithm that is also able to obtain an optimial solution if the initial parameters are far from this solution. However, the approach is usually not successfull for identification of a global minimum.

3. **Multiple shooting:** A multiple shooting approach can be used for parameter identification of a system of ODEs. The code PAREST [41] has been evaluated and some first experimental parameter identifications have been performed. The code seems to require good initial parameter estimates for successfull parameter identification which are usually not available.

4. **Genetic algorithms:** For genetic algorithms, parameter vectors are considered as a vector of genes. Genetic vectors can mutate and an evolutionary approach is used to improve the genetic vectors. It cannot be proven that the algorithm performs well on a problem. Collections of genetic algorithms can be found in [54] and [34].

5.4.2. Improving the performance of parameter fitting algorithms

In order to get a good performance of parameter fitting algorithms when searching for global minima, the following methods are frequently used:

1. **Restart the parameter fitting:** Because the algorithms depend on the initial parameter vector $\mathbf{P}_0$, and often end up in a local minimum that depends on this initial values, it is useful to restart the parameter fitting after a local minimum is reached. For the restart, the initial parameter vector $\mathbf{P}_0$ can be changed randomly, where each parameter should remain in its range in order to fulfill the constraints.

2. **Range extensions:** After an parameter fitting is finished, analysis of parameter values can give indications which parameters are close to their upper or lower bounds. A range extension can be performed for parameters that are uncertain from a biological point of view. For other parameters, however, a range extension should not be performed because it would detach the model from its biological meaning.

3. **Parameter space reduction:** The dimension of the parameter fitting problem, i.e. the parameter space can be reduced. As described before, this can be done by using simpler enzyme kinetic equations, or reducing by fixing certain steady state compound concentrations. At the parameter fitting stage, however, additional parameter space reduction can be performed. This can be done by fixing certain parameters where good guesses are available. Also, parameters with low sensitivities can be neglected. The methods implemented in the MMT software are described in Chapter 4.

Another approach proved to be helpful: If parameters were scaled according to their sensitivities, data fitting efficiency sometimes increased significantly. If $\mathbf{P}_{i,0}$ is the initial value of parameter $\mathbf{P}_i$ and the parameter fitting algorithm sets the parameter to a value $\tilde{\mathbf{P}}_i$, then the scaling method sets the parameter to

$$\mathbf{P}_i = (\tilde{\mathbf{P}}_i - \mathbf{P}_{i,0}) \cdot \frac{S_{max}}{S_{P,i}} \tag{5.22}$$

where $S_{P,i}$ is the sensitivity of parameter $\mathbf{P}_i$ and S_{max} is the maximum sensitivity of all values $S_{P,i}$. I.e., the lower a sensitivity is, the more the increase in the difference of the parameter to its initial value $\mathbf{P}_{i,0}$. The idea is that parameters with low sensitivities have to be changed more in order to influence the result.

A problem of the scaled parameters is the fact that some parameters change by a large amount. This can only be justified for parameters where no good literature values exist. Otherwise, the approach of scaled parameters leads to unrealistic results.

5.4.3. Reasons for failing parameter fittings

Parameter fitting of a model can fail for several reasons. Numerical methods as the SUBPLEX algorithm cover the whole parameter range while fitting a model. Therefore if a model cannot be fitted to experimental data after sufficient computation time (which depends on the dimensions of the problem), then the numerical algorithm is most likely not the reason. Instead, one of the three following situations may have occurred:

Underfitting: If a model cannot be parameterized because too many of the parameters have been fixed or the ranges of the parameters have been chosen too small, then the model is underfitted. This problem can be overcome by allowing more parameters to change or allowing parameters to change over larger ranges.

Overfitting: A model is overfitted if parameter fitting does not lead a unique result. i.e. if several parameter vectors $\mathbf{P}$ exist that fit the model to experimental data with the same quality. Overfitting can be a result of too many parameters for the available data or too large ranges of the parameters. The problem can be overcome by fixing parameters with low sensitivities that cannot be identified by the available experimental data. These parameters then have to be estimated using other sources (Literature, other experimental data).

Incorrect Model: If a model cannot be parameterized even though all its parameters are allowed to change without limits, then the model itself is incorrect. The model does not describe the system correctly and changes in the model setup have to be made.

5.4.4. Performance and problems of parameter identifiability: an example

An example is given that shows performance and the problems of parameter identification. The following steps have been carried out and will be discussed in this section:

1. A simple model with one ODE is setup.

2. A simulation is performed with the model. The simulated compound concentrations are modified randomly to add a hypothetical measurement error.

3. A Monte-Carlo simulation is performed in order to parameterize the model to the randomly modified simulation output.

Figure 5.3 shows the fluxes into and out of the intracellular G6P metabolite pool of *Escherichia coli*. The following system was setup:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{v}_{\mathrm{GLC} \rightarrow \mathrm{G6P}} \\ \mathbf{v}_2 &= \mathbf{v}_{\mathrm{G6P} \rightarrow \mathrm{6PGL}} \\ \mathbf{v}_3 &= \mathbf{v}_{\mathrm{G6P} \rightarrow \mathrm{F6P}} \end{aligned} \tag{5.23}$$

$\mathbf{v}_1$ describes the uptake of GLC with the PT-System, $\mathbf{v}_2$ the flux from G6P into the PP-Pathway, and $\mathbf{v}_3$ the flux from G6P into glycolysis. One ODE was setup for the concentration of G6P:

$$\dot{\mathrm{G6P}} = \mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 \tag{5.24}$$

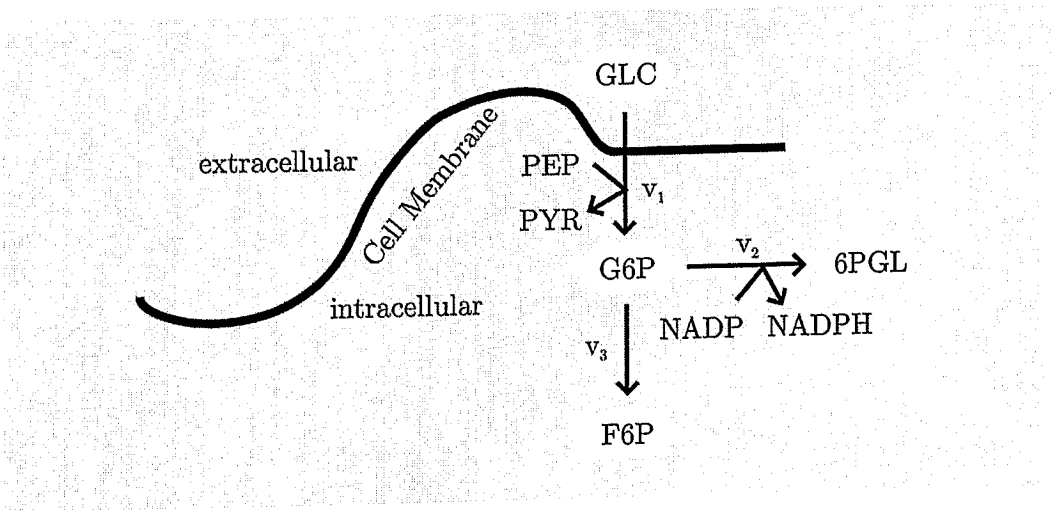


Figure 5.3.: Fluxes in and out of the intracellular metabolite pool G6P as included in the example model.

For $\mathbf{v}_1$ and $\mathbf{v}_3$ a Michaelis-Menten kinetic is assumed, whereas the flux $\mathbf{v}_2$ is assumed to depend proportionally on the G6P concentration. The system then is

$$\dot{\text{G6P}} = v_{max,1} \cdot \frac{\text{GLC}}{K_{m,1} + \text{GLC}} - v_{max,2} \cdot \text{G6P} - v_{max,3} \cdot \frac{\text{G6P}}{K_{m,3} + \text{G6P}} \quad (5.25)$$

This model with three kinetic equations contains five parameters and is of non-linear form. For the specific case of GLC uptake with the PT-System (flux $\mathbf{v}_1$), a molecule of PEP is needed for each GLC molecule taken up. Therefore, a correct representation of the rate equation $r_{\text{GLC}\rightarrow\text{G6P}}$ needs to include the PEP concentration and then is a function of at least two concentrations. This would add more complexity and parameters to the model. For illustration of the parameter fitting problem only the simple model in Equation 5.25 is discussed here.

A simulation with the example model

Concentration of GLC was used as input to the system. It was set in the same way as in the GLC pulse experiments described in Section 2.1:

$$\text{GLC}(t) = \begin{cases} 0.1 \quad \text{mM} & \text{, for } 0.0s \leq t < 5.0s \\ 15.0 \quad \text{mM} & \text{, for } 5.0s \leq t \leq 40.0s \end{cases} \quad (5.26)$$

In order to solve the IVP generated by the model, parameter values of the model have been set to the values shown in Table 5.1. The table also shows the ranges of the parameters that are used in the parameter fitting described in the next paragraph. The parameter values have been chosen to obtain an uptake flux $\mathbf{v}_1 = 0.35$ mM/s of GLC before and $\mathbf{v}_1 = 1.45$ mM/s after the pulse of GLC. These fluxes are in a reasonable range from a biological point of view.

Measurements of intracellular compound concentrations are not exact. In order to account for this effect, the obtained simulated compound concentration of G6P from solution of the IVP is provided with an error. It is assumed that a normally distributed error of 5% per average is present in the measurements. This "noise" was added to the simulated values for G6P. Figure 5.4 shows the exact model output and the G6P concentrations with added noise.

Table 5.1.: Parameter values used for the simulation with the experimental model. Shown are also the ranges of the parameters that are used during parameter fitting.

Parameter	Unit	Initial value	Lower limit	Upper Limit
$v_{max,1}$	mM/s	1.5000	0.1500	15.0000
$K_{m,1}$	mM	0.5000	0.0500	5.0000
$v_{max,2}$	mM/s	0.1500	fixed	fixed
$v_{max,3}$	mM/s	1.3650	0.1365	13.6500
$K_{m,3}$	mM	0.7500	0.0750	7.5000

Monte-Carlo simulation to test parameter fitting

In order to test if the parameters used for generation of the data shown in Figure 5.4 can be identified during parameter fitting, a Monte-Carlo simulation has been performed. The following setup was used:

- $v_{max,1}$, $K_{m,1}$, $v_{max,3}$, and $K_{m,3}$ have been used for parameter fitting. $v_{max,2}$ has been assumed constant. Initial values for the parameters subject to parameter fitting have been chosen randomly by assuming a uniform distribution of the initial parameters over their ranges. The ranges used are shown in Table 5.1.
- As a fitting criteria, the relative least square d_1 was used.
- 25000 parameter fitting runs have been performed with the randomly changed initial parameter values. Each parameter fitting was carried out with the SUBPLEX algorithm until a relative tolerance of 0.0001 was reached.

Performance of the parameter fittings

Each parameter fitting of the model requires several hundreds or thousands of solutions of the IVP. The exact number depends on the initial parameter values. During the Monte-Carlo simulation performed, one parameter fitting could be performed per average in 1.5 seconds on a Pentium III with 800 Mhz.

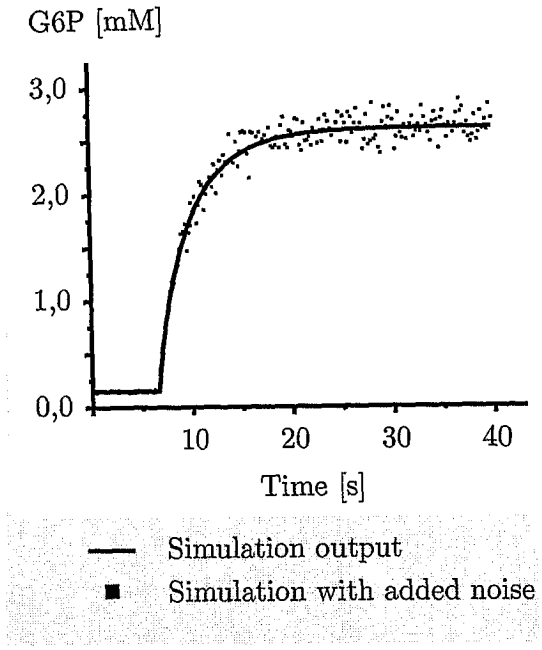


Figure 5.4.: Exact solution of the IVP and the solution with added noise. The pulse of GLC was assumed at time $t = 0$.

The Monte-Carlo simulation shows that the parameters of the example model cannot be identified uniquely. Figures 5.5 and 5.6 summarize the results obtained. In Figure 5.5 for each of the four parameters that had to be identified the final parameter is printed against the final value of the relative least square. Each dot in the graph represents a local minima identified by the parameter fitting algorithm.

If the parameters would be clearly identifiable, the correct parameter value would have to coincide with small values of the relative least square. This is clearly not the case for all four parameters. Parameters $v_{max,1}$ and $v_{max,3}$ show an accumulation of dots close to the correct value with a small least square error. $K_{m,1}$ and $K_{m,3}$ show accumulation of dots for small least square error over the whole range of parameter values. This indicates that these parameters are very unlikely to be identified correctly. The figure also shows dependencies

In Figure 5.6 histograms are shown for each parameter showing the frequency of the result parameter. It can be seen that the frequency of the result parameters is high close to the correct value. This indicates that a large number of parameter fittings

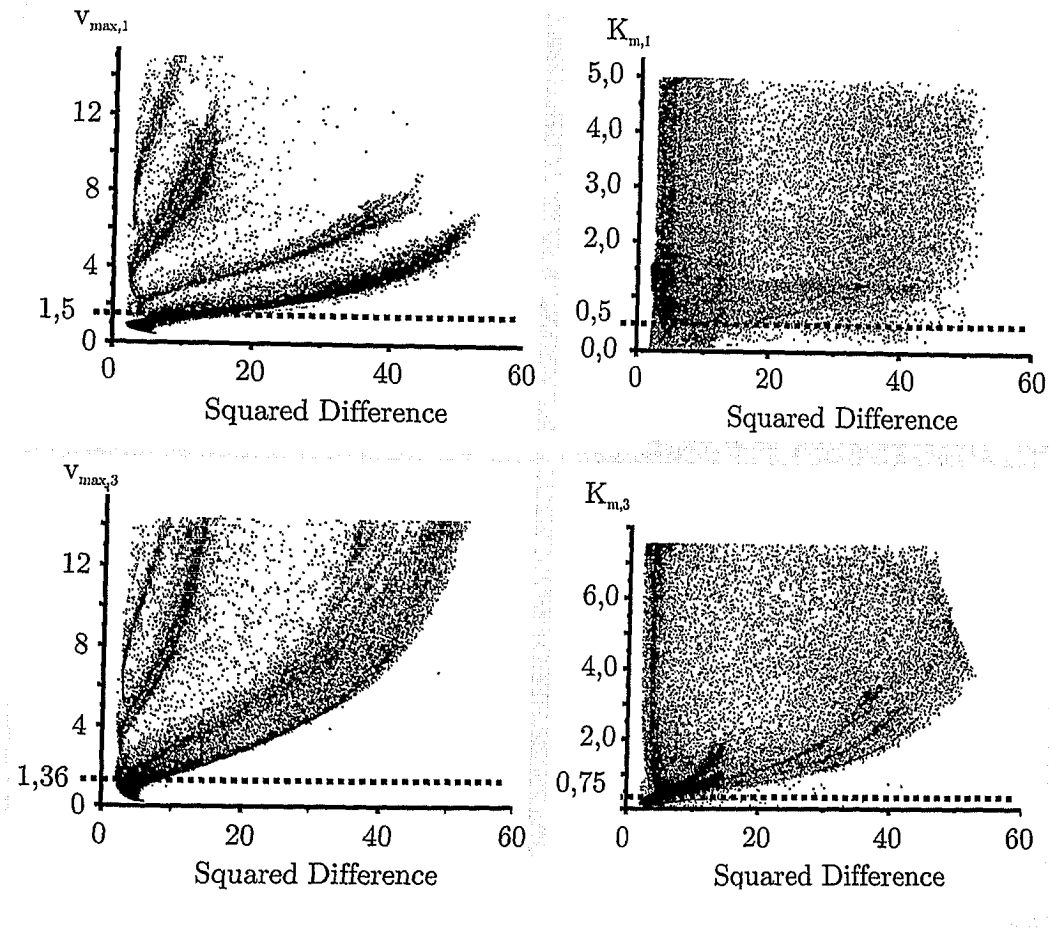


Figure 5.5.: Parameter values that resulted from 25000 parameter fittings versus relative least square error. Vertical dotted lines represent the initial parameter value that should have been identified.

increases the accuracy of the fitting. At least it would allow a conclusion about a range of a certain parameter. Notice, however, that it requires high computational costs to perform thousands of parameter fittings with a larger model than the one used for this illustrative example. It should be also noted that a different error measurement d_i did not lead to significant changes in parameter identifiability.

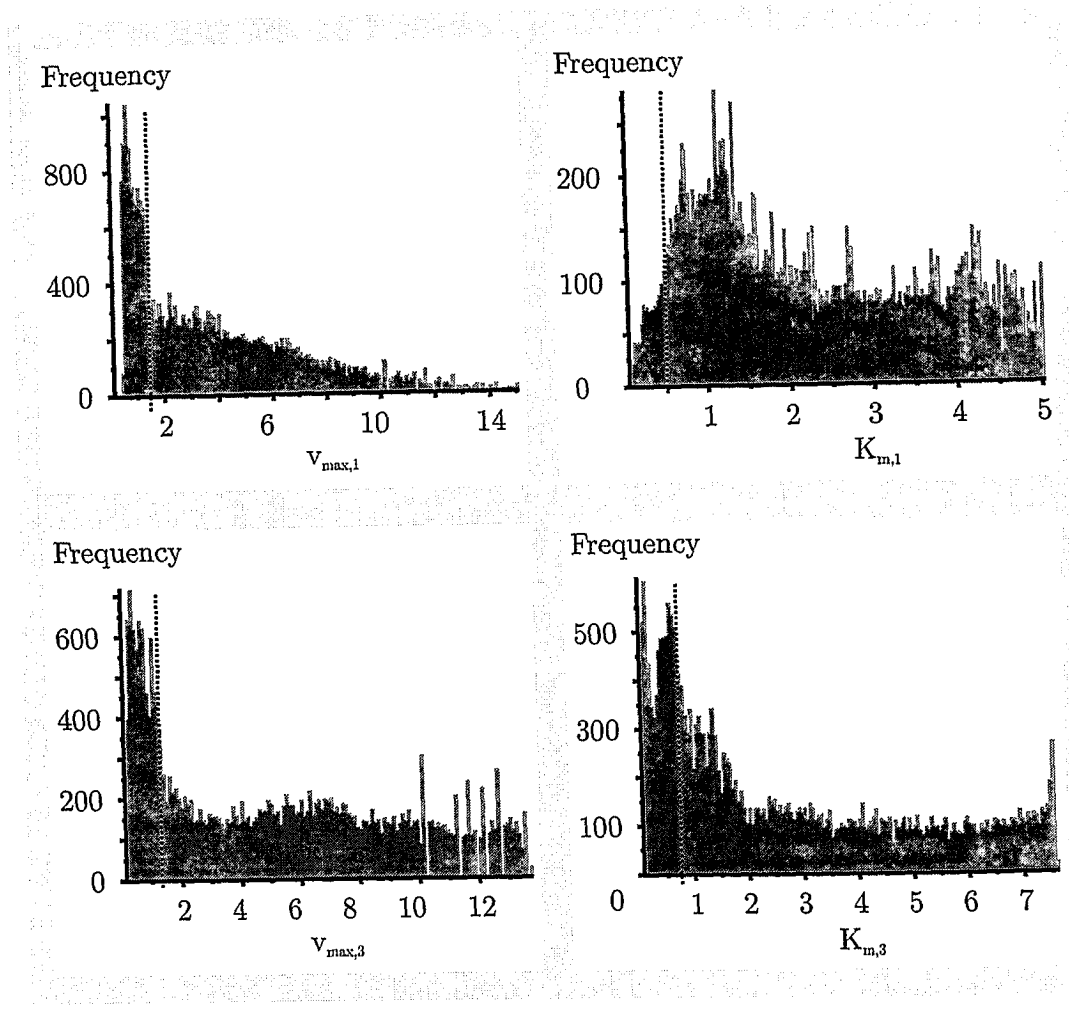


Figure 5.6.: Histograms showing parameters values that resulted from 25000 parameter fittings as optimal solutions. Vertical dotted lines represent the initial parameter value that should have been identified.

Summary of the example model

It has been shown that the MMT software can be used efficiently for parameter fitting of models that represent chemical reaction networks. 25000 parameter fittings could be performed in 10 hours which means that one parameter fitting requires per average 1.5 seconds of computation time.

The example system also shows that problems occur when identifying parameters of chemical reaction systems. In the presented example the parameters are not

clearly identifiable. A parameter fitting leads to a local minimum quickly but the result is not useful without further investigations. It can be also concluded that a large number of parameter fittings allows the identification of a range where the parameters are likely to be located.

As a conclusion, it is important to investigate parameter identifiability e.g. with a Monte-Carlo simulation. Some help also comes from sensitivity analysis. However, sensitivity analysis depends on the initial parameter values of an parameter fitting. Therefore excluding parameters with low sensitivities from parameter fitting is not a solution. E.g. calculation of the sensitivities according to Equation 5.7 using the correct parameter values at time $t = 0$ leads to the conclusion that $K_{m,1}$ and $K_{m,3}$ are more sensitive parameters than $v_{max,1}$ and $v_{max,3}$. The results from the Monte-Carlo simulation suggests the contrary.

5.5. Summary of solution techniques

This chapter describes the numerical methods used for solution of the IVP and the parameter fitting problem. Whereas solution of the IVP does not provide any problems, the parameter fitting problem is difficult to solve. A variety of methods is combined in order to obtain a good performance. Many times a combination of different error functionals, weight factors, or parameter space reductions provided the best solution for the parameter fitting problem. The MMT Software allows to change such strategies during an parameter fitting and therefore provides a good base for the strategy mix. More details about the software setup can be found in Chapter 4. The example model, however, clearly showed that the model setup is crucial for parameter identifiability.

6. Modeling the *Escherichia coli* Network

The MMT software presented in Chapter 4 was used to develop models for chemical reaction pathways. In order to test the tool models ranging from 3 dynamic variables with 13 parameters to 26 dynamic variables with 302 parameters have been developed. More than 50 models have been created and analyzed.

This chapter first presents an isolated system for enzyme production with 7 differential equations. Following a top down approach (see Section 3.4.4), first two small models for the PT-System and the reaction from F6P to FBP catalyzed by phosphofructokinase (PFK) of *Escherichia coli* are shown. Than two complex models of glycolysis, PP-Pathway, and TCA-cycle of *Escherichia coli* with 16 respectively 26 dynamic variables are presented.

6.1. A model for enzyme production

The MMT software can be used to model chemical reaction systems. Its intention is the modeling of complex *in vivo* situations. The software, however, can also be used for *in vitro* systems. Two models describing *in vitro* reactions for production of an enzyme have been set up in order to check the functioning of the MMT software. The system shown in Figure 6.1 produces 7 differential equations and 24 parameters. Differential equations are defined for B1, C1, C2, C3, S, I and P. B2 does not have any known influence on the system. Therefore, it is not integrated in the model. The model was fitted to experimental data gained at the Institute of Biotechnology [12]. During the experiment, 2 mM of substance S are converted to P via an intermediate I. The cosubstrate C2 of the conversion reactions is recycled through C1. It uses an additional substrate B1. The amount of B1 in the system at time 0 is 200 mM. Because only 2 mM can be used (substrate S is only present with 2 mM and C1 + C2 + C3 = 0.4 mM), B1 does not effect the system significantly. An additional flux v_5 exists in the model that describes the degradation of C3. Therefore C1 + C2 + C3 is decreasing with time.

Figure 6.2 summarizes experimental data and parameter fitting results obtained

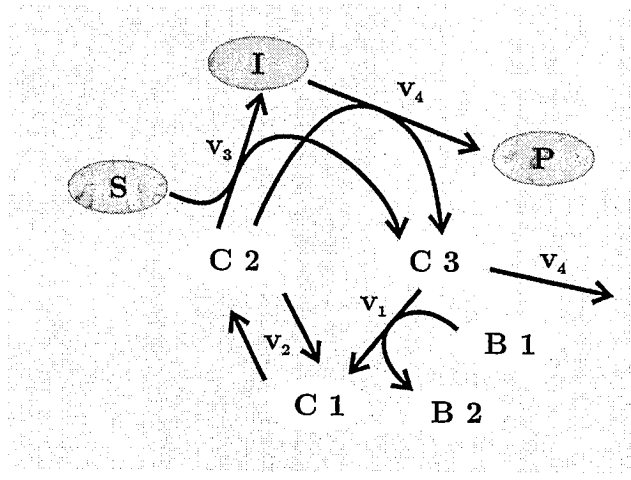


Figure 6.1.: An *in vitro* chemical reaction system that is used to produce compound P from substrate S through an intermediate I. The cosubstance C2 of the two reactions is recycled by two reactions catalyzed by enzymes added to the *in vitro* system. Substance B1 is converted to B2 during this process.

with the model. It can be seen that the initial parameters could not represent any effects from the experiment. The reason is that for about two thirds of the 24 parameters of the model no literature values or experimental data was available. Therefore, the initial guesses of these parameters have been set to 1.0. In order to get the fitted parameters, a strategy was applied that randomly changes parameter values by a large percentage before each parameter fitting run. Using this strategy, the presented result was calculated in less than 15 minutes. The identified parameters can now be optimized further with a strategy more suitable for parameter values close to an optimal solution.

6.2. Glycolytic oscillations

Oscillations in the glycolysis of yeast have been reported and modeled before [95], [35], [77], [94], [43]. The oscillatory behavior usually is assumed to result from inhibitions of the enzyme PFK. This enzyme catalyzes the glycolytic reaction from F6P to FBP. PFK is one of the most analyzed enzymes and is considered to play a significant role in metabolic regulation. However, the amount of available data does not allow a unique interpretation. The entry for PFK in the public database BRENDA lists 67 possible inhibitors. Most of these inhibitors cannot be measured in the cell. It is therefore not possible to find a correct mathematical representation

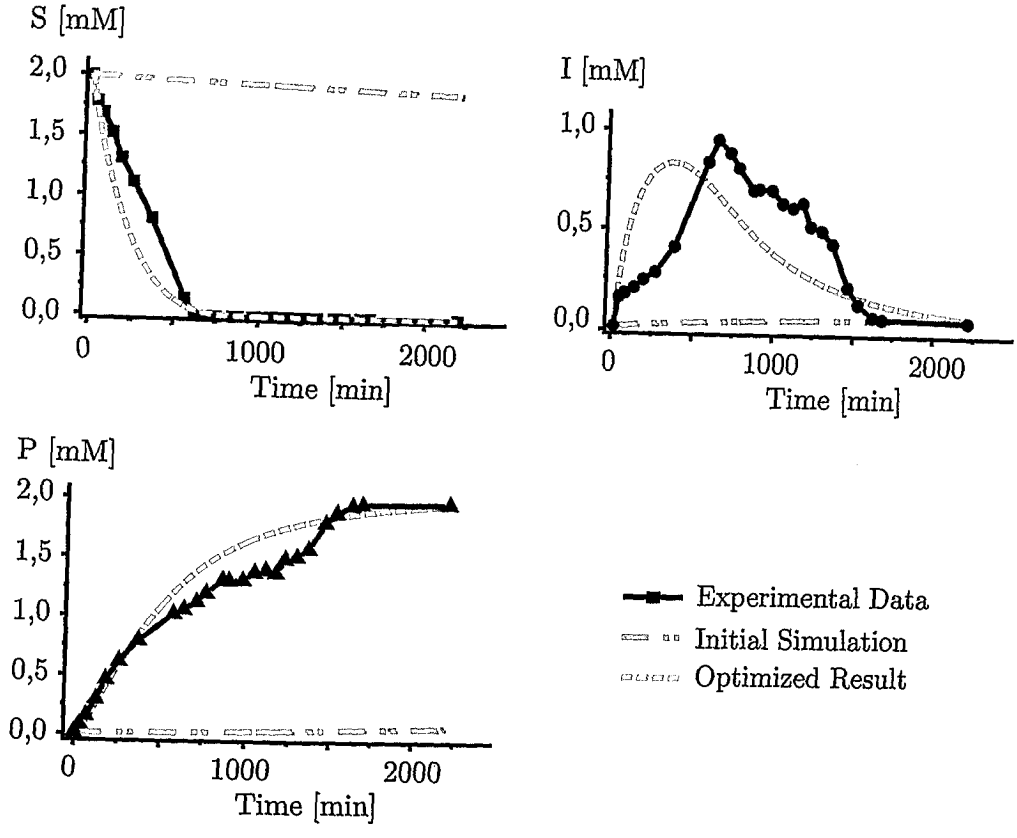


Figure 6.2.: Experimental data and simulation results before and after parameter fitting of parameters. Shown are values for the substrate S, intermediate I and final product P.

for the PFK. Also, a complete model including all these effects would include too many parameters to be useful in a pathway model.

The models for PFK in the literature mentioned focus on the main inhibitors and activators of the enzyme. Goldbeter and Lefever [36] developed a model that describes the allosteric behavior of the enzyme (Figure 6.3). The term allosteric is used to describe a reaction in which the binding of a compound to the enzyme influences the binding of the second. For PFK, a positive feedback exists, i.e. the reaction rate from F6P to FBP increases if more FBP is available. Additionally, cooperative behavior between the enzyme subunits is a necessary assumption for instability in the PFK kinetics [94].

It has been demonstrated by the authors with a simple mathematical model that these two effectors on the enzyme PFK are sufficient for an oscillatory behavior.

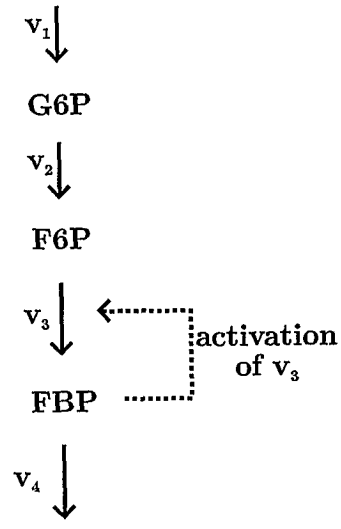


Figure 6.3.: The model for PFK that results in oscillations. v_1 is a constant input flux into the G6P pool.

A system with three dynamic concentrations was derived from this description in order to examine the findings of the models. In addition to the concentrations of F6P and FBP, the G6P concentration is included as a dynamical variable in the model. Together, a constant flux into G6P, the reactions G6P $\rightarrow$ F6P, F6P $\rightarrow$ FBP, and a flux exiting FBP are included in the model.

In the following, all concentrations are non-dimensionalized according to Goldbeter and Lefever. The following equations result:

$$\begin{aligned} \dot{\text{G6P}} &= v_{\text{G6PIn}\rightarrow\text{G6P}} - v_{\text{G6P}\rightarrow\text{F6P}} = v_1 - v_2 \\ \dot{\text{F6P}} &= v_{\text{G6P}\rightarrow\text{F6P}} - v_{\text{F6P}\rightarrow\text{FBP}} = v_2 - v_3 \\ \dot{\text{FBP}} &= v_{\text{F6P}\rightarrow\text{FBP}} - v_{\text{FBP}\rightarrow\text{FBPOut}} = v_3 - v_4 \end{aligned} \tag{6.1}$$

For v_2 a reaction rate proportional to the concentration of G6P was assumed

$$v_2 = v_{max,2} \cdot \text{G6P}$$

For the flux from F6P to FBP, Goldbeter and Lefever developed the following equa-

tion:

$$\mathbf{v_3} = \frac{2 \cdot b \cdot D_1 \cdot \frac{e}{e+1} \cdot \mathrm{F6P} \cdot (1+\mathrm{FBP})^2 \cdot (1+\frac{\mathrm{F6P}}{e+1})}{L \cdot (1+\mathrm{F6P} \cdot c)^2 + (1+\mathrm{FBP})^2 \cdot (1+\frac{\mathrm{F6P}}{e+1})^2} \tag{6.2}$$

For the flux out of the F6P pool again a kinetic rate proportional to the F6P concentration is assumed

$$\mathbf{v_4} = -v_{max,4} \cdot \mathrm{FBP} \tag{6.3}$$

Details of the kinetic parameters can be found in Goldbeter and Lefever [36]. The authors use parameters from kinetic studies with PFK from yeast and *Escherichia coli*. Figure 6.4 shows simulation results with the parameters from the authors.

The obtained results were compared with the experimental data from GLC pulse experiments that are discussed in Section 2.2. This data suggests an oscillatory behavior of the intracellular G6P concentration after the pulse of GLC was given. However, the F6P concentration in this experiment was below the detection limit.

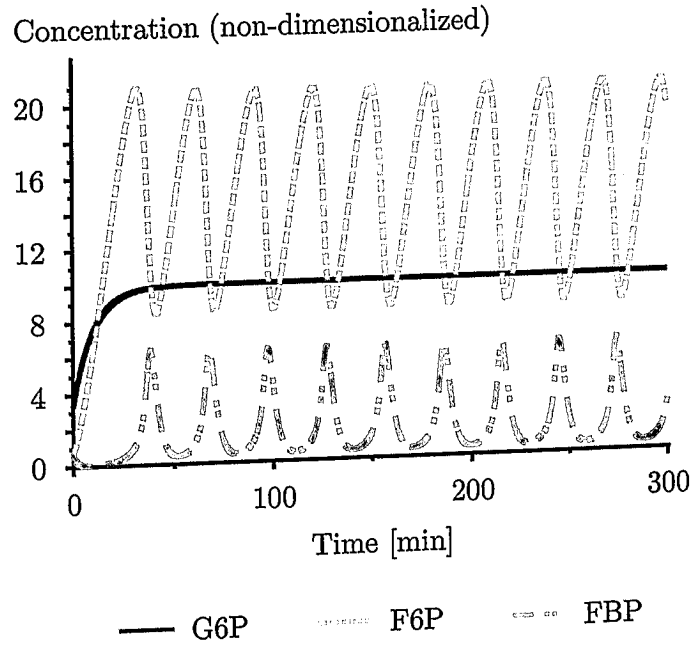


Figure 6.4.: Simulation results of the PFK model from Goldbeter and Lefever.

The original system of ODEs from Goldbeter and Levefer consists of two differential equations for F6P and FBP only. This allows analytic stability analysis of the system. This analysis is based on a linearization at an equilibrium point. The equilibrium points of the system can be calculated explicitly. With the help of this stability analysis the occurrence and frequency of oscillations could be estimated. The MMT software provides the derivatives of the model equations that are needed for stability analysis.

In order to apply the PFK model to the experimental data in this work, parameter values were sought that maintain oscillations in measured ranges of the concentrations. Due to the large number of parameters, this problem could not be solved explicitly. Instead, a large amount of simulations was performed. Parameters varied in large ranges for these simulations. No parameter values could be found that resulted in oscillations by maintaining a concentration relation of $\frac{F6P}{FBP} < 0.002$ as found in the experiments. It was concluded that the PFK feedback inhibition by FBP and the cooperative interactions between the subunits of PFK could not be the reason for the oscillations suggested by the experimental data.

6.3. Models for GLC uptake of *Escherichia coli*

The PT-System for GLC uptake of *Escherichia coli* consists of three phosphotransfer reactions, Figure 6.5. Next to the PFK, the PT-System is often thought to have the strongest influence on metabolic regulation.

Two models for the system can be found in [83] and [82]. Both models describe the system in a detailed way by including all three phosphotransfer reactions. Each reaction step is divided into two steps:

1. **Building of the enzyme-substrate complex:** The first step of each reaction is the binding of the substances to the enzyme.

2. **Dissociation of compounds from the complex:** The second step of the reaction is the dissociation of the products from the enzyme-substrate complex.

This straightforward approach is not based on the pseudo-steady-state assumption which would assume the substrate-enzyme complex concentration to be constant [21]. This assumption is part of most kinetic rate equations. Without this assumption, each reaction step can be modeled with a simple kinetic rate equation: assume that the reaction rate is proportional to the substrate concentration of the

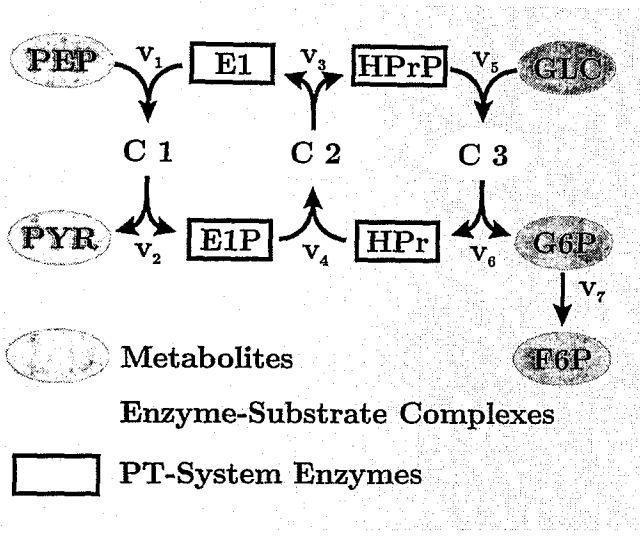


Figure 6.5.: The three phosphotransfer reactions of the PT-System in *Escherichia coli*. E1, HPr: unphosphorylated enzymes of the PT-System. E1P, HPrP: phosphorylated enzymes. Arrows indicate the direction of the irreversible reactions assumed in the model.

educt, or to the product of the concentrations if two educts are present. This is based on a common assumption that the substance concentration is proportional to the reaction rate in chemical reactions. The stoichiometric matrix of the PT-System reads as

	G6P	F6P	E1	E1P	HPr	HPrP	C1	C2	C3
v_1	0	0	−1	0	0	0	1	0	0
v_2	0	0	0	1	0	0	−1	0	0
v_3	0	0	1	0	0	1	0	−1	0
v_4	0	0	0	−1	−1	0	0	1	0
v_5	0	0	0	0	0	−1	0	0	1
v_6	1	0	0	0	1	0	0	0	−1
v_7	−1	1	0	0	1	0	0	0	−1

and the fluxes are modeled as

$$v_1 = k_1 \cdot \text{PEP} \cdot \text{E1}$$

$$v_2 = k_2 \cdot \text{C1}$$

$$v_3 = k_3 \cdot \text{C2}$$

$$
\begin{aligned}
\mathbf{v}_4 &= k_4 \cdot \mathrm{E1P} \cdot \mathrm{HPr} \\
\mathbf{v}_5 &= k_5 \cdot \mathrm{HPrP} \cdot \mathrm{GLC} \\
\mathbf{v}_6 &= k_6 \cdot \mathrm{C3} \\
\mathbf{v}_7 &= \frac{V_f \frac{\mathrm{G6P}}{K_{mS}} - V_r \frac{\mathrm{F6P}}{K_{mP}}}{1 + \frac{\mathrm{G6P}}{K_{mS}} + \frac{\mathrm{F6P}}{K_{mP}}}
\end{aligned}
\tag{6.4}
$$

A detailed model that includes enzyme substrate complexes as dynamic variables can lead to different results if compared to models based on the pseudo-steady-state assumptions. This, however, strongly depends on the total enzyme concentration. If the concentration is high, a large amount of enzyme could be e.g. stored in the enzyme-substrate complex. A sudden change in the growth conditions could change this situation and free a significant amount of enzyme and substrate from the complex. This phenomenon can only be explained with detailed models of these processes.

For the experimental conditions of the available data (pulse experiments) it is likely that the above mentioned effects occur. Before the pulse was given in the experiments, cells were grown on GLC limited conditions. Relatively high PEP concentrations coincided with low G6P concentrations. This suggests that the concentration of the PEP-enzyme I complex is high whereas concentration of the GLC-HprP complex remains low. Due to the sudden and large increase in the extracellular GLC concentration this situation is likely to change. In order to evaluate the effects of such changes, both models presented in [83] and [82] have been analyzed with the experimental data sets. The models use GLC, PEP and PYR as input variables and only model the PT-System plus the flux from G6P to F6P. For the flux from G6P to F6P, a reversible Michaelis-Menten equation was assumed.

In Figure 6.6 two simulations produced with the model in [83] are shown and data from a GLC pulse experiment as described in Section 2.1. Initial parameters were also taken from [83]. The upper graph in the Figure shows a result from an parameter fitting for the first 17 seconds of the experiment. The result shows good agreement between the model and the data for this time frame. The second part of the experiment cannot be reproduced by the model. If the complete data file until 36 seconds was used during parameter fitting, no satisfactory fit was achieved. The lower part of the figure shows the best result obtained with parameter fitting using the SUBPLEX algorithm. It can be clearly seen that the first part of the model does not agree very well. The parameter fitting algorithm could not identify parameters that fit the first and the second part of the data.

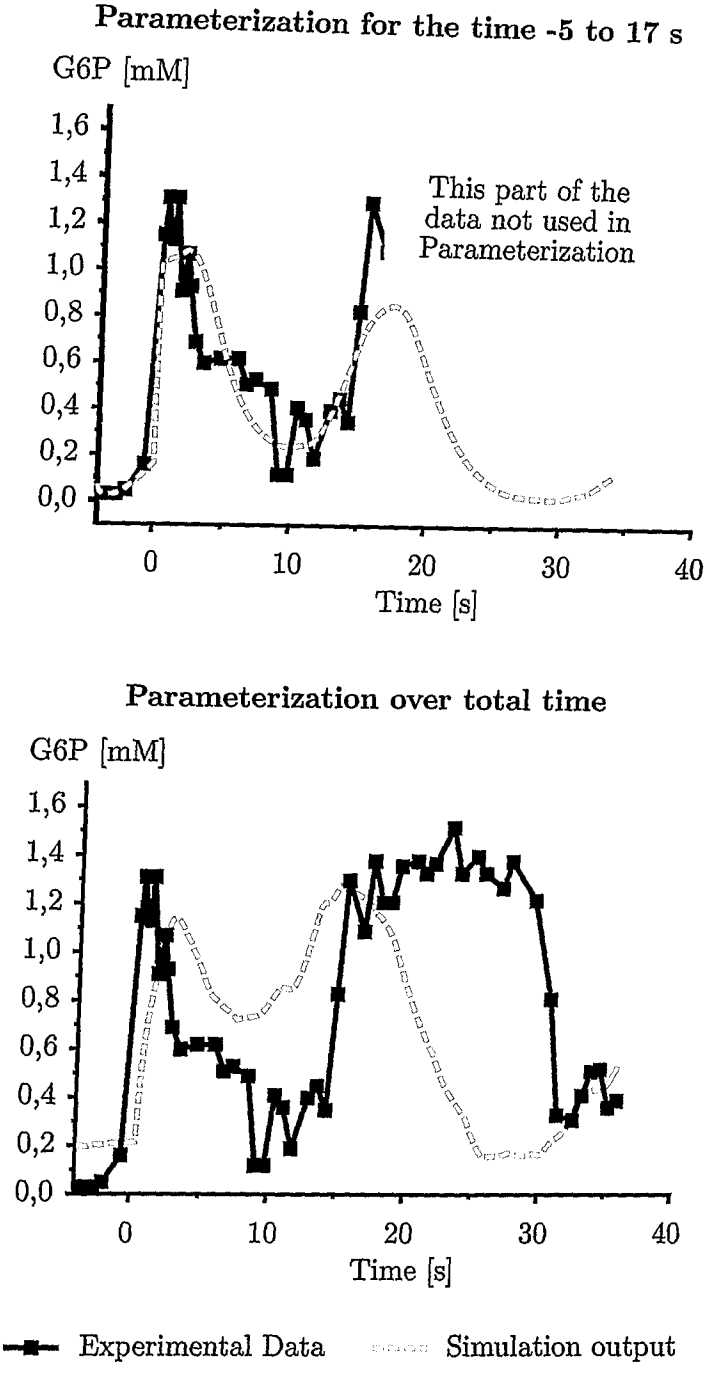


Figure 6.6.: Simulation results of the PT-System model. The upper graph results from a parameter fitting for the time frame -5 to 17 seconds. The lower graph is from an parameter fitting for the complete time frame from -5 to 36 seconds.

It has been discussed previously in Section 2.2 that there is no biological interpretation for the time window 17 to 32 seconds where a high G6P concentration coincides with a low PEP concentration. Both models for the PT-System also were not able to explain this phenomenon. The hypothesis that effects of the three phosphotransfer reactions and the enzyme-substrate-complexes alone are responsible for such behavior has to be rejected.

6.4. A model of glycolysis, PP-Pathway, and TCA-cycle

A complex model for *Escherichia coli* was constructed. The goal was to find the effects of metabolic regulation that are able to explain the behavior seen in the data. Figure 6.7 shows the reactions of the central metabolic pathway of *Escherichia coli* that were included in the model.

Simplifications have been made at the following points (compare with the more detailed pathway in Figure B.1):

- The PP-Pathway has been reduced to one compound called PPP (black-box formulation). A flux is assumed from G6P to PPP and a flux from PPP directly to PEP. This is a simplification because there are additional pathways from the PP-Pathway back to glycolysis (F6P, GAP). Additionally, a flux from PPP to biomass production is assumed in the model.
- The two reactions PYR $\rightarrow$ MAL and MAL $\rightarrow$ OAA are reduced to one reaction.
- The TCA-cycle is reduced to OAA, ACoA, CoA, CIT, and a compound called TCA (black-box assumption). From CIT, a flux enters TCA. TCA is then directly converted to OAA. Fluxes to biomass from the TCA-cycle are summarized in one equation, a flux from CIT into biomass.
- Fluxes into biomass are summarized in three kinetics: from PPP, CIT, and PYR. It keeps the number of parameters small if compared to an approach where many compounds have a flux into biomass.

Additionally, a reaction that includes maintenance and production of ATP has been assumed. Together, the model contains 16 compound concentrations. The kinetic rate equations used contain a total of 159 parameters. The rate equations are summarized in Table 6.1. Formulas for these equations are summarized in Appendix C.

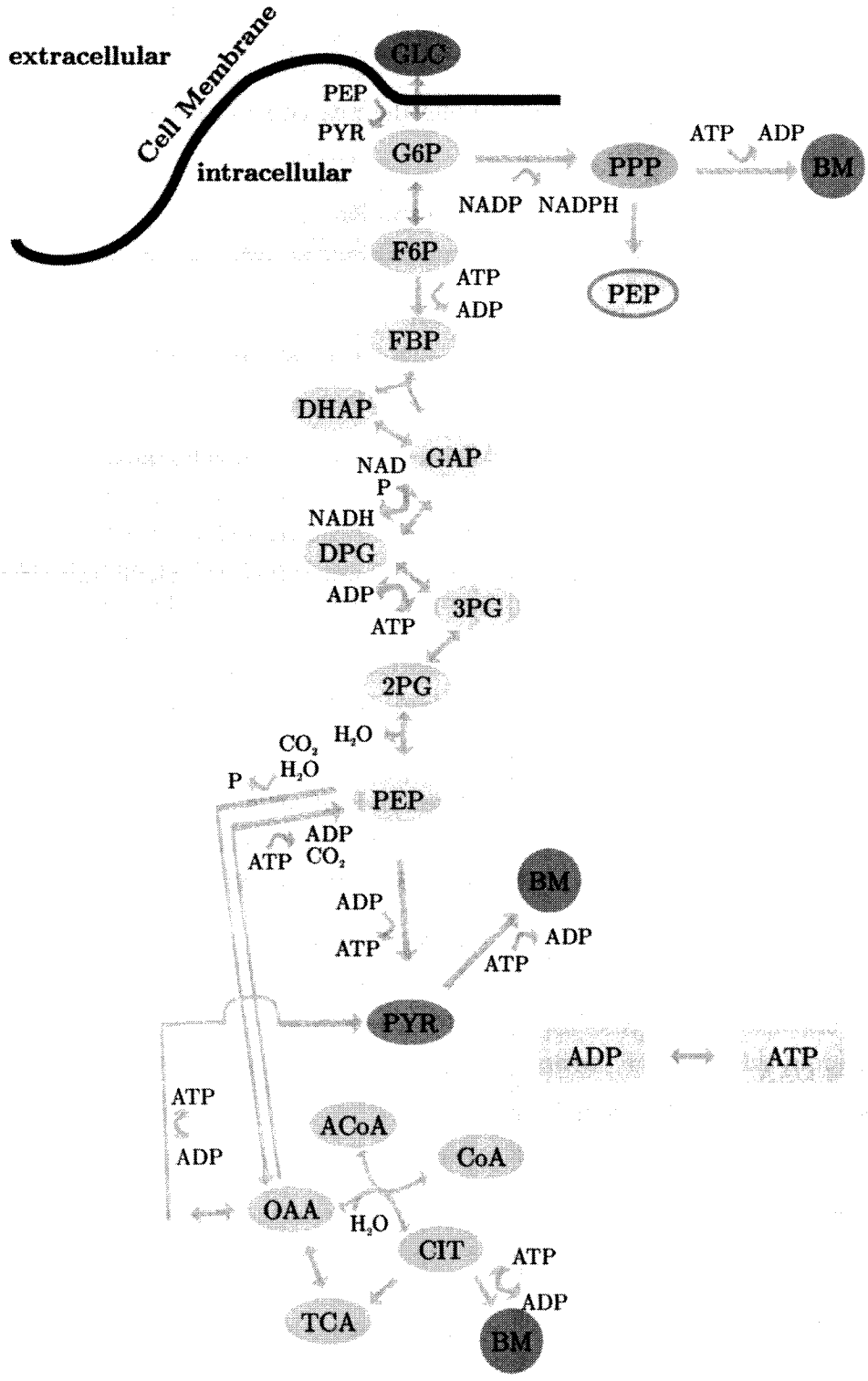


Figure 6.7.: Pathway of the model including black-box formulations of the PP-Pathway and the TCA-cycle.

Table 6.1.: All reactions of the model and the rate equations used in the mathematical model.

Reaction Rates $\mathbf{v}_i$	Rate equation
GLC + PEP $\leftrightarrow$ G6P + PYR	see text
G6P $\leftrightarrow$ F6P	reversible Michaelis-Menten
G6P $\rightarrow$ PPP	Michaelis-Menten
PPP + ATP $\rightarrow$ BM + ADP	Michaelis-Menten, two substrates, irreversible
PPP $\leftrightarrow$ PEP	reversible Michaelis-Menten
F6P + ATP $\leftrightarrow$ FBP + ADP	ordered inhibited Bi-Uni-Kinetic (inhibition by PEP)
FBP $\leftrightarrow$ DHAP + GAP	ordered Uni-Bi-Kinetic
DHAP $\leftrightarrow$ GAP	reversible Michaelis-Menten
GAP + NAD $\leftrightarrow$ DPG + NADH	ordered Bi-Bi-Kinetic
DPG + ADP $\leftrightarrow$ 3PG + ATP	ordered Bi-Bi-Kinetic
3PG $\leftrightarrow$ 2PG	reversible Michaelis-Menten
2PG $\leftrightarrow$ PEP	reversible Michaelis-Menten
PEP + ADP $\leftrightarrow$ PYR + ATP	ordered Bi-Bi-Kinetic
PYR + ATP $\leftrightarrow$ OAA + ADP	ordered Bi-Bi-Kinetic
PYR + ATP $\leftrightarrow$ BM + ADP	ordered Bi-Bi-Kinetic
PEP $\rightarrow$ OAA	see text
OAA + ATP $\leftrightarrow$ PEP + ADP	see text
OAA + ACoA $\leftrightarrow$ Cit + CoA	ordered Bi-Bi-Kinetic
CIT $\rightarrow$ TCA	Michaelis-Menten, two substrates, irreversible
CIT + ATP $\leftrightarrow$ BM + ADP	ordered Bi-Uni-Kinetic
CIT $\leftrightarrow$ OAA	reversible Michaelis-Menten
ATP $\leftrightarrow$ ADP	reversible Michaelis-Menten

As listed in the table, the reaction from F6P to FBP is described with an ordered inhibited Bi-Uni-Kinetics which has the following form [63]:

$$\begin{aligned} v_{\mathrm{F6P}\to\mathrm{FBP}} = v_{max} \cdot \frac{(\mathrm{F6P}\cdot\mathrm{ATP} - \frac{\mathrm{FBP}}{K_{eq}})}{\mathrm{F6P}\cdot\mathrm{ATP} + K_{m,\mathrm{F6P}}\cdot\mathrm{ATP} + K_{m,\mathrm{ATP}}\cdot\mathrm{F6P}...} \\ \frac{}{... + \frac{v}{v_r K_{eq}} \cdot \left(K_{m,\mathrm{FBP}} + \mathrm{FBP}\cdot(1 + \frac{\mathrm{F6P}}{K_{i,\mathrm{F6P}}}) + (\frac{\mathrm{PEP}}{K_I})^n\right)} \end{aligned} \tag{6.5}$$

This reaction assumes an inhibition by PEP. The importance of this inhibition was analyzed in detail.

For GLC uptake, PEP $\to$ OAA, and OAA $\to$ PEP, equations from [59] have been used. In [59], the derivation of these rate equations is based on a combination of algebraic equations and fuzzy logic in order to incorporate qualitative knowledge into the enzyme kinetic equations. For the reaction from OAA $\to$ PEP an algebraic equation was combined with a set of fuzzy rules. These fuzzy rules described a strong inhibition of the reaction by high concentrations of ATP, a phenomenon derived from experimental data presented also in [59]. The authors suggested to include the fuzzy rules in the algebraic equation by multiplying v_{max} with a factor α_{pck} of the form

$$\alpha_{pck} = \begin{cases} 1.0 & ,\text{for } \mathrm{ATP} \le 0.9 \\ (1 - \frac{\mathrm{ATP}-0.9}{0.2}) \cdot b \cdot \mathrm{ATP} & ,\text{for } 0.9 < \mathrm{ATP} \le 1.1 \\ 1.0 - \alpha_{pck} & ,\text{for } 1.1 < \mathrm{ATP} \end{cases}$$

where b was estimated by the authors as 0.24 mM^{-1}. Notice that even though the the authors derived the equation from fuzzy rules, they have chosen to formulate the equation as a simple piecewise function. The equation for α_{pck} is problematic because it attains negative values for large ATP concentration. To avoid this problem, the following equation has been used for α_{pck} to include a strong inhibition by ATP:

$$\alpha_{pck} = \frac{1}{1 + e^{b\cdot\mathrm{ATP}}} \tag{6.6}$$

The overall equation for the reaction OAA $\to$ PEP is then

$$v_{\mathrm{OAA}\to\mathrm{PEP}} = v_{max} \cdot \frac{K_1 \cdot \mathrm{ATP}\cdot\mathrm{OAA} - K_2 \cdot \mathrm{PEP}\cdot\mathrm{ADP}}{(1 + e^{b\cdot\mathrm{ATP}})\cdot P} \tag{6.7}$$

with

$$\begin{aligned} P &= 1 + K_2 \cdot \mathrm{ATP} + K_3 \cdot \mathrm{OAA} + K_4 \cdot \mathrm{ATP} \cdot \mathrm{OAA} \\ &\quad + K_5 \cdot \mathrm{PEP} + K_6 \cdot \mathrm{ADP} + K_7 \cdot \mathrm{PEP} \cdot \mathrm{ADP} \end{aligned} \tag{6.8}$$

The same equation was used for the uptake of GLC with the PT-System:

$$\mathrm{v}_{\mathrm{GLC}\rightarrow\mathrm{G6P}} = v_{max} \cdot \frac{K_1 \cdot \mathrm{PEP} \cdot \mathrm{GLC} - K_2 \cdot \mathrm{G6P} \cdot \mathrm{PYR}}{(1 + e^{b \cdot \mathrm{PEP}}) \cdot P} \tag{6.9}$$

with

$$\begin{aligned} P &= 1 + K_2 \cdot \mathrm{PEP} + K_3 \cdot \mathrm{GLC} + K_4 \cdot \mathrm{PEP} \cdot \mathrm{GLC} \\ &\quad + K_5 \cdot \mathrm{G6P} + K_6 \cdot \mathrm{PYR} + K_7 \cdot \mathrm{G6P} \cdot \mathrm{PYR} \end{aligned} \tag{6.10}$$

Notice that this equation assumes a strong inhibition of the PT-System for high PEP-concentrations which is not reported in the literature. This assumption will be discussed later.

The second equation used from [59] describes the reaction from PEP to OAA. In this case, however, the algebraic form of the equation was used because it shows the same behavior as the equation based on fuzzy logic. The equation reads as

$$\mathrm{v}_{\mathrm{PEP}\rightarrow\mathrm{OAA}} = \frac{\mathrm{PEP}}{K_m + \mathrm{PEP}} \cdot \frac{K_1 + K_2 \cdot \mathrm{ACoA} + K_3 \cdot \mathrm{FBP} + K_4 \cdot \mathrm{ACoA} \cdot \mathrm{FBP}}{1 + K_5 \cdot \mathrm{ACoA} + K_6 \cdot \mathrm{FBP}} \tag{6.11}$$

The equation is based on a Michaelis-Menten type kinetics. The second fraction is a formula for the maximum reaction rate v_{max} of the reaction. The equation assumes inhibitions of the reaction by ACoA and FBP which is modeled by variation of v_{max}.

6.4.1. Fitting the model to data from the steady state

After defining initial parameters, a parameter fitting was defined for fitting the model to the steady state of the underlying experimental data, i.e. to the time frame -5 to 0 seconds. As expected, the results showed an overfitting problem. The steady state measurements are not sufficient for estimation of all kinetic parameters. Therefore the obtained fitted parameter values are only useful as initial values for a parameter fitting of the dynamic data (after the GLC pulse).

Figure 6.8 shows an example of a steady state simulation for PEP. All other compounds also reached steady state about 0.5 seconds after the simulation was started. Because some compound concentrations do not start at time $t = -5s$ at their steady state values, the system has first to evolve to the steady state. For the compounds where measurements are available, the initial conditions have been taken from these experiments. As shown for PEP in Figure 6.8, the model calculates a slightly different steady state. It is reached after about 0.5 seconds.

Figure 6.9 summarizes the differences between the simulated steady states and the steady states measured in the experiment for each individual compound concentration. The steady state of the reaction system is represented by the model with good accuracy. Problematic is the fact that this good representation of the steady-state is usually lost if the model is fitted to the data of the pulse experiment.

The results obtained for the steady state show a GLC uptake rate of $0.58\frac{mM}{s}$. The estimated range for this value from the experimental data was calculated as $0-0.9\frac{mM}{s}$. If flux estimation are available, it is also possible to include this flux as a constraint in the parameter fitting for the steady state. Because in the experimental data, no measurements from flux-analysis are available, such a constraint was not used.

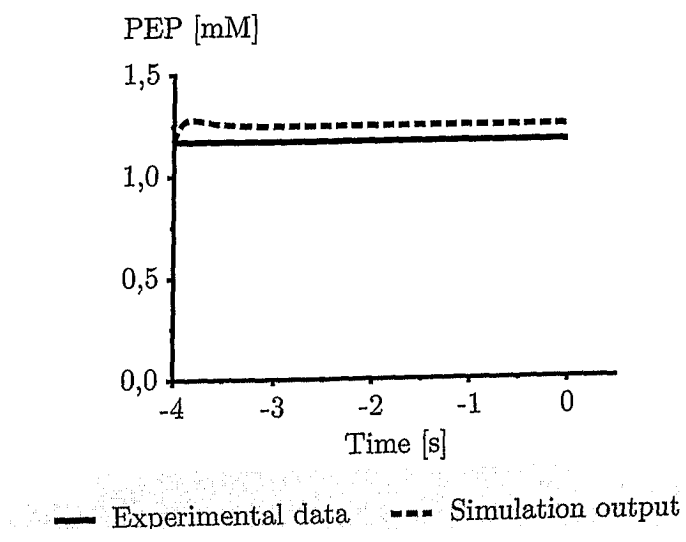


Figure 6.8.: Simulation results for the steady state. Shown is the intracellular PEP concentration.

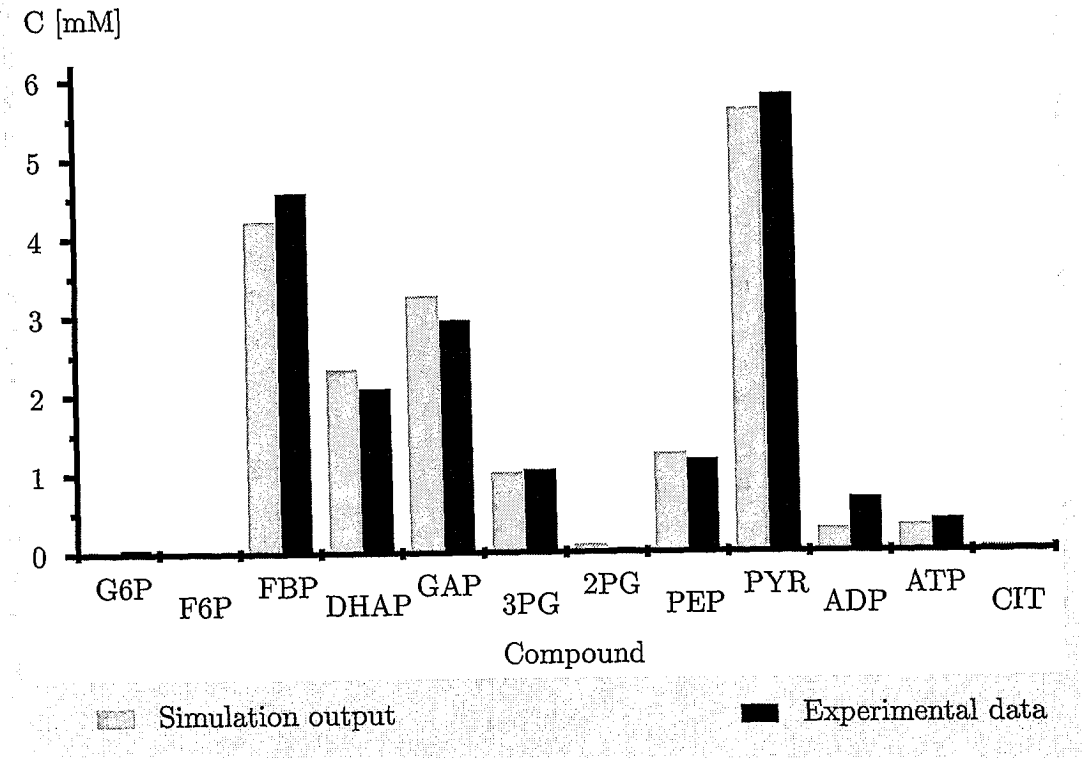


Figure 6.9.: Steady state concentrations of the intracellular compounds and values from experiments.

Additionally, a futile cycling by the reactions PEP $\rightarrow$ PYR, PYR $\rightarrow$ OAA, OAA $\rightarrow$ PEP is calculated by the model [18]. The size of this flux is 0.25 $\frac{mM}{s}$. Due to the setup of the system, one ATP is lost in this futile cycling. Notice that this result can depend on the simplifications made for the model for the reactions from PYR to the TCA-cycle. Also, energy limitation does not occur in the steady state situation because the additional reaction ADP $\leftrightarrow$ ATP can provide enough energy. This reaction is assumed to include maintenance terms (i.e. usage of ATP) and the respiratory chain (i.e. production of ATP). The respiratory chain was included in the model with this single reaction. As a result, the steady state situation does not give new insights in the system.

6.4.2. Fitting the model to data from a pulse experiment

The described model was fitted to the experimental data of the GLC pulse experiment shown earlier. The focus was again on the PT-System and the PFK inhibition. Both reaction rates include a strong inhibition by PEP. For PFK, a strong inhibition by PEP is reported in [8]. The inhibition of the PT-System by PEP is hypothetical. It could represent a situation where the cell needs to decrease the uptake rate because intracellular compound concentrations are too high.

Oscillations of the compounds

During simulations, the inhibition constants K_I in the Equations 6.9 and 6.7 have been set to 1.0 mM indicating a weak inhibitory effect. In contrast, parameter n in Equation 6.5 has been set to 8.0 with an value for K_I of 1.0, indicating a strong inhibition of PFK by PEP as reported in the literature. As in the Goldbeter models described earlier, the inhibition of PFK leads to oscillations in glycolysis. Figure 6.10 shows simulated values for G6P, PEP, and ATP. All other compounds show oscillations of the same frequency. The period of oscillations is in the range of 20 seconds and therefore in a range that is of interest for the data from the GLC pulse experiments.

Dynamic behavior for various inhibition constants

Inhibition of PFK by PEP leads to oscillations, however, this effect was not sufficient to fit the model to the experimental data. parameter fittings were performed that also included inhibition of the PT-System by PEP as mentioned before. In general results from the parameter fitting using the dynamic data from the pulse experiment do not represent the steady state situation, even if the initial parameter vector showed good representation of the steady state. Figure 6.11 compares three different results from parameter fittings:

1. **Inhibition of the PT-System and PFK:** If both reactions are assumed to be inhibited, a dynamic behavior was achieved that comes somewhat close to the experimental data. Notice that a steady state for the PEP concentration actually seems to evolve right before the pulse (see Figure 6.11, top right, PEP concentration right before the pulse at time $t = 0$). Apparently the system has to reach a steady state first as could be seen in the steady state simulations above.

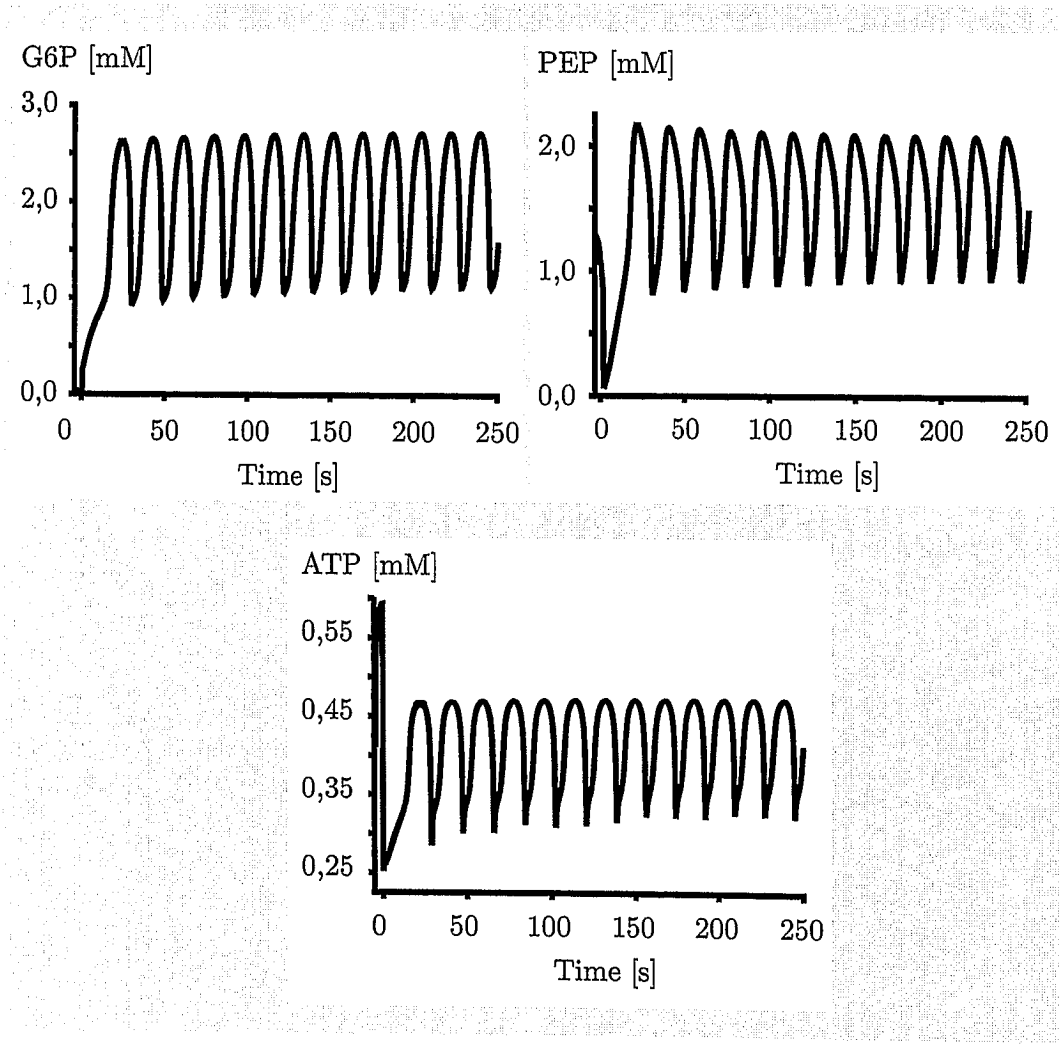


Figure 6.10.: Simulated values over a time period of 250 seconds. Oscillations result from the strong inhibition of PFK by PEP.

2. **Inhibition of PFK only:** If the inhibition of the PT-System is not assumed but only an inhibition of PFK by PEP, the qualitative behavior of the system does not match the experimental data anymore. The result shown is obtained through an additional parameter fitting without the PT-System inhibition. Especially the dynamics of the G6P concentration are effected.

3. **Inhibition of the PT-System only:** The third results shown in the figure is obtained by assuming inhibition of the PT-System by PEP and neglecting inhibition of the PFK by PEP. The result from parameter fitting shows that the dynamics seen in the experiment cannot be explained with this model. As in the second case, the dynamics of the G6P concentration change significantly.

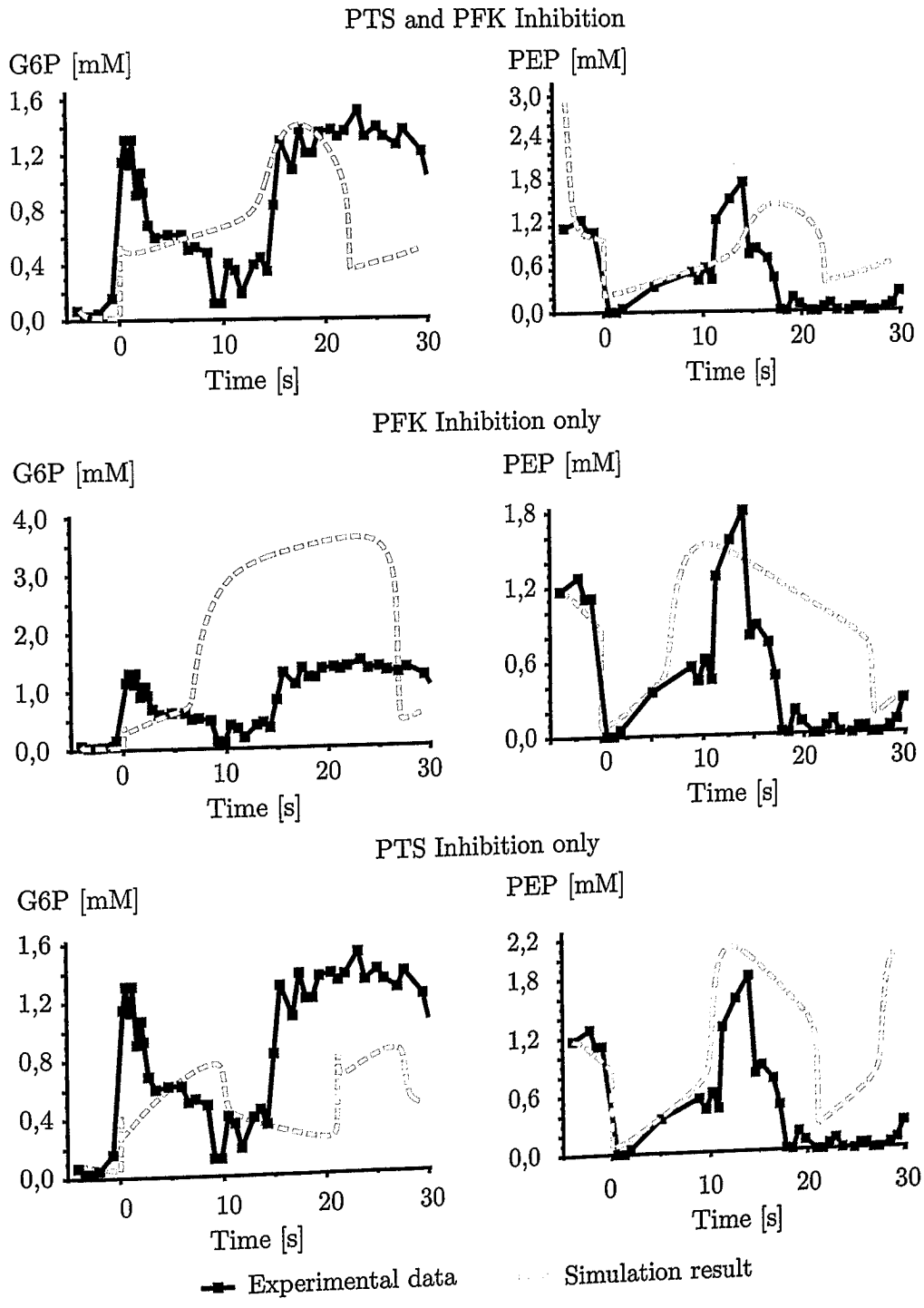


Figure 6.11.: Experimental data and simulated values for G6P and PEP over a time period of 35 seconds. From -5 to 0 seconds a steady state situation was present in the experiment, the pulse of GLC was given at time $t = 0s$. A qualitative behavior close to the experimental data occurs only if inhibitions for the PT-System and the PFK are assumed.

6.4.3. Summary of the complex model of glycolysis, PP-Pathway, and the TCA-cycle

From the complex model with 16 dynamic variables and 159 parameters of the central metabolic pathway of *Escherichia coli* the following conclusions were obtained:

1. The model provides a good description for the steady state before the GLC pulse.
2. The model fails to explain dynamic behavior of the system after the GLC pulse.
3. The system shows strong dynamics and oscillations. These effects can result from one single inhibition (PT-System or PFK) and increase if more effects are included in the model.
4. The fit obtained for the steady state situation resulted from parameter fittings with little computing time. This suggests that the parameter fitting algorithm might not be the problem for the unsatisfactory fit in the dynamic situation. It is concluded that the model must be modified in order to explain the measured values.

6.5. A detailed model of the central metabolic pathways

The complex model described in the last section suggests that inhibitory effects have a strong influence on the dynamics of the system. Following this conclusion, the model was extended to include more regulating phenomena of the pathway. This extensions mainly focused on more detailed descriptions of the PP-Pathway and the TCA-cycle and additional fluxes to biomass. Also, NAD, NADH, NADP, NADPH, and AMP were included in the model to follow energy charges. Figure 6.12 and 6.13 show all reactions of the system. The following assumptions have been made for the model:

1. The flux in the PP-Pathway was modeled with a reaction from G6P to 6PGL, replacing the black-box assumption of the above model. The flux back to glycolysis was modeled by the reactions from 6PG $\rightarrow$ GAP and 6PG $\rightarrow$ PYR.

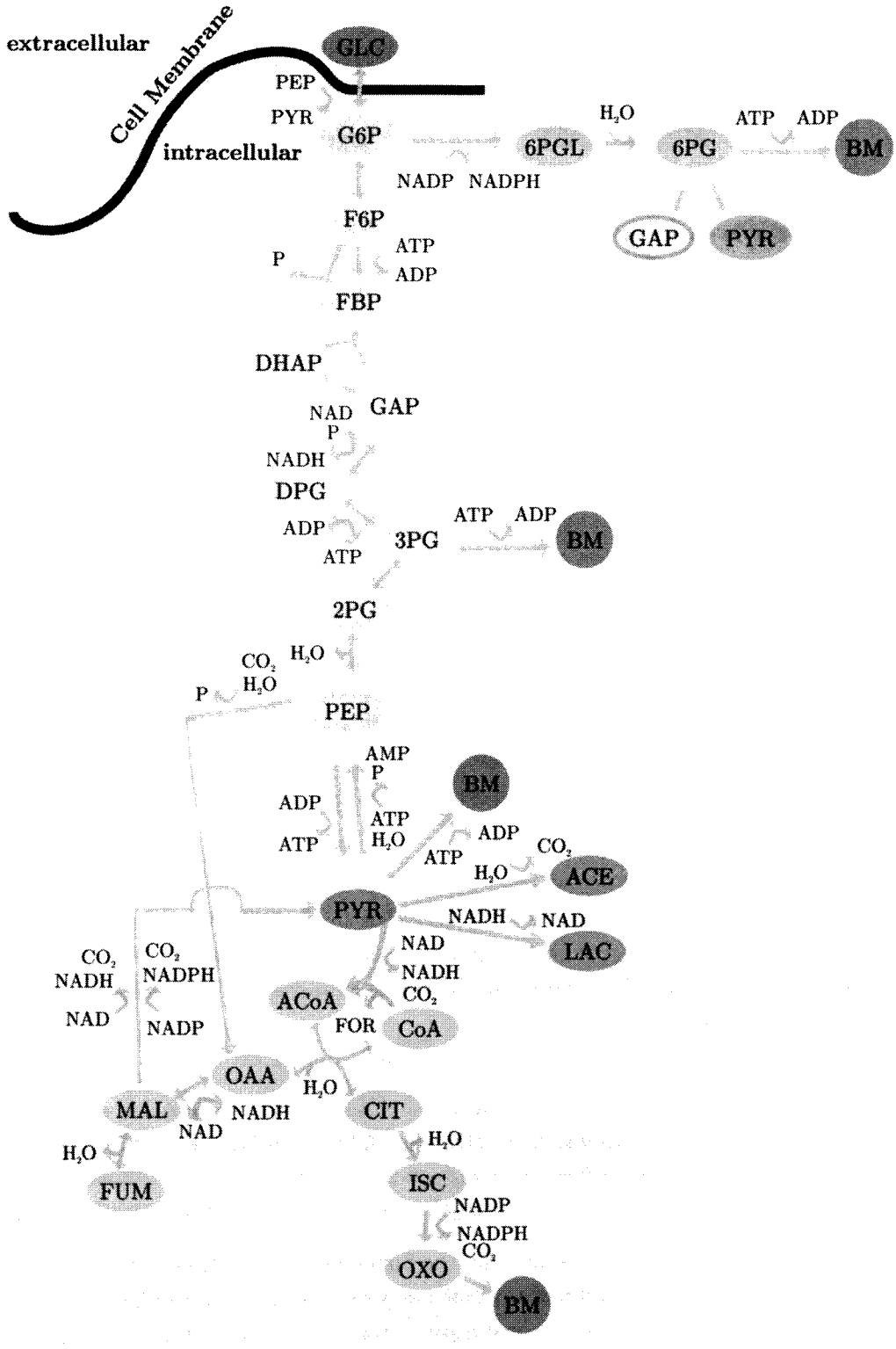


Figure 6.12.: Pathway of the extended model with 26 dynamic variables.

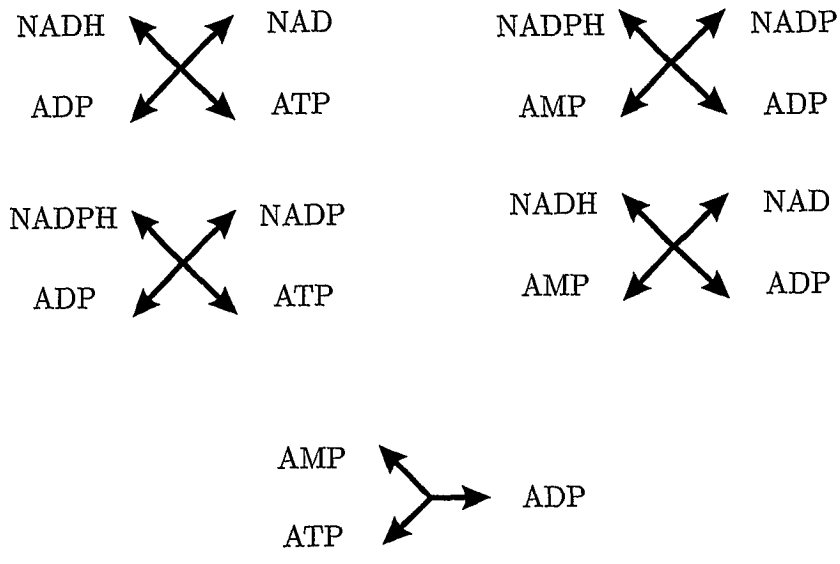


Figure 6.13.: Reactions for the nucleotides included in the extended model. NADH and NADPH each are assumed to convert 3 ADP or 3 AMP into 3 ATP respectively 3 ADP. AMP and ATP can be converted into 2 ADP. These reactions simulate the respiratory chain and energy conversion reactions in *Escherichia coli.*

2. The TCA-cycle was assumed to be active in the branched form, i.e. the cycle is not closed. See Figure 6.12, where no connection between OXO and FUM exists. For growth of *Escherichia coli* this is supposed to be the form of action of the cycle [65].

3. In Figure 6.13 all reactions for interconversion of the nucleotides are shown. Reactions for the production of ADP and ATP from NADH, NADPH, AMP, and ADP have been assumed for energy balancing.

4. Biomass production represented by the fluxes from 6PG, 3PG, PYR and OXO to biomass has been assumed to consume ATP. For the reaction from 3PG and PYR to biomass, the ATP consumption was a parameter that was adjusted during the process of parameter fitting. NAD, NADP and NADH, NADPH were not assumed to be necessary for biomass production but are only included in the reactions in the central metabolic pathway.

Equation 6.9 for the PT-System has been changed to the following form:

$$v_{\mathrm{GLC}\rightarrow\mathrm{G6P}} = v_{max} \cdot \frac{K_1 \cdot \mathrm{PEP} \cdot \mathrm{GLC} - K_2 \cdot \mathrm{G6P} \cdot \mathrm{PYR}}{(1 + \frac{\mathrm{PYR}}{K_I}) \cdot P} \tag{6.12}$$

with

$$\begin{aligned} P &= 1 + K_2 \cdot \mathrm{PEP} + K_3 \cdot \mathrm{GLC} + K_4 \cdot \mathrm{PEP} \cdot \mathrm{GLC} \\ &\quad + K_5 \cdot \mathrm{G6P} + K_6 \cdot \mathrm{PYR} + K_7 \cdot \mathrm{G6P} \cdot \mathrm{PYR} \end{aligned} \tag{6.13}$$

The equation includes a product inhibition by PYR of the PT-System. Because the phosphotransfer reactions of the uptake system are reversible, this assumption has a biological background. However, kinetic constants of the system from [82] mentioned earlier suggest that the inhibition is not strong.

Equation 6.5 for the reaction from F6P to FBP was replaced by an equation from [47]. The equation is based on a four state allosteric enzyme model and includes AMP and ADP as activators and ATP as an inhibitor. The equation takes the following form:

$$\begin{aligned} v_{\mathrm{F6P}\rightarrow\mathrm{FBP}} &= v_{max} \cdot \frac{\mathrm{ATP}}{\mathrm{ATP} + K_{m,\mathrm{ATP}} \cdot (1 + \frac{\mathrm{ADP}}{K_{A,\mathrm{ADP}}})} \cdot \\ &\quad \frac{\mathrm{F6P}}{(\mathrm{F6P} + P) \cdot (1 + \frac{L}{(1+\frac{\mathrm{F6P}}{P})^8})} \end{aligned} \tag{6.14}$$

where

$$P = K_{m,\mathrm{F6P}} \cdot \frac{1 + \frac{\mathrm{ATP}}{K_{1,\mathrm{ATP}}} + \frac{\mathrm{ADP}}{K_{1,\mathrm{ADP}}} + \frac{\mathrm{AMP}}{K_{1,\mathrm{AMP}}}}{1 + \frac{\mathrm{ADP}}{K_{2,\mathrm{ADP}}} + \frac{\mathrm{AMP}}{K_{2,\mathrm{AMP}}}} \tag{6.15}$$

For the reaction rate from 6PG to biomass and from PYR to biomass, the following equation for synthesis of monomeric building blocks was used [5]:

$$v_{\mathrm{6PG}\rightarrow\mathrm{BM}} = v_{max} \cdot \frac{\mathrm{6PG}}{K_{m,\mathrm{6PG}}} \cdot \frac{\mathrm{ATP}}{K_{m,\mathrm{ATP}}} \tag{6.16}$$

A hypothetical equation was used for the reaction from PYR to ACoA. The equation is derived from the ordered Bi-Bi-Kinetics but includes CoA as a third substrate:

$$\mathbf{v}_{\mathrm{PYR}\rightarrow\mathrm{ACoA}} = \frac{v_f \cdot (\mathrm{PYR}\cdot\mathrm{NAD}\cdot\mathrm{CoA} - \frac{\mathrm{ACoA}\cdot\mathrm{NADH}}{K_{eq}})}{P} \tag{6.17}$$

$$\begin{aligned} P &= \mathrm{PYR}\cdot\mathrm{NAD}\cdot(1+\frac{\mathrm{ACoA}}{K_{i,\mathrm{ACoA}}}) + K_{m,\mathrm{NAD}}\cdot(\mathrm{PYR}+K_{i,\mathrm{PYR}}) + \ldots \\ &+ K_{m,\mathrm{PYR}}\cdot\mathrm{NAD} + K_{m,\mathrm{CoA}}\cdot\mathrm{CoA} + \ldots \\ &+ \frac{v_f}{v_r\cdot K_{eq}}\cdot\left[K_{m,\mathrm{ACoA}}\cdot\mathrm{ACoA}\cdot(1+\frac{\mathrm{PYR}}{K_{i,PYR}}) + \ldots\right. \\ &+ \left.\mathrm{NADH}\cdot\left(K_{m,\mathrm{ACoA}}\cdot(1+\frac{K_{m,\mathrm{PYR}}\cdot\mathrm{NAD}}{K_{iA}\cdot K_{mB}}) + \mathrm{ACoA}\cdot(1+\frac{\mathrm{NAD}}{K_{i,\mathrm{NAD}}})\right)\right] \end{aligned}$$

Reaction kinetics of the model are summarized in Table 6.2.

Table 6.2.: All reactions of the model and the rate equations used in the mathematical model.

Reaction rates $\mathbf{v}_i$	Rate equation
GLC + PEP $\leftrightarrow$ G6P + PYR	see text
G6P $\leftrightarrow$ F6P	reversible Michaelis-Menten
G6P $\rightarrow$ 6PGL	ordered Bi-Bi-Kinetic
6PGL $\leftrightarrow$ 6PG	reversible Michaelis-Menten
6PG $\leftrightarrow$ PYR	reversible Michaelis-Menten
6PG $\leftrightarrow$ GAP	reversible Michaelis-Menten
6PG + ATP $\leftrightarrow$ BM + ADP	see text
F6P + ATP $\leftrightarrow$ FBP + ADP	see text
FBP $\rightarrow$ F6P	Hill-Kinetic
FBP $\leftrightarrow$ DHAP + GAP	ordered Uni-Bi-Kinetic
DHAP $\leftrightarrow$ GAP	reversible Michaelis-Menten

continued on next page

Table 6.2.: (continued)

Reaction rates $\mathbf{v}_i$	Rate equation
GAP + NAD $\leftrightarrow$ DPG + NADH	ordered Bi-Bi-Kinetic
2PG + ADP $\leftrightarrow$ 3PG + ATP	ordered Bi-Bi-Kinetic
3PG $\leftrightarrow$ 2PG	reversible Michaelis-Menten
3PG + ATP $\leftrightarrow$ BM + ADP	ordered Bi-Uni-Kinetic
2PG $\leftrightarrow$ PEP	reversible Michaelis-Menten
PEP + ADP $\leftrightarrow$ PYR + ATP	ordered Bi-Bi-Kinetic
PYR + ATP $\leftrightarrow$ PEP + AMP	ordered Bi-Bi-Kinetic
PYR + NADH $\leftrightarrow$ LAC + NAD	ordered Bi-Uni-Kinetic
PYR $\rightarrow$ ACE	Hill-Kinetic
PYR + ATP $\leftrightarrow$ BM + ADP	see text
PYR + NAD + CoA $\leftrightarrow$ ACoA + NADH	see text
MAL + NAD $\leftrightarrow$ PYR + NADH	ordered Bi-Bi-Kinetic
MAL + NADP $\leftrightarrow$ PYR + NADPH	ordered Bi-Bi-Kinetic
MAL + NAD $\leftrightarrow$ OAA + NADH	ordered Bi-Bi-Kinetic
PEP $\rightarrow$ OAA	see text in the model of the last section
OAA + ACoA $\leftrightarrow$ CIT + CoA	ordered Bi-Bi-Kinetic
CIT $\leftrightarrow$ ISC	reversible Michaelis-Menten
ISC + NADP $\leftrightarrow$ OXO + NADPH	ordered Bi-Bi-Kinetic
MAL $\rightarrow$ BM	Hill-Kinetic
MAL $\rightarrow$ FUM	Hill-Kinetic
ATP + AMP $\leftrightarrow$ ADP	ordered Bi-Uni-Kinetic
ADP + NADH $\leftrightarrow$ ATP + NAD	ordered Bi-Bi-Kinetic
ADP + NADPH $\leftrightarrow$ ATP + NADP	ordered Bi-Bi-Kinetic
AMP + NADH $\leftrightarrow$ ADP + NAD	ordered Bi-Bi-Kinetic
AMP + NADPH $\leftrightarrow$ ADP + NADP	ordered Bi-Bi-Kinetic

6.5.1. Fitting the model to data from a pulse experiment

The presented model includes 26 dynamic variables and 302 parameters. The only data input to the simulation is the extracellular GLC concentration and initial conditions for the compounds. All other compound concentrations are calculated from the model equations. Solutions of the parameter fitting problem were intended to give insight in the dynamic behavior of the system after the GLC pulse. Due to the low number of data points compared to the number of parameters, an parameter fitting is difficult. During the parameter fitting, the weights for the error functional have been set to high values for G6P and PEP. For these metabolites, weights factors 100 to 500 times larger than for all other metabolites have been used. During the parameter fitting using the SUBPLEX algorithm of the MMT software, the following strategies have been applied:

- Scale parameter changes by their sensitivities as described in Section 5.4.
- Restart the parameter fitting and increase ranges of fitted initial conditions and parameters.
- Perform the parameter fitting stepwise, i.e. fit for the first 5 seconds of the steady state. Then use this result to fit for the first 8 seconds that includes the steady state and the initial reaction of the system to the GLC pulse. Continue to increase the time window after the parameter fitting is finished.

Using this strategy, a reasonable fit of the simulated G6P and PEP concentrations to the measured values could be achieved for the time frame -5 to 25 seconds. Figure 6.14 shows the best result obtained with parameter fitting. Figure 6.15 shows the other compounds of the model where experimental data was available. Due to the low priority during the parameter fitting, the simulated metabolite concentrations only remain in a reasonable range if compared to the experimental data. Figure 6.16 shows the GLC uptake rate and the fluxes around G6P.

If the parameter fitting was extended to a time frame of more than -5 to 25 seconds, the model failed to represent the experimental data. Therefore, the model is not capable to explain the complete dynamics of the experiment.

6.5.2. Conclusions and summary of the model

The model shown was able to describe well the intracellular G6P and PEP concentrations for the time -5 to 25 seconds. The following conclusion are made:

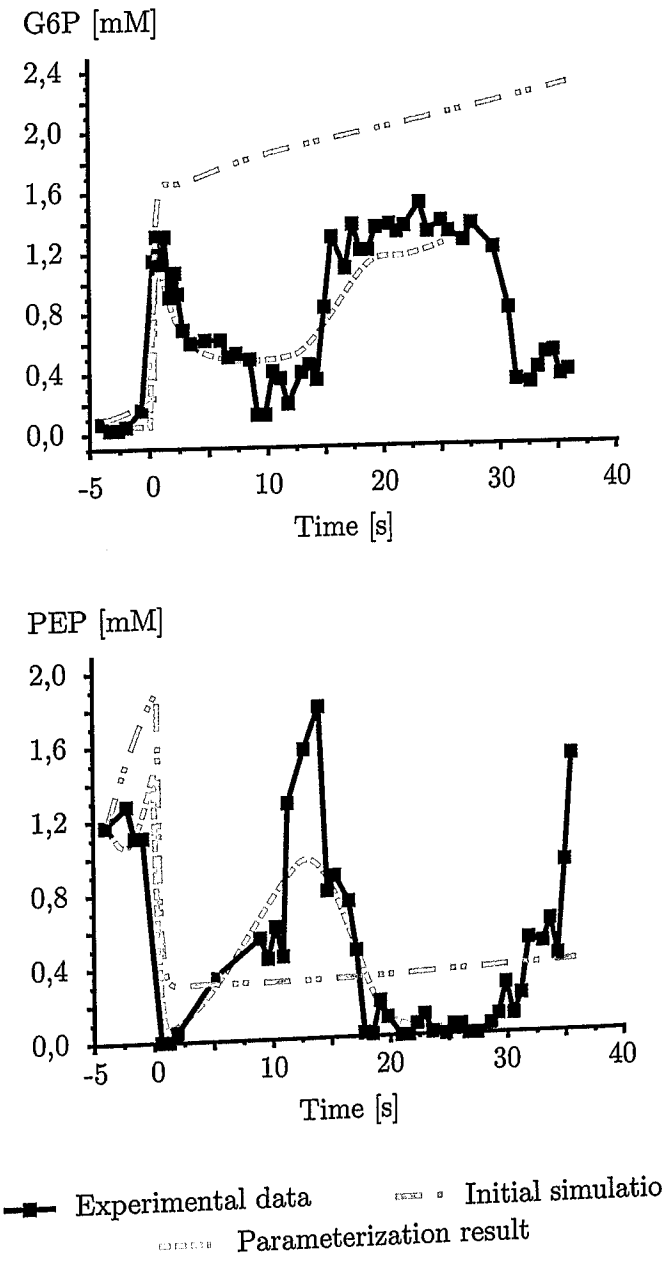


Figure 6.14.: Initial simulation results and results obtained from parameter fittings. parameter fittings for a time window larger then -5 to 25 seconds was not successful.

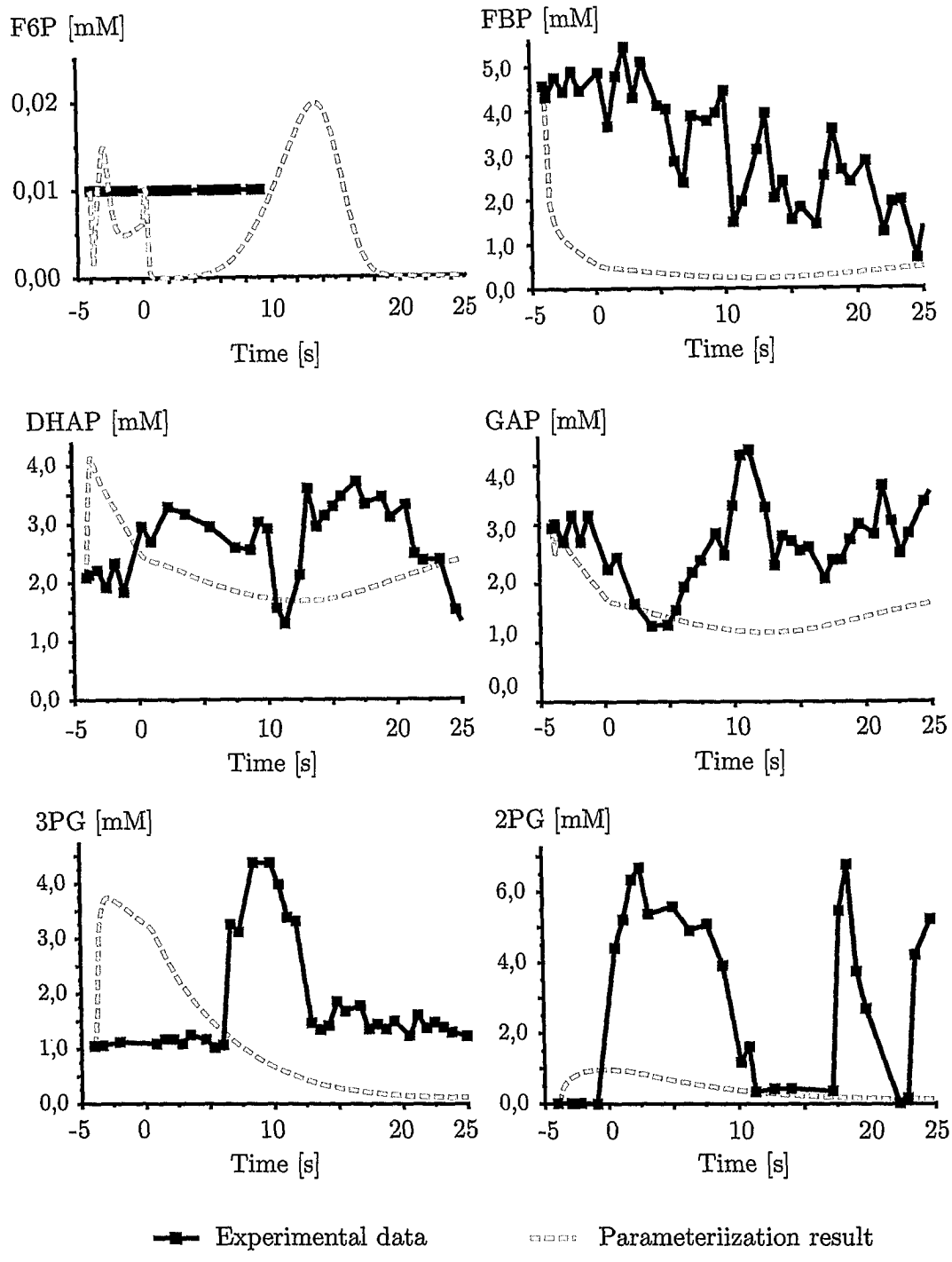


Figure 6.15.: Results obtained from parameter fittings for metabolites with low weighting factor in the error functional.

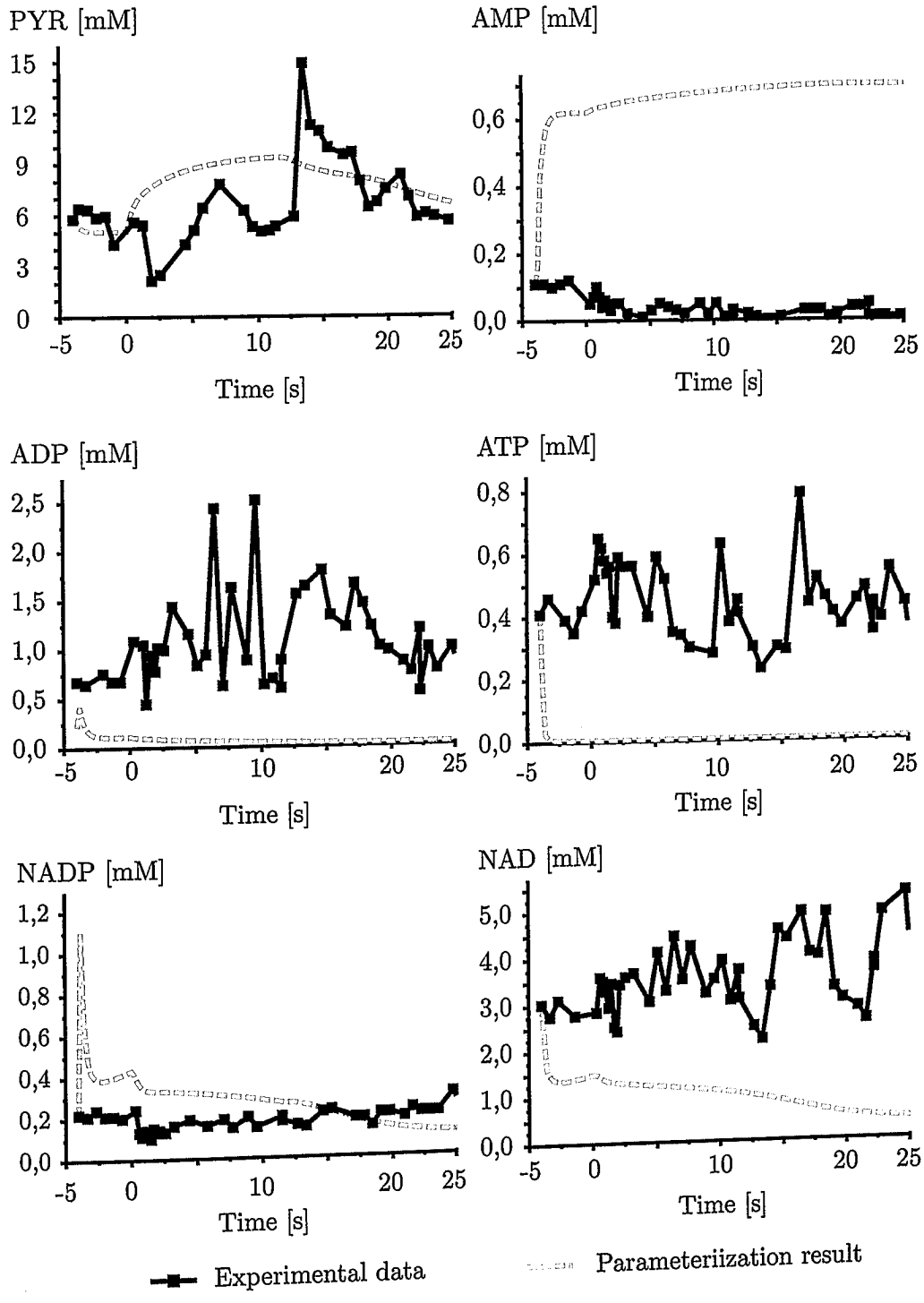


Figure 6.15, continued

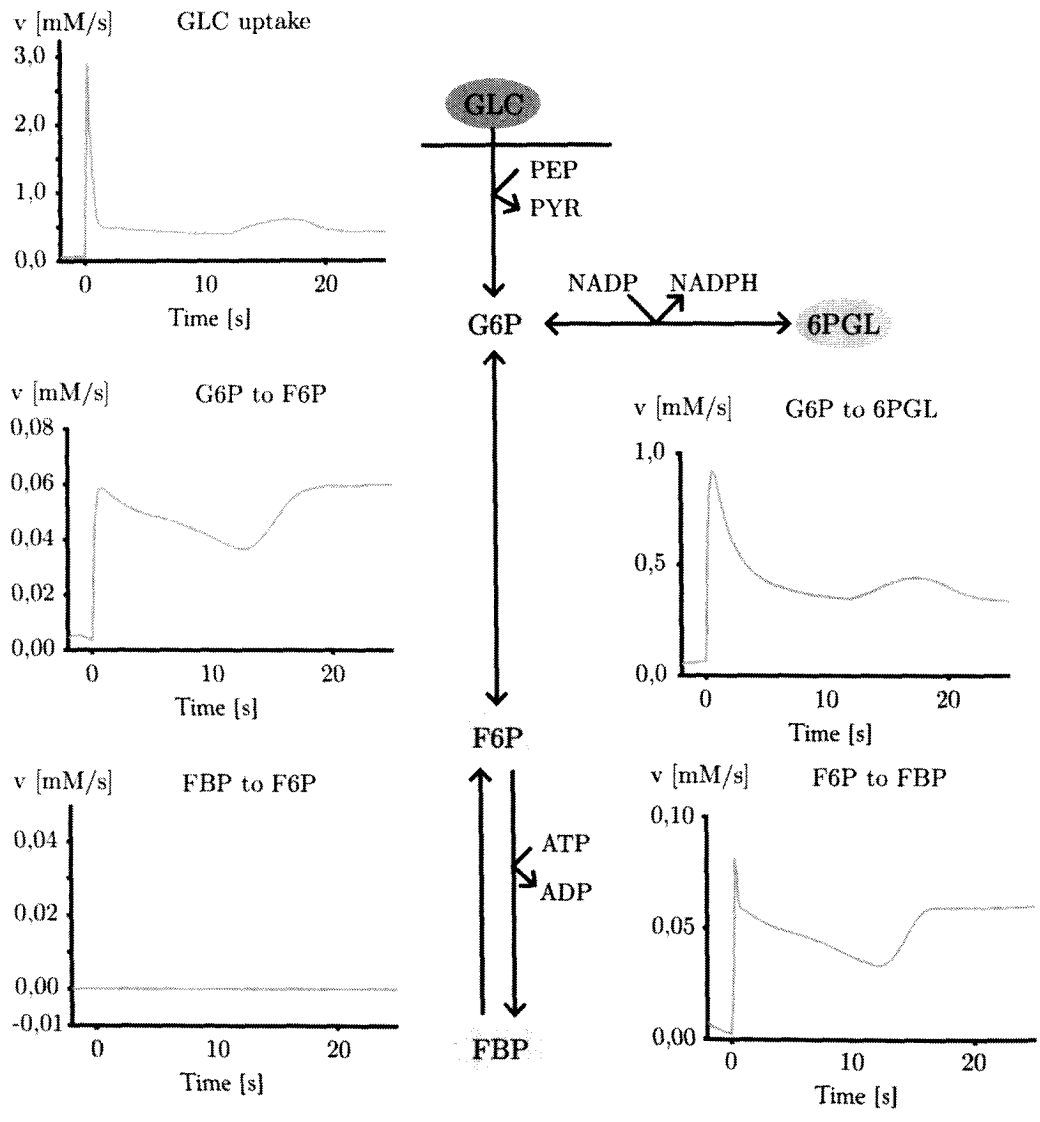


Figure 6.16.: Fluxes of GLC uptake, the reactions from G6P to the PP-Pathway and into glycolysis.

1. The model describes the dynamic behavior of G6P and PEP after the GLC pulse. This includes the increase of the PEP concentration at time 10 seconds and the high G6P concentration for low PEP concentrations in the time frame 18 to 25 seconds.

2. Analysis of the fluxes that occur in the simulation shows that the flux through the PP-Pathway is very high (Figure 6.16). During the steady state, about 50% of the GLC that is taken up enters the PP-Pathway whereas the other 50% are processed by glycolysis. After the GLC pulse, however, the flux through glycolysis remains almost the same whereas the flux through the PP-Pathway increases almost 10-fold. Therefore about 90% of the GLC taken up is processed by the PP-Pathway.

3. The increase in PEP after the initial drop is a result of a high flux from PYR to PEP. There was no net-flux from OAA to PEP in the model. Instead, a net-flux in the other direction resulted from the parameter fitting. Additionally, a strong flux was present from OAA to MAL and from MAL to PYR. Therefore OAA is one of the sources for the PEP production. This conversion from OAA to MAL and further to PYR and PEP is already present during the steady state. However, the rate of these anaplerotic reactions is estimated to high because the OAA concentration decreases already during the steady state. In general OAA may be important to provide PEP under certain environmental conditions. Intracellular measurements of OAA could help to answer the question of intracellular OAA concentrations but no data was available for this parameter fitting.

4. The high G6P concentration that coincides with the low PEP concentration between time 18 and 25 seconds is an effect of a lower flux from G6P to 6PGL. As mentioned above, about 90% of the GLC taken up enters the PP-Pathway through this reaction in the simulations. This reaction is NADP dependent. NADP decreases after the pulse and may be a reason that leads to a decrease in the reaction rate from G6P to 6PGL. The high G6P concentration at time 18-25 second follows.

5. Simulation of the nucleotide concentrations AMP, ADP, ATP, NAD, NADH, NADP, and NADPH were problematic. Finding a balance between the reactions is difficult. Either a situation occurred where high ADP, AMP, NADH, and NADPH concentrations and low ATP, NAD, and NADP concentrations were present or the contrary. Even though the results suggest that the decreasing NADP concentration is a reason for the decreased flux from G6P to 6PGL, a correct model for nucleotide concentrations has not been found yet. Also, the nucleotide concentrations that could be measured show large errors and it is not clear if dynamic changes in these concentrations are present at all.

7. Summary and Future Development

The large increase in available quantitative and qualitative knowledge has been mentioned at the beginning of this work. The data results from genomics, proteomics, transcriptomics, metabolomics and fluxeomics. These techniques are more and more automated and create large amount of data that has to be analyzed.

During this work, a modeling framework was developed. The approach is based on a Metabolic Modeling Tool (MMT) software. The MMT software was especially developed for a specific kind of models for chemical reaction kinetics: non-stationary pathway models of microorganisms based on chemical reaction kinetics. The models lead to a system of ODEs. Simulations require then the solution of an IVP and parameter fittings require numerical algorithm due to the non-linearity of the ODEs. These algorithms are also part of the MMT software.

All information of the models, including the chemical reaction networks, mathematical equations, parameters, and results of simulations and parameter fittings, are stored in a relational database. The software allows to build new models or to modify existing models in a very short time. Because the software sets up the IVP automatically and creates a C-code using MAPLE 6, the user only needs to define the model and specify the parameters. The approach significantly reduces the number of errors usually made when typing equations by hand into a C-code.

Summary of the presented models

It is demonstrated in this work how the software is set up and what the advantages are. An example is given that illustrates the database setup and the mathematical problem that results from the modeling approach. An example is also given to illustrate the problems of fitting a complex pathway model to experimental data.

Models of different dimension are shown. The results show that the MMT software is useful

- for modeling chemical reaction kinetics of different dimensions.
- for simulations with defined models.
- to fit models with several hundred parameters to experimental data.
- to define a large amount of models in a short time and compare their results obtained through parameter fitting.

The analyzed and developed models show that

- the exact knowledge of metabolic control effects as inhibitions and activations has very strong effects on the dynamics of a metabolic pathway.
- small models are able to describe phenomena of isolated systems well.
- large models are able to describe some effects at least in a qualitative way but produce questionable results due to the lack of available data, e.g. for the flux through the PP-Pathway.

Project continuation

The Institute of Biotechnology is further developing the MMT software. A parallel analysis of competing models will be included in the near future. Further, the performance of the software will be increased by parallelization methods and recoding of parts of the routines. Two major points for improvement have already been mentioned. One is the storage of results. Instead of storing all results, the C-code produced by the MMT and parameter values should be stored. This saves the large amount of time needed for the setup of the system of ODEs from the database content as well as the storage space needed for the simulation results. Results can then be reproduced in a few seconds. The second point is the inclusion of smoothing splines for experimental data.

In cooperation with the University of Siegen a project "Modellierung Metabolischer Netzwerke auf der Grundlage von *in vivo*, *invitro* and *in silico* Daten", financed by the Deutsche Forschungsgemeinschaft (DFG), will begin this year. During the project, an extended software tool for models of cellular systems will be developed.

Additionally, experiments with new and existing strains are performed and analyzed. Methods from genomics, proteomics, transcriptomics, fluxeomics, and metabolomics are used at the Institute of Biotechnology. An additional focus is the optimization of fermentation conditions for the new strains. The data generated from metabolomics

will be analyzed with the MMT software. For model set up and parameter identification data from the other methods will be helpful.

The future of metabolic engineering

In the future, tools similar to the MMT software will be used to analyze the large amount of available data. These tools will be able to describe cellular systems with increasing accuracy and detail. A prerequisite is, however, continued effort from biologists, chemists, engineers, mathematicians, physicists, and computer scientists in this field. A close cooperation of scientists from these research areas is absolutely necessary for further progress.

A second prerequisite is a continued inflow of money that is necessary for support of such research activities. Methodological projects in the field of biotechnology usually require high investments over a long period of time. Because many products in the field are used in medical treatments, significant revenues in this field are likely to produce a continued inflow of money for research projects.

A final comment

During the time this thesis is written there are already new methods for analyzing microorganisms coming into focus. A method termed interactomics is used with increasing success to examine all possible interactions between proteins in a cell. The method is based on analysis with coimmunoprezipitations or the Yeast-Two Hybrid-Analysis [30]. For *Escherichia coli*, 5.6 Billion so called "Bait-Prey" combinations have been tested with the Yeast-Two Hybrid-System and 1200 protein-protein interactions have been identified.

Another promising field is the single-molecule analysis [92]. The method is based on fluorescent analysis of molecules stimulated by a laser beam. Thus intracellular concentrations of molecules can be measured with high accuracy. Additionally, the method allows to view the location of molecules inside a cell. The method has a great potential for analyzing compartmentation in cells.

Interactomics increases the complexity of models significantly because protein-protein interactions have to be included in models for an exact description of regulatory effects. Data about distribution of molecules inside cells will lead to models that include spatial and conceptual compartmentation of cells. These two examples show how fast the field of metabolic engineering is progressing and again empha-

sizes the need of software tool for developing complex models of chemical reaction kinetics.

Bibliography

[1] R.D. Appel, A. Bairoch, and D.F. Hochstrasser. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server (http://www.expasy.ch/). *Trends Biochem Sci.*, 19(6):258–260, 1994.

[2] A.S. Asratian, T.M.J. Denley, and R. Haggkvist. *Bipartite Graphs and their Applications. Cambridge Tracts in Mathematics, No. 131.* Cambridge University Press, 1998.

[3] J.E. Bailey. Toward a science of Metabolic Engineering. *Science-Wash.*, 252(5013):1668–1675, 1991.

[4] J.E. Bailey and D.F. Ollis. *Biochemical engineering fundamentals.* McGraw-Hill, second edition, 1986.

[5] M. Baltes. *Entwurf eines strukturierten Modells zur dynamischen Simulation des Katabolismus von Saccharomyces cerevisiae.* VDI-Verlag Düsseldorf, Reihe, Vol. 17, Biotechnik, No. 148., Universität Stuttgart, Germany, 1996.

[6] B.A. Barshop, R.F. Wrenn, and C. Frieden. Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM - a flexible, portable system. *Anal Biochem.*, 130(1):134–145, 1983.

[7] D. Bingmann, H.G. Lipinski, G. Hagemann, E.J. Speckmann, and P. Tetsch. Time courses of lidocaine effects on sodium membrane currents in small and large neurons. *Gen Physiol Biophys.*, 9(4):331–342, 1990.

[8] D. Blangy, H. Buc, and J Monod. Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli* . *J Mol Biol.*, 31:13–35, 1968.

[9] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* K-12. *Science*, 5, 277(5331):1453–74, 1997.

[10] H.P.J. Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trend Biotechnol.*, 15(8):308–314, 1997.

[11] BRENDA. The enzyme database. http://www.brenda.uni-koeln.de/.

[12] N. Brinkmann. Personal communication, paper in preparation.

[13] A. Buchholz. *Reaktionstechnische Untersuchungen zur Dynamik intrazellulärer Metabolite in glycerollimitierten* Escherichia coli *K12.* Diplomarbeit. Mathematisch-Naturwissenschaftliche Fakultät der Rheinischen-Friedrich-Wilhelms-Universität Bonn, 1998.

[14] A. Buchholz, J. Hurlebaus, R. Takors, and C. Wandrey. Metabolomics: quantification of intracellular metabolite dynamics. Submitted.

[15] S.J. Cain and P.C. Chau. Transition probability cell cycle model with product formation. *Biotechnol Bioeng.*, 58(4):387–399, 1998.

[16] S. Cayley, B.A. Lewis, H.J. Guttman, and M.T. Record Jr. Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *J Mol Biol.*, 222(2):281–300, 1991.

[17] F.E. Cellier. *Continous system modeling.* Springer, 1991.

[18] Y-P. Chao and J.C. Liao. Metabolic responses to substrate futile cycling in *Escherichia coli. J Biol Chem.*, 269(7):5122, 1994.

[19] J.M. Chushing. *An introduction to structured population dynamics.* Cbms-Nsf Regional Conference Series in Applied Mathematics, 71.Society for Industrial & Applied Mathematics, 1998.

[20] B.L. Clarke. Stoichiometric network analysis. *Cell Biophys.*, 12:237–253, 1988.

[21] A Cornish-Bowden. *Fundamentals of Enzyme Kinetics.* Portland Press, 1995.

[22] A. Cornish-Bowden and M.L. Cardenas, editors. *Control of metabolic processes*, volume 190. NATO ASI Series A: Life Sciences, Plenum Press, 1990.

[23] A. Cornish-Bowden and J.H. Hofmeyr. `MetaModel`: a program for modelling and control analysis of metabolic pathways on the IBM PC and compatibles. *Comput Appl Biosci.*, 7(1):89–93, 1991.

[24] C.F. Curtiss and J.O. Hirschfelder. Integration of stiff equations. *Proc Nat Acad Sci.*, 38:235–243, 1952.

[25] Q. Dang and C. Frieden. New PC versions of the kinetic simulation and fitting programs, `KINSIM` and `FITSIM`. *Trends Biochem Sci.*, 22(8):317, 1997.

[26] J.E. Dennis, D.M. Gay, and R.E. Welsch. An adaptive nonlinear least-squares algorithm. *ACM Trans Math Software*, 7(3):348–368, 1981.

[27] M.M. Domach, S.K. Leung, R.E. Cahn, G.G. Cocks, and M.L. Shuler. Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-S. *Biotechnol Bioeng.*, 36(3):203–216, 1984.

[28] E. coli Index. The E. coli Index: Part of the WWW Virtual Library. http://web.bham.ac.uk/bcm4ght6/res.html.

[29] M. Ehlde and G. Zacchi. MIST: a user-firendly metabolic simulator. *Comput Appl Biosci.*, 11(2):201–207, 1995.

[30] S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.

[31] C.J. Gadgil, P.J. Bhat, and K.V. Venkatesch. Cybernetic model for the growth of *Saccharomyces cerevisiae* on melibiose. *Biotechnol Prog.*, 12(6):744–750, 1996.

[32] E.D. Gilles, M. Reuss, J. Wang, and K. Mauch. Rechnergestütze Stoffwechselmodellierung bioverfahrenstechnischer Prozesse. http://www.isr.uni-stuttgart.de/sfb412/sfb412_tp_a3.html, 1997.

[33] M.L. Giuseppin and N.A. van Riel. Metabolic modeling of *Sacharomyces cerevisiae* using the optimal control of homeostasis: a cybernetic model definition. *Metab Eng.*, 2(1):14–33, 2000.

[34] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley, 1953.

[35] Goldbeter. *Biochemical oscillations and cellular rhythms.* Cambridge University Press, 1996.

[36] A Goldbeter and R Lefever. Dissipative structures for an allosteric model. application to glycolytic oscillations. *Biophys J.*, 12(10):1302–1315, 1972.

[37] I. Goryanin, T.C. Hodgman, and E. Selkov. Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics*, 9, 1999.

[38] van W.M. Gulik, ten H. Hoopen, and J.J. Heijnen. A structured model describing carbon and phosphate limited growth of *Catharanthus roseus* plant cell suspensions in batch and chemostat culture. *Biotechnol Bioeng.*, 41(8):771–780, 1993.

[39] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II*. Springer Verlag, 1991.

[40] V. Hatzimanikatis and K.H. Lee. Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab Eng.*, 1(4):275–281, 1999.

[41] A. Heim and O. Stryk. *Documentation of* PAREST - *A multiple shooting code for optimization problems in differential-algebraic equations.* Fakultät für Mathematik, Technische Universität München, report tum-m96-16 edition, 1996.

[42] R. Heinrich and T.A. Rapoport. A linear steady state treatment of enzymatic chains. general properties, control and effector strength. *Eur J Biochem.*, 42:89–95, 1974.

[43] R. Heinrich and S. Schuster. *The regulation of cellular systems.* Chapman and Hall, NY, 1996.

[44] C.J. Hewitt, G.N. Caron, B. Axelsson, C.M. McFarlane, and A.W. Nienow. Studies related to the scale-up of high-cell density *E.coli* fed-batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnology and Bioengineering*, 70(4):381–390, 2000.

[45] A.C. Hindmarsh. LSODE and LSODI, two new initial value ordinary differential equation solvers. *ACM-SIGNUM Newsletter*, 15:10–11, 1980.

[46] R. Hofestädt. Modelling molecular counters with stochastic cellular grammars. *Informatik Biometrie und Epidemiologie in Medizin und Biologie*, 4:169–177, 1993.

[47] E. Hofmann and G. Kopperschläger. Phosphofructokinase in yeast. In W.A. Wood, editor, *Methods in enzymology*, pages 49–60. Academic Press, New York, 1982.

[48] A. Holmberg. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Math Biosci.*, 62:23–43, 1982.

[49] D.H. Irvine, E. Voit, and M.A. Savageau. Analysis of complex dynamic networks with ESSYNS. A program for the efficient solution of nonlinear differential equations in S-systems form. In E.O. Voit, editor, *Canonical Nonlinear Modeling. S-system approach to understanding complexity.*, pages 133–141. Van Nostrand Reinhold, 1991.

[50] H. Kacser and J.A. Burns. The control of flux. *Symp Soc Exp Biol.*, 27:65–104, 1973.

[51] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 28:29–34, 2000.

[52] KEGG. Kyoto Encyclopedia of Genes and Genomes. http://www.genome.ad.jp /kegg/.

[53] S.T. Kellogg, D.K. Chatterjee, and A.M. Charkrabarty. Plasmid-assisted molecular breeding: new technique for enhanced biodegradation of persistent toxic chemicals. *Science*, 214(4525):1133–1135, 1981.

[54] W. Kinnebrock. *Optimierung mit genetischen und selektiven Algorithmen.* Oldenbourg Verlag, 1994.

[55] A. Kremling, K. Jahreis, J.W. Lengeler, and E.D. Gilles. The organization of metabolic reaction networks: a signal-oriented approach to cellular models. *Metab Eng.*, 2(3):190, 2000.

[56] H. Kuo and J.D. Keasling. A Monte Carlo simulation of plasmid replication during the bacterial division cycle. *Biotechnol Bioeng.*, 52:633–647, 1996.

[57] P. Kuzmic. Program `DYNAFIT` for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem.*, 237:260–273, 1996.

[58] L. Laffend and M.L. Shuler. Structured model of genetic control via the lac promoter in *Escherichia coli* . *Biotechnol Bioeng.*, 43:399–410, 1994.

[59] B. Lee, J. Yen, L. Yang, and J.C. Liao. Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnol Bioeng.*, 62(6):722, 1999.

[60] J.W. Lengeler, G. Drews, and H.G. Schlegel. *Biology of the Prokaryotes.* Georg Thieme Verlag, 1999.

[61] N.V. Mantzaris, J.-J. Liou, P. Daoutidis, and F. Srienc. Numerical solution of a mass structured cell population balance model in an envrionment of changing substrate concentration. *J Biotechnol.*, 71(1-3):157–174, 1999.

[62] J Marsch, J Fritze. *SQL - eine praxisorientierte Einführung* . Vieweg und Sohn, 1993.

[63] P. Mendes. `GEPASI`: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comp Appl Biosci.*, 9(5):563–571, 1993.

[64] J.J. Moré. The Levenberg-Marquardt algorithm: Implementation and theory. In G.A. Watson, editor, *Numerical Analysis. Proceedings Biennial Conference Dundee 1977*, volume 630 of *Lecture Notes in Mathematics*, pages 105–116. Springer-Verlag, Berlin, 1978.

[65] F.C. Neidhardt. *Escherichia coli and Salmonella : Cellular and Molecular Biology.* American Society for Microbiology, 1999.

[66] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[67] J Nielsen, J Villadsen. *Bioreaction engineering principles.* Plenum Pub Corp, 1994.

[68] M. Otter. MODELICA - die Modellierungssprache der nächsten Generation für den Modellaustausch. In A. Kuhn and S. Wenzel, editors, *Simulationstechnik. 11. Symposium in Dortmund, November 1997, Volume 11 of Fortschritte in der Simulationstechnik*, pages 212–217. Vieweg, 1997.

[69] M. Papageorgiou. *Optimierung: Statische, dynamische, stochastische Verfahren für die Anwendung.* Oldenbourg Verlag, 1991.

[70] L Perko. *Differential Equations and dynamical systems.* Springer Verlag, 1986.

[71] L.R. Petzold and A.C. Hindmarsh. Livermore solver for ordinary differential equations. www.netlib.org.

[72] Pfeiffer, T. and Sánches-Valdenbro, I. and Nuño, J.C., and Montero, F. and Schuster, S. METATOOL: For studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.

[73] P.W. Postma, J.W. Lengeler, and G.R. Jacobsen. Phosphoenolpyruvate:Carbohydrate phosphotransferase systems. In F.C. Neidhardt, editor, Escherichia coli and Salmonella. *Second Edition.*, chapter 75. ASM Press, Washington, 1996.

[74] T. Preiss and F. Breitenecker. Definition of a formal language to describe models by means of relations. In Möller, P.F., editor, *ESS 2000 in cooperation with ASIM 2000: Simulation in industry. 12th eruopean simulation symposium.* Society for Computer Simulation International, 2000.

[75] N.V. Reddy, Mavrovouniotis M.L., and M.N. Liebman. Petri net representations in metabolic pathways. In L. Hunter, D. Searls, and J. Shavlik, editors, *ISMB-93, Proceedings of the First International Conference on Intelligent Systems for Molecular Biology.*, pages 328–336. AAAI Press, 1993.

[76] B. Richey, D.S. Cayley, M.C. Mossing, C. Kolka, C.F. Anderson, T.C. Farrar, and M.T. Record Jr. Variability of the intracellular ionic environment of *Escherichia coli. J Biol Chem.*, 262:7157–7164, 1987.

[77] O. Richter, A. Betz, and C. Giersch. The response of oscillating glycolysis to perturbations in the NADH/NAD system: a comparision between experiments and a computer model. *BioSystems*, 7:137–146, 1975.

[78] T. Ritzau. `mysqltcl`. Can be obtained from `MySQL` homepage http://www.mysql.com.

[79] M Rizzi, M Baltes, U Theobald, and M Reuss. In-vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnol Bioeng.*, 55(4):592–608, 1997.

[80] M. Rizzi, Baltes M., U. Theobald, and M. Reuss. In-vivo analysis of metbolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnol Bioeng.*, 55(2):305–316, 1997.

[81] M. Rizzi, U. Theobald, E. Querfurth, T. Rohrhirsch, M. Baltes, and M. Reuss. In vivo investigations of glucose transport in saccharomyces cerevisiae. *Biotechnol Bioeng.*, 49:316–327, 1996.

[82] J.M. Rohwer, N.D. Meadow, S. Roseman, H.V. Westerhoff, and P. Postma. Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements in vitro. *J Biol Chem.*, 275(45):34909–34921, 2000.

[83] J.M. Rohwer, P.W. Postma, B.N. Kholodenko, and H.V. Westerhoff. Implications of macromolecular crowding for signal transduction and metabolite channeling. *Proc Natl Acad Sci US*, 95:10547, 1998.

[84] T. Rowan. *Functional stability analysis of numerical algorithms.* PhD thesis, University of Texas at Austin, 1990.

[85] H.M. Sauro. `SCAMP`: a general-purpose simulator and metabolic control analysis program. *Comput Applic Biosci.*, 9:441–450, 1993.

[86] M. Sauro. `JARNAC - SCAMP II`. http://members.tripo.co.uk/sauro/biotech.htm.

[87] M.A. Savageau. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor Biol.*, 25:365–369, 1969.

[88] M.A. Savageau. Biochemical system analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol.*, 26(2):215–226, 1970.

[89] M.A. Savageau. A theory of alternative designs for biochemical control systems. *Biodem Biochim Acta*, 44(6):875–880, 1985.

[90] U. Schäfer. *Automatisierte Probenahme zur Messung intrazellulärer Metabolitdynamiken.* PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät der Rheinischen-Friedrich-Wilhelms-Universität Bonn, 1999.

[91] S.M. Schoberth, N-K. Bär, R. Krämer, and Kärger, J. Pulsed high-field gradient *in vivo* NMR spectroscopy to measure diffusional water permeability in *Corynebacterium glutanicum. Analytical Biochemistry*, 279:100–105, 2000.

[92] P. Schwille. Einzelmolekül-Analysemethoden für die evolutive Biotechnologie. Vortrag, DECHEMA Jahrestagung der Biotechnologen, Leipzig, Germany., März 2001.

[93] Šefčik, J. and Šefčiková, M. and Báleš, V. Bubble distribution and liquid circulation in airlift reactors. In Báleš, B., editor, *Modelling for improved bioreactor performance, 27.-28. September, 1993, Papiernička, Slovakia*, pages 86–94. Malú Centrum Publisher, 1993.

[94] L.A. Segel. *Biological Kinetics.* Cambridge University Press, 1991.

[95] E.E. Sel'kov. Self-oscillations in glycolysis. 1. a simple kinetic model. *Eur J Biochem.*, 4(1):79–86, 1968.

[96] G. Stephanopoulos, J. Nielsen, and A. Aristidou. *Metabolic Engieering: Principles and methodologies.* Academinc Press, Inc., 1998.

[97] H.J. Stoffers, E.L. Sonnhammer, G.J. Blommestijn, N.J. Raat, and H.V. Westernhoff. `METASIM`: object-oriented modelling of cell regulation. *Comput Appl Biosci.*, 8(5):443–449, 1992.

[98] L. Stryer. *Biochemistry.* W.H. Freeman Company, 4th edition edition, 1995.

[99] The Mathworks Inc. `MATLAB` and `SIMULINK`. http://www.mathworks.com.

[100] M. Tomita, K. Hashimoto, K. Takahashi, T.S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, and C.A. Hutchison. III. `E-cell`: a software environment for whole-cell simulation. *Bioinformatics*, 15(1), 1999.

[101] `MySQL` Relational Database System. www.mysql.com.

[102] A. Varma and B.O. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio-Technology*, 12(8):994–998, 1994.

[103] D. Visser, R. Heijden, K. Mauch, K.M. Reuss, and S. Heijnen. Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae. Metab Eng.*, 2(3):252–275, 2000.

[104] E.O. Voit, editor. *Canonical nonlinear modeling. S-system approach to understanding complexity.* Van Nostrand Reinhold, 1991.

[105] von Grünbert, H.H. and Bechinger, C. Die Attraktivität von Unordnung. *Spektrum der Wissenschaft*, pages 16–18, Juni 2000.

[106] B. Von Hohenbalken, B.L. Clarke, and J.E. Lewis. Least distance methods for the frame of homogenous equations systems. *J. Comput. Appl. Math.*, 19:231–241, 1987.

[107] Waterloo Maple. MAPLE 6. http://www.maplesoft.com.

[108] J. Werner. *Numerische Mathematik 1.* vieweg studium, 1992.

[109] W Wiechert. Diskrete Analyse von Stoffwechselnetzwerken. In F. Haubensack and J. Shnel, editors, *Bioinformatik - Computereinsatz in den Biowissenschaften, 9. Workshop und 2. Fachtagung Informatik in den Biowissenschaften, 5.9.-7.9.1994, Jena*, pages 100–111. Institut für Molekulare Biotechnologie, 1994.

[110] W. Wiechert and A de Graaf. In vivo stationary flux analysis by 13C labeling experiments . *Adv Biochem Eng Biotechnol.*, 54:109–154, 1996.

[111] W. Wiechert and A. de Graaf. Bidirectional reactions steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol Bioeng.*, 55(1):101, 1997.

[112] W. Wiechert, B. Joksch, R. Wittig, A. Hartbrich, T. Höner, and M. Möllney. Object oriented programming for the biosciences. *Comput Appl Biosci.*, 11(5):517–534, 1995.

[113] WIT. A WWW-based system to support the curation of function assignments made to genes and the development of metabolic models. http://wit.mcs.anl.gov/WIT2/.

[114] T.C. Zangirolami, C.L. Johansen, J. Nielsen, and S.B. Joergensen. Simulation of penicillin production in fed-batch cultivations using a morphologically structured model. *Biotechnol Bioeng.*, 56(6):593–604, 1997.

A. Table of Symbols and Abbreviations

A.1. Metabolites and nucleotides

Table A.1 summarizes the symbols used in this work and shows the page number where the symbol is defined or introduced. Table A.2 shows abbreviations of the metabolites, Table A.3 abbreviations of nucleotides and other compounds. The last column "Substance ID Internet" contains the official compound ID as used in public databases on the Internet.

Table A.1.: Symbol table. The page number in column 3 refers to the context of the definition. Therefore the definition of the specific symbol might sometimes be found on the following page.

Symbol	Meaning	Definition
C, $\mathbf{C}_i$	concentration of a compound (for a list of compounds see Table A.2)	see page 36
$\dot{C}$, $\dot{\mathbf{C}}_i$	derivative of a compound concentration with respect to time (for a list of compounds see Table A.2)	see page 41
$\mathbf{C}$	vector of intracellular compounds	see page 40
$\mathbf{C}_0$	vector of intracellular compounds at initial time t_0	see page 102
d_k	measure for the distance between calculated and measured values for calculation of the error functional F	see page 103
$D_{\mathbf{X}}$	total differential operator	see page 93
$\varepsilon_{\mathbf{i}}$	distance between calculated and measured value	see page 94
$\mathbf{E}$	vector of enzyme concentrations	see page 40

continued on next page

Table A.1.: (continued)

Symbol	Meaning	Definition
$\mathbf{E}_i$	elements of the vector of enzyme concentrations $\mathbf{E}$	see page 40
F	error functional	see page 103
IVP	initial value problem	see page 81
$\mathbf{K}_m$, $\mathbf{K}_{m,*}$	kinetic parameter, saturation constant	see page 40
$\mathbf{K}_I$, $\mathbf{K}_{I,*}$	kinetic parameter, inhibition constant	see page 46
$\mathbf{M}$	measure matrix	see page 94
MMT	metabolic modeling tool	see page 8
$\mathbf{N}$	stoichiometric matrix	see page 38
N_P	number of parameters of the vector $\mathbf{P}$	see page 107
N_t	number of intermediate time steps t_i where measurements exist	see page 94
N_v	number of reaction rates $\mathbf{v}_i$ summarized in the flux vector $\mathbf{v}$	see page 36
ODE	ordinary differential equation	see page 41
$\mathbf{P}$	vector of kinetic parameters	see page 40
PFK	phosphofructokinase	see page 119
PT-system	phosphotransferase system	see page 18
$\mathbf{P}_0$	vector of initial kinetic parameters	see page 109
PP-Pathway	pentose-phosphate pathway	see page 18
$\mathbf{S}$	vector of extracellular compounds	see page 40
SQL	standard query language	see page 60
TCA-Cycle	citric-acid-cycle	see page 18
$\mathbf{v}$	vector of fluxes (reaction rates)	see page 36
$\mathbf{v}_i$	element i of the flux vector $\mathbf{v}$	see page 36
$\mathbf{v}_{A\to B}$, $\mathbf{v}_{A\leftrightarrow B}$	flux vector (reaction rate) corresponding to the reaction from A to B. $\to$ ($\leftrightarrow$) indicates an irreversible (reversible) reaction	see page 36
$\mathbf{v}(\mathbf{C}, \mathbf{S}, \mathbf{E}, \mathbf{P})$	vector of reaction rates $\mathbf{v}_i$ as functions of $\mathbf{C}$, $\mathbf{S}$, $\mathbf{E}$, $\mathbf{P}$.	see page 40
$\mathbf{v}_{max}$, $\mathbf{v}_{max,*}$	kinetic parameter, maximal velocity of a kinetic rate $\mathbf{v}$	see page 40
$\mathbf{w}_j$	weighting factor in the error functional F	see page 103

continued on next page

Table A.1.: (continued)

Symbol	Meaning	Definition
$\mathbf{y}_i$	calculated value at time t_i	see page 94
$\tilde{\mathbf{y}}_i$	experimental data at time t_i	see page 94
$\tilde{\mathbf{Y}}$	matrix of all experimental data points	see page 94

Table A.2.: Table of abbreviations: Metabolites.

Abbreviation	Full Name	chemical Composition	Substance ID Internet
2PG	2-Phospho-D-glycerate, D-Glycerate 2-phosphate	$C_3H_7O_7P$	C 00631
3PG	3-Phospho-D-glycerate, D-Glycerate 3-phosphate	$C_3H_7O_7P$	C 00197
6PG	6-Phospho-D-gluconate	$C_6H_{13}O_{10}P$	C 00345
6PGL	D-Glucono-1,5-lactone 6-phosphate	$C_6H_{11}O_9P$	C01236
ACE	Acetate, Acetic acid, Ethanoic acid	$C_2H_4O_2$	C 00033
ACoA	Acetyl-CoA, Acetyl coenzyme A	$C_{23}H_{38}N_7O_{17}P_3S$	C 00024
ACP	Acetyl phosphate	$C_2H_5O_5P$	C 00227
ADH	Acetaldehyde, Ethanal	$C2H4O$	C 00084
CIT	Citrate, Citric acid, 2-Hydroxy-1,2,3-propanetricarboxylic acid, 2-Hydroxytricarballylic acid	$C_6H_8O_7$	C 00158
CoA	Coenzyme A, CoA-SH	$C_{21}H_{36}N_7O_{16}P_3S$	C 00010
DHAP	Glycerone phosphate, Dihydroxyacetone phosphate	$C_3H_7O_6P$	C 00111

continued on next page

Table A.2.: (continued)

Abbreviation	Full Name	chemical Composition	Substance ID Internet
DPG	3-Phospho-D-glyceroyl phosphate, 1,3-Bisphospho-D-glycerate, (R)-2-Hydroxy-3-(phosphonooxy)-1-monoanhydride	$C_3H_8O_{10}P_2$	C 00236
F6P	beta-D-Fructose 6-phosphate	$C_6H_{13}O_9P$	C 05345
FBP	beta-D-Fructose 1,6-bisphosphate	$C_6H_{14}O_{12}P_2$	C 05378
FOR	Formate, Methanoic acid, Formic acid	CH_2O_2	C 00058
FUM	Fumarate, Fumaric acid, Trans-butenedioic acid	$C_4H_4O_4$	C 00122
G3P	sn-Glycerol 3-phosphate, Glycerophosphoric acid, Phosphatidyl glycerol, sn-Gro-1-P, Glycerol-3-phosphate	$C_3H_9O_6P$	C 00093
GAP	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal, D-Glyceraldehyde 3-phosphate	$C_3H_7O_6P$	C00118
G6P	alpha-D-Glucose 6-phosphate	$C_6H_{13}O_9P$	C 00668
GLC	D-Glucose extracellular, Grape sugar, Dextrose	$C_6H_{12}O_6$	C 00031
GLO	Glycerone, Dihydroxyacetone, 1,3-Dihydroxyacetone, 1,3-Dihydroxy-2-propanone, 1,3-Dihydroxypropan-2-one	$C_3H_6O_3$	C 00184
GLY	Glycerol, Glycerin, 1,2,3-trihydroxypropane, 1,2,3-propanetriol	$C_3H_8O_3$	C00116

continued on next page

Table A.2.: (continued)

Abbreviation	Full Name	chemical Composition	Substance ID Internet
ISC	Isocitrate, Isocitric acid, 1-Hydroxytricarballylic acid, 1-Hydroxypropane-1,2,3-tricarboxylic acid	$C_6H_8O_7$	C 00311
LAC	(R)-Lactate, D-Lactate, D-Lactic acid, D-2-Hydroxypropanoic acid, D-2-Hydroxypropionic acid	$C_3H_6O_3$	C 00256
MAL	(S)-Malate, L-Malate, L-Apple acid, L-Malic acid, L-2-Hydroxybutanedioic acid	$C_4H_6O_5$	C 00149
MGL	Methylglyoxal, Pyruvaldehyde, Pyruvic aldehyde, 2-Ketopropionaldehyde, 2-Oxopropanal	$C_3H_4O_2$	C 00546
OAA	Oxaloacetate, Oxalacetic acid, Oxaloacetic acid, 2-oxobutanedioic acid, Oxosuccinic acid	$C_4H_4O_5$	C 00036
OXA	Oxalosuccinate, Oxalosuccinic acid	$C_6H_6O_7$	C 05379
OXO	2-Oxoglutarate, Oxoglutaric acid, 2-Ketoglutaric acid, alpha-Ketoglutaric acid	$C_5H_6O_5$	C 00026
PEP	Phosphoenolpyruvate, Phosphoenolpyruvic acid	$C_3H_5O_6P$	C 00074
PYR	Pyruvate, Pyruvic acid, 2-Oxopropanoate, 2-Oxopropanoic acid, Pyroracemic acid	$C_3H_4O_3$	C 00022
R5P	D-Ribose 5-phosphate, Ribose 5-phosphate	$C_5H_{11}O_8P$	C 00117

continued on next page

Table A.2.: (continued)

Abbreviation	Full Name	chemical Composition	Substance ID Internet
RL5P	D-Ribulose 5-phosphate	$C_5H_{11}O_8P$	C 00199
SCoA	Succinyl-CoA, Succinyl coenzyme A	$C_{25}H_{40}N_7O_{19}P_3S$	C00091
SDHP	S-Succinyldihydrolipoamide	$C_{12}H_{21}NO_4S_2$	C 01169
SDP	D-Sedoheptulose 7-phosphate, D-altro-Heptulose 7-phosphate	$C_7H_{15}O_{10}P$	C 05382
SUC	Succinate, Succinic acid, Butanedionic acid, Ethylenesuccinic acid	$C_4H_6O_4$	C 00042
X5P	D-Xylulose 5-phosphate	$C_5H_{11}O_8P$	C 00231

Table A.3.: Table of abbreviations: Nucleotides and others.

Abbreviation	Full Name	chemical Composition	Substance ID Internet
ADP	Adenosine 5'-diphosphate	$C_{10}H_{15}N_5O_{10}P_2$	C00008
AMP	Adenosine 5'-monophosphate, Adenylic acid, Adenylate, 5'-AMP, 5'-Adenylic acid, 5'-Adenosine monophosphate, Adenosine 5'-phosphate	$C_{10}H_{14}N_5O_7P$	C00020
ATP	Adenosine 5'-triphosphate	$C_{10}H_{16}N_5O_{13}P_3$	C00002
cAMP	3',5'-Cyclic AMP, Cyclic adenylic acid, Cyclic AMP, Adenosine 3',5'-phosphate	$C_{10}H_{12}N_5O_6P$	C00575
NAD	Nicotinamide adenine dinucleotide	$C_{21}H_{28}N_7O_{14}P_2$	C00003
NADH		$C_{21}H_{29}N_7O_{14}P_2$	C00004
NADP	Nicotinamide adenine dinucleotide phosphate, beta-Nicotinamide adenine dinucleotide phosphate	$C_{21}H_{28}N_7O_{17}P_3$	C00006
NADPH		$C_{21}H_{30}N_7O_{17}P_3$	C00005
P	Orthophosphate, Phosphate	H_3O_4P	C00009

A.2. Enzymes

Table A.4 summarizes enzyme abbreviations, names, and official E.C. numbers. Shown is also the main reaction that is catalyzed by each enzyme.

Table A.4.: Table of abbreviations: Enzymes.

EC Number	Systematic Name	Recommended Name	Reaction Catalysed
1.1.1.6	Glycerol:NAD+ 2-oxidoreductase	Glycerol dehydrogenase	GLY + NAD ↔ GLO + NADH
1.1.1.8	Sn-Glycerol-3-phosphate:NAD+ 2-oxidoreductase	Glycerol-3-phosphate dehydrogenase (NAD+)	G3P + NAD ↔ DHAP + NADH
1.1.1.28	(R)-Lactate:NAD+ oxidoreductase	D-Lactate dehydrogenase	PYR + NADH ↔ LAC + NAD
1.1.1.37	(S)-Malate:NAD+ oxidoreductase	Malate dehydrogenase	OAA + NADH ↔ MAL + NAD
1.1.1.38	(S)-Malate:NAD+ oxidoreductase (oxaloacetate-decarboxylating)	Malate dehydrogenase (oxaloacetate-decarboxylating)	MAL + NAD → PYR + NADH + CO_2
1.1.1.40	(S)-Malate:NADP+ oxidoreductase (oxaloacetate-decarboxylating)	Malate dehydrogenase (oxaloacetate-decarboxylating, NADP+)	MAL + NADP → PYR + NADPH + CO_2
1.1.1.42	Isocitrate:NADP+ oxidoreductase (decarboxylating)	Isocitrate dehydrogenase (NADP+)	ISC + NADP → OXO + NADPH + CO_2
1.1.1.44	6-Phospho-D-gluconate:NADP+ 2-oxidoreductase (decarboxylating)	Phosphogluconate dehydrogenase (decarboxylating)	6PG + NADP → RL5P + NADPH + CO_2

continued on next page

EC Number	Systematic Name	Recommended Name	Reaction Catalysed
1.1.1.49	D-Glucose-6-phosphate:NADP+ 1-oxidoreductase	Glucose-6-phosphate 1-dehydrogenase	G6P + NADP ↔ 6PGL + NADPH
1.2.1.3	Aldehyde:NAD+ oxidoreductase	Aldehyde dehydrogenase (NAD+)	ADH + NAD + H_2O ↔ ACE + NADH
1.2.1.10	Acetaldehyde:NAD+ oxidoreductase (CoA-acetylating)	Acetaldehyde dehydrogenase (acetylating)	ACoA + NADH ↔ CoA + ADH + NAD
1.2.1.12	D-Glyceraldehyde-3-phosphate: NAD+ oxidoreductase (phosphorylating)	Glyceraldehyde-3-phosphate dehydrogenase	GAP + NAD + P ↔ DPG + NADH
1.2.1.22	(S)-Lactaldehyde:NAD+ oxidoreductase	Lactaldehyde dehydrogenase	MGL + NAD + H_2O ↔ PYR + NADH
1.2.2.2	Pyruvate: ferricytochrome-b1 oxidoreductase	Pyruvate dehydrogenase (cytochrome)	PYR + H_2O → ACE + CO_2
1.2.4.2	2-Oxoglutarate: lipoamide 2-oxidoreductase (decarboxylating and acceptor-succinylating)	Oxoglutarate dehydrogenase (lipoamide)	OXO → SDHP + CO_2
1.3.99.1	Succinate: (acceptor) oxidoreductase	Succinate dehydrogenase	FUM + NADH ↔ SUC + NAD, FUM + NADPH ↔ SUC + NADP
2.2.1.1	Sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glycolaldehydetransferase	Transketolase	X5P + R5P ↔ GAP + SDP, X5P + GAP ↔ F6P

continued on next page

EC Number	Systematic Name	Recommended Name	Reaction Catalysed
2.3.1.54	Acetyl-CoA:formate C-acetyltransferase	Formate C-acetyltransferase	CoA + PYR + NAD → ACoA + FOR + CO_2
2.3.1.61	Succinyl-CoA:dihydrolipoamide S-succinyltransferase	Dihydrolipoamide S-succinyltransferase	SCoA ↔ SDHP + CoA
2.7.1.11	ATP:D-fructose-6-phosphate 1-phosphotransferase	6-Phosphofructokinase	F6P + ATP → FBP + ADP
2.7.1.30	ATP:glycerol 3-phosphotransferase	Glycerol kinase	GLY + ATP ↔ G3P + ADP
2.7.1.40	ATP:pyruvate 2-O-phosphotransferase	Pyruvate kinase	PEP + ADP ↔ PYR + ATP
2.7.2.3	ATP:3-phospho-D-glycerate 1-phosphotransferase	Phosphoglycerate kinase	DPG + ADP ↔ 3PG + ATP
2.7.9.2	ATP:pyruvate,water phosphotransferase	Pyruvate,water dikinase	PYR + ATP + H_2O → PEP + AMP + P
3.1.1.31	6-Phospho-D-glucono-1,5-lactone lactonohydrolase	6-Phosphogluconolactonase	6PGL + H_2O ↔ 6PG
3.1.3.1	Orthophosphoric-monoester phosphohydrolase (alkaline optimum)	Alkaline phosphatase	GLO + P ↔ DHAP + H_2O
3.1.3.11	D-Fructose-1,6-bisphosphate 1-phosphohydrolase	Fructose-bisphosphatase	FBP → F6P + P

continued on next page

EC Number	Systematic Name	Recommended Name	Reaction Catalysed
4.1.1.31	Orthophosphate: oxaloacetate carboxy-lyase (phosphorylating)	Phosphoenolpyruvate carboxylase	PEP + H_2O + CO_2 → OAA + P
4.1.1.49	ATP:oxaloacetate carboxy-lyase (transphosphorylating)	Phosphoenolpyruvate carboxykinase (ATP)	OAA + ATP → ADP + CO_2
4.1.2.13	D-Fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate-lyase	Fructose-bisphosphate aldolase	FBP ↔ DHAP + GAP
4.1.3.7	Citrate oxaloacetate-lyase ((pro-3S)-CH2COO—¿ acetyl-CoA)	Citrate (si)-synthase	OAA + ACoA + H_2O ↔ CIT + CoA
4.2.1.2	(S)-Malate hydro-lyase	Fumarate hydratase	MAL ↔ FUM + H_2O
4.2.1.3	Citrate (isocitrate)hydro-lyase	Aconitate hydratase	CIT ↔ ISC + H_2O
4.2.1.11	2-Phospho-D-glycerate hydro-lyase	Enolase	2PG ↔ PEP +H_2O
4.2.99.11	Glycerone-phosphate phospho-lyase	Methylglyoxal synthase	DHAP ↔ MGL + P
5.1.3.1	D-Ribulose-5-phosphate 3-epimerase	Ribulose-phosphate 3-epimerase	RL5P ↔ X5P

continued on next page

EC Number	Systematic Name	Recommended Name	Reaction Catalysed
5.3.1.1	D-Glyceraldehyde-3-phosphate ketol-isomerase	Triose-phosphate isomerase	DHAP $\leftrightarrow$ GAP
5.3.1.6	D-Ribose-5-phosphate ketol-isomerase	Ribose-5-phosphate isomerase	RL5P $\leftrightarrow$ R5P
5.3.1.9	D-Glucose-6-phosphate ketol-isomerase	Glucose-6-phosphate isomerase	G6P $\leftrightarrow$ F6P
5.4.2.1	D-Phosphoglycerate 2,3-phosphomutase	Phosphoglycerate mutase	3PG $\leftrightarrow$ 2PG
6.2.1.5	Succinate:CoA ligase (ADP-forming)	Succinate-CoA ligase (ADP-forming)	SUC + ATP $\leftrightarrow$ SCoA + ADP + P

B. Pathway Maps

Figure B.1 shows the main metabolic pathways of *Escherichia coli*. Figure B.2 shows the reactions of the nucleotides and the legend for all pathway figures in this work. Figure B.3 shows the same pathway as Figure B.1 but includes the official enzyme number for each reaction step.

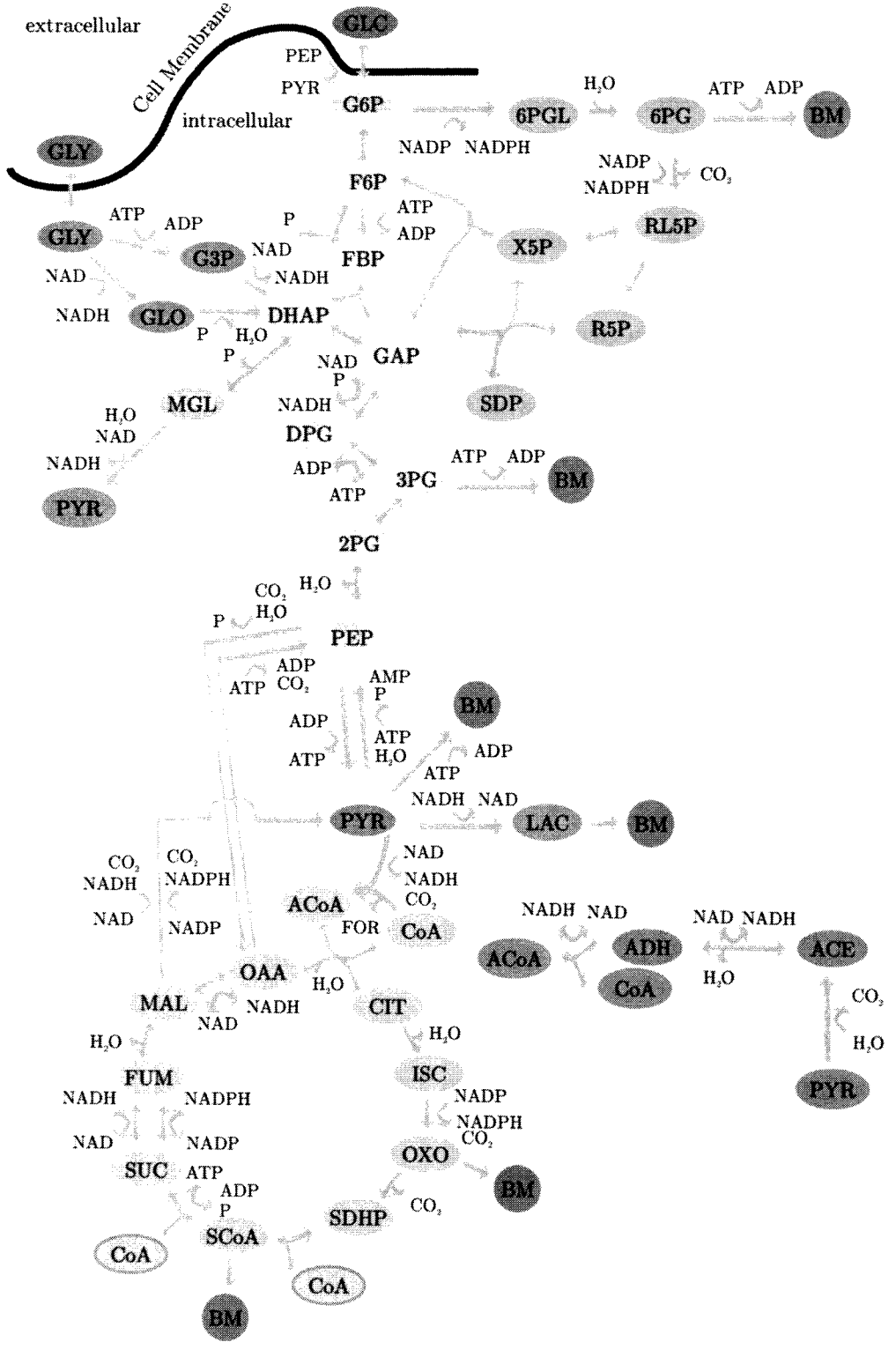


Figure B.1.: The main pathways of *Escherichia coli*. See Figure B.2 for the legend.

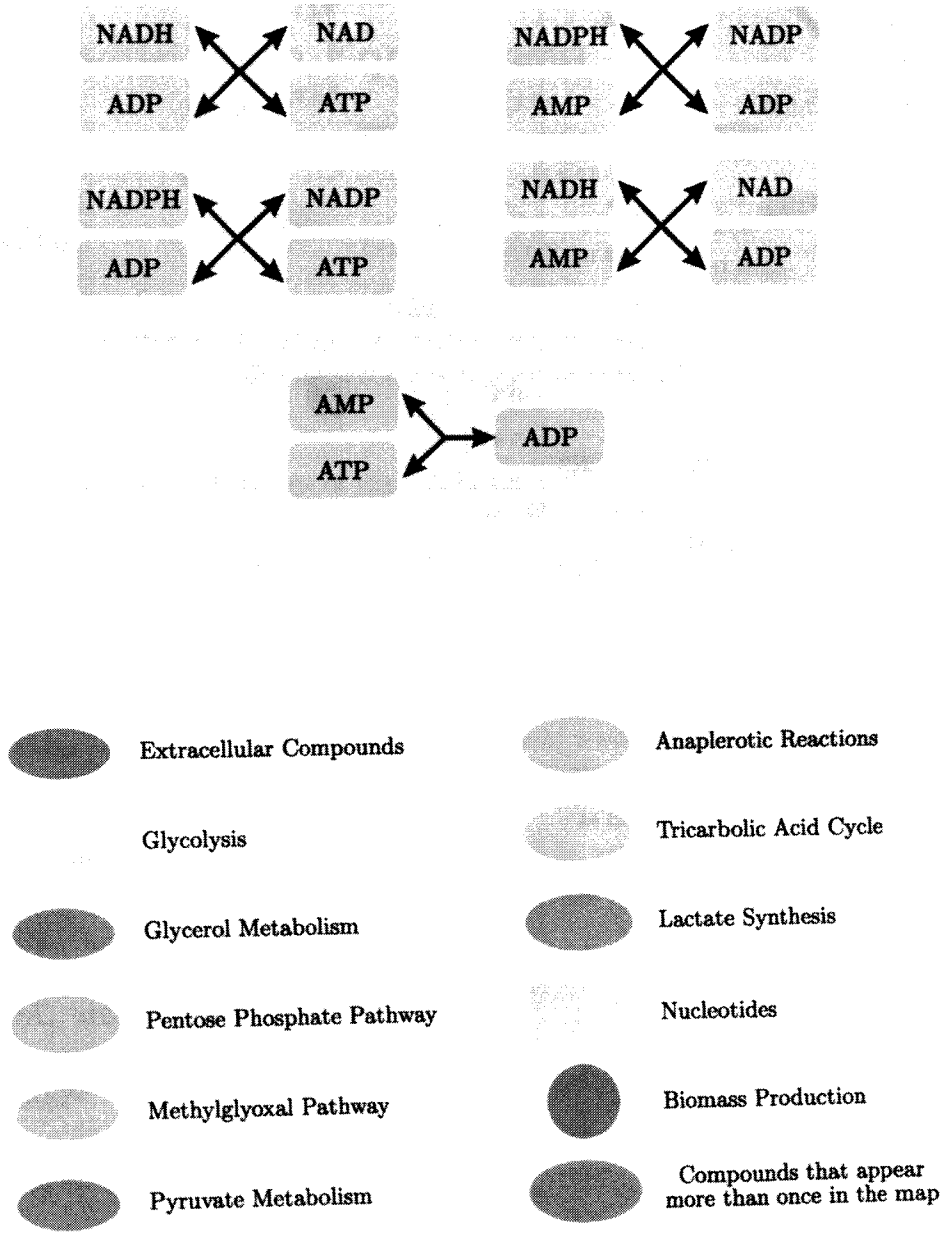


Figure B.2.: The nucleotide interconversion reactions of *Escherichia coli* and the legend for all pathway maps in this work.

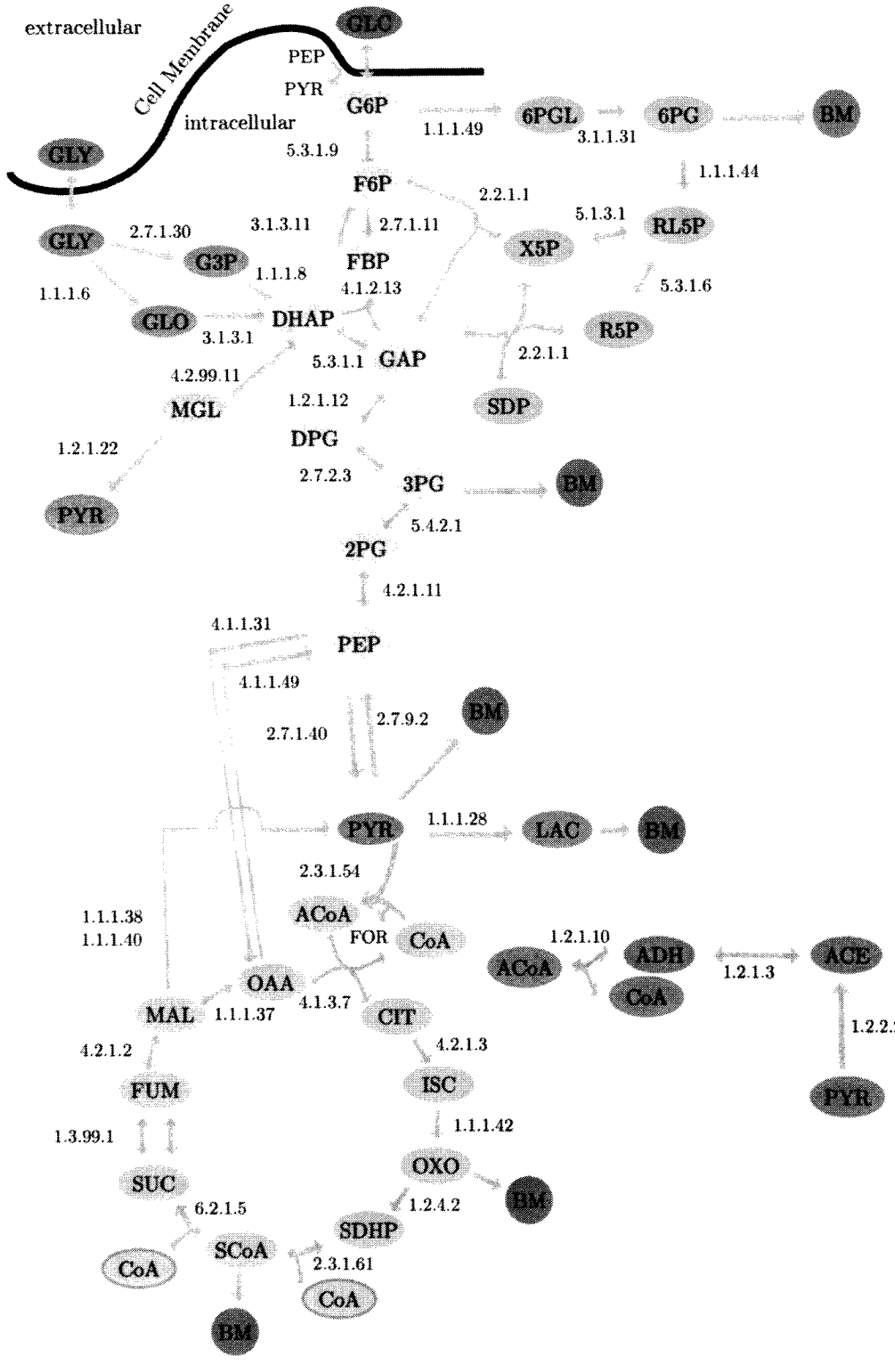


Figure B.3.: The main pathways of *Escherichia coli* with catalyzing enzymes (official enzyme number displayed). See Figure B.2 for the legend.

C. Kinetic Rate Equations

Notation

This appendix shows the kinetic rate equations used in the metabolic pathway models in this work. The following notation is used in this appendix:

- A and B are substrates of the reactions.
- P and Q are products of the reactions.
- I is an inhibitor.
- v is the reaction rate. v_{max} is a maximum reaction rate in irreversible reactions. v_f and v_r are the forward respectively reverse reaction rates in reversible reactions.
- K_{eq} is the equilibrium constant in reversible reactions.
- $K_{m_\cdot}$ are half saturation constants.
- $K_{i_\cdot}$ are inhibition constants.

Irreversible Michaelis-Menten Equation:

$$v = v_{max} \cdot \frac{A}{K_m A + A} \tag{C.1}$$

Irreversible Michaelis-Menten Equation with two Substrates

$$v = \frac{v \cdot A \cdot B}{K_{m,A} \cdot K_{m,B} + K_{m,B} \cdot A + K_{m,A} \cdot B + A \cdot B} \tag{C.2}$$

Reversible Michaelis-Menten Equation

$$v = \frac{v_f \cdot \frac{A}{K_{mA}} - v_r \cdot \frac{P}{K_{mP}}}{1 + \frac{A}{K_{mA}} + \frac{P}{K_{mP}}} \tag{C.3}$$

Hill-Kinetics

$$v = v_{max} \cdot \frac{A^h}{K_m^h + A^h} \tag{C.4}$$

h: Hill coefficient

Ordered Uni-Bi Kinetics

$$v = \frac{v_f \cdot (A - \frac{P \cdot Q}{Keq})}{K_{mA} + A \cdot (1 + \frac{P}{K_{iP}}) + \frac{v_f}{v_r \cdot K_{eq}} \cdot (K_{mQ} \cdot P + K_{mP} \cdot Q + P \cdot Q)} \tag{C.5}$$

Ordered Bi-Uni Kinetics

$$v = \frac{v_f \cdot (A \cdot B - \frac{Q}{K_{eq}})}{A \cdot B + K_{mA} \cdot B + K_{mB} \cdot A + \frac{v_f}{v_r \cdot K_{eq}} \cdot \left(K_{mP} + Q \cdot (1 + \frac{A}{K_{iA}})\right)} \tag{C.6}$$

Ordered Inhibited Bi-Uni Kinetics

$$v = \frac{v_f \cdot (A \cdot B - \frac{Q}{K_{eq}})}{A \cdot B + K_{mA} \cdot B + K_{mB} \cdot A + \frac{v_f}{v_r \cdot K_{eq}} \left(K_{mP} + Q \cdot (1 + \frac{A}{K_{iA}})\right) + (\frac{I}{K_i})^n} \tag{C.7}$$

n: strength of inhibition, usually dependent on number of bindings sites of the enzyme.

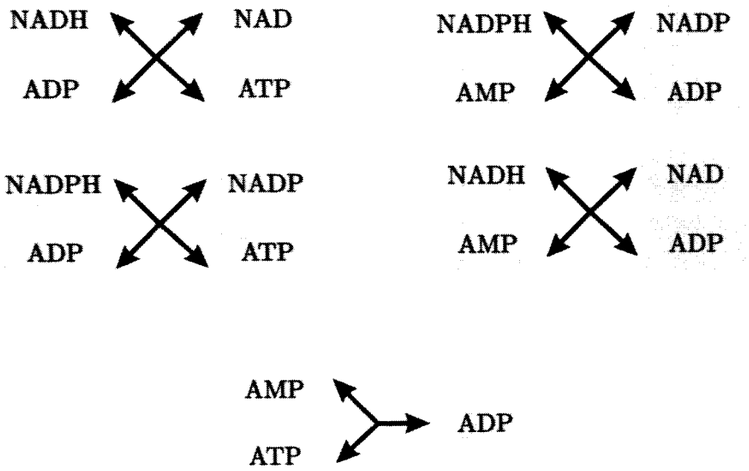
Ordered Bi-Bi Kinetics:

$$v, = \frac{v_f \cdot (\mathrm{A} \cdot \mathrm{B} - \frac{\mathrm{P} \cdot \mathrm{Q}}{K_{eq}})}{X} \tag{C.8}$$

where the denominator is

$$X = \mathrm{A}\cdot\mathrm{B}\cdot(1+\frac{\mathrm{P}}{K_{i,\mathrm{P}}}) + K_{m,\mathrm{B}}\cdot(\mathrm{A}+K_{i,\mathrm{A}}) + K_{m,\mathrm{A}}\cdot\mathrm{B} + \ldots$$

$$\ldots + \frac{v_f\cdot\left(K_{m,\mathrm{P}}\cdot\mathrm{P}\,(1+\frac{\mathrm{A}}{Ki,A})\right) + \mathrm{Q}\cdot\left(K_{m,\mathrm{P}}\cdot(1+\frac{K_{m,\mathrm{A}}\cdot\mathrm{B}}{K_{iA}\cdot K_{mB}}) + \mathrm{P}\cdot(1+\frac{\mathrm{B}}{K_{i,\mathrm{B}}})\right)}{v_r\cdot K_{eq}} \tag{C.9}$$



Extracellular Compounds

Glycolysis

Glycerol Metabolism

Pentose Phosphate Pathway

Methylglyoxal Pathway

Pyruvate Metabolism

Anaplerotic Reactions

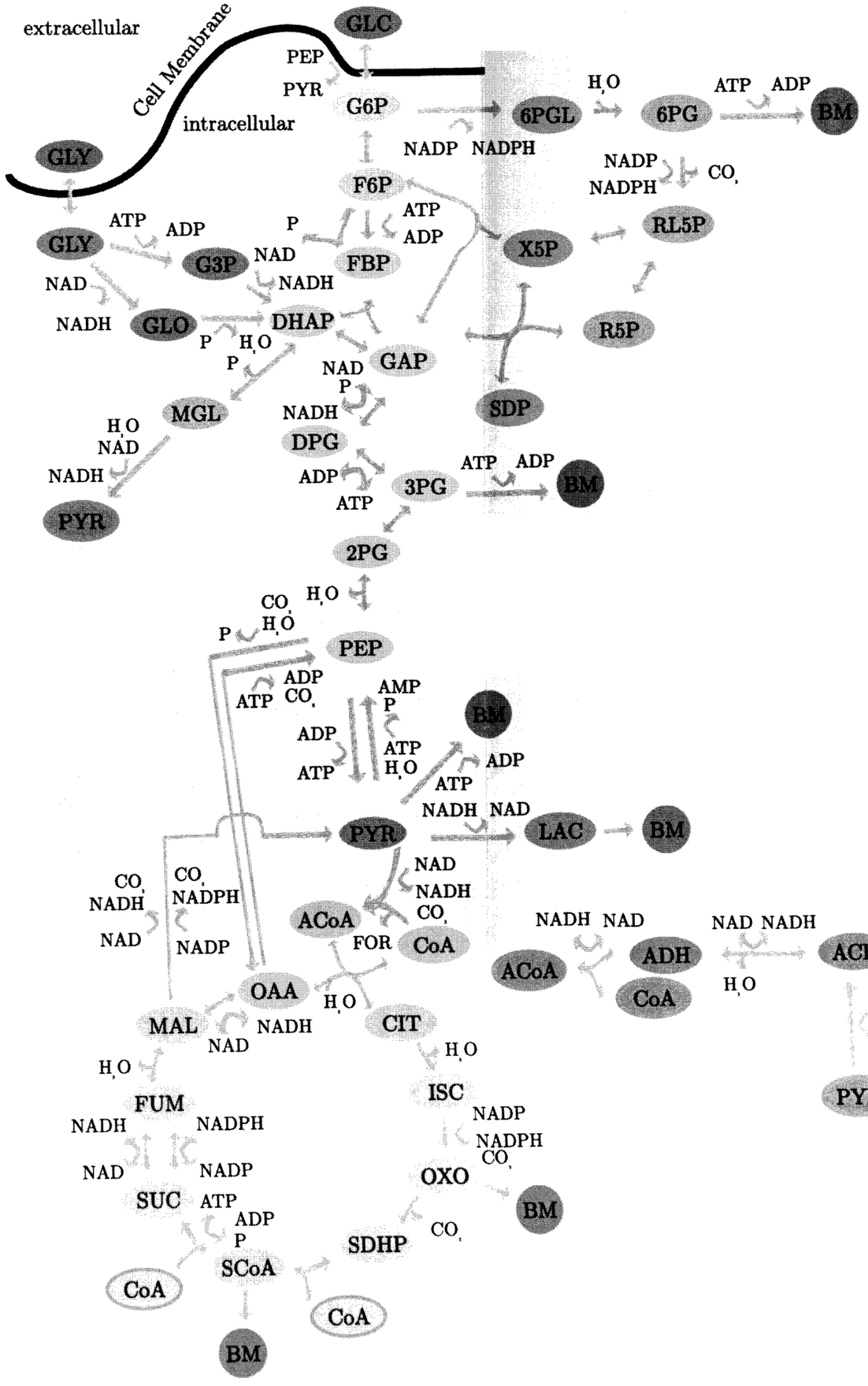
Tricarbolic Acid Cycle

Lactate Synthesis

Nucleotides

Biomass Production

Compounds that appear more than once in the map



Forschungszentrum Jülich

Jül-3912
Oktober 2001
ISSN 0944-2952