MATHEMATICAL MODELLING OF ENZYMATIC REACTIONS, SIMULATION AND PARAMETER ESTIMATION

SÜREYYA ÖZÖĞÜR

January 2005

MATHEMATICAL MODELLING OF ENZYMATIC REACTIONS, SIMULATION AND PARAMETER ESTIMATION

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF APPLIED MATHEMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

SÜREYYA ÖZÖĞÜR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF SCIENTIFIC COMPUTING

JANUARY 2005

Approval of the Graduate School of Applied Mathematics

Prof. Dr. Aydın AYTUNA
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Bülent Karasözen
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Bülent Karasözen
Supervisor

Examining Committee Members

Prof. Dr. Bülent Karasözen _____

Prof. Dr. Gerhard Wilhelm Weber _____

Dr. Hakan Öktem _____

Prof. Dr. Münevver Tezer _____

Prof. Dr. Feza Korkusuz _____

Abstract

MATHEMATICAL MODELLING OF ENZYMATIC REACTIONS, SIMULATION AND PARAMETER ESTIMATION

Süreyya Özöğür
M.Sc., Department of Scientific Computing
Supervisor: Prof. Dr. Bülent Karasözen

January 2005, 75 pages

A deep and analytical understanding of the human metabolism grabbed attention of scientists from biology, medicine and pharmacy. Mathematical models of metabolic pathways offer several advances for this deep and analytical understanding due to their incompensable potential in predicting metabolic processes and anticipating appropriate interventions when required. This thesis concerns mathematical modelling analysis and simulation of metabolic pathways. These pathways include intracellular and extracellular compounds such as enzymes, metabolites, nucleotides and cofactors. Experimental data and available knowledge on metabolic pathways are used in constituting a mathematical model. The models are either in the form of nonlinear ordinary differential equations (ODE's) or differential algebraic equations (DAE's). These equations are composed of kinetic parameters such as kinetic rate constants, initial rates and concentrations of metabolites. The nonlinear nature of enzymatic reactions and large number of parameters cause trouble in efficient simulation of those

reactions. Metabolic engineering tries to simplify these equations by reducing the number of parameters. In this work, an enzymatic system which includes Creatine Kinase, Hexokinase and Glucose 6-Phosphate Dehydrogenase (CK-HK-G6PDH) is modelled in the form of DAE's, solved numerically and the system parameters are estimated. The numerical results are compared with the results from an existing work in literature. We demonstrate that our solution method based on direct solution of the CK-HK-G6PDH system significantly differs from simplified solutions. We also show that a genetic algorithm (GA) for parameter estimation provides much more clear results to the experimental values of the metabolite, especially, with NADPH.

Keywords: metabolic engineering, kinetic modelling, biochemical reactions, enzymatic reactions, differential algebraic equations, parameter estimation, optimization, genetic algorithm.

Öz

ENZİMATİK REAKSİYONLARIN MATEMATİKSEL MODELLENMESİ, SİMULASYONU VE PARAMETRE KESTİRİMİ

Süreyya Özöğür Yüksek Lisans, Bilimsel Hesaplama.
Tez Yöneticisi: Prof. Dr. Bülent Karasözen

Ocak 2005, 75 sayfa

İnsan metabolizmasının detaylı ve analitik olarak incelenmesi, biyoloji, tıp ve eczacılık alanlarındaki bilim adamlarının dikkatlerini üzerine toplamıştır. Metabolik patikaların matematiksel modellemesi, metabolik süreçlerin gelecekteki davranışlarınıın tahmini ve gerektiğinde en uygun müdahelelerin öngörülebilmesi için kullanılabilirliği nedeniyle bu detaylı ve analitik incelemeye yeri doldurulamaz katkılar sağlamaktadır. Bu çalışma metabolik patikaların matematiksel modeli, analizi ve simülasyonu üzerinedir. Metabolik patikalar, hücre içi ve hücre dışı enzimler, metabolitler, nükleotidler ve kofaktörler gibi bileşikleri içermektedir. Deneysel veriler ve metabolik patikalar hakkındaki bilgiler matematiksel modelin oluşturulmasında kullanılmaktadır. Model denklemleri, lineer olmayan adi diferensiyel denklemler ya da cebirsel diferensiyel denklemlerden oluşmaktadır. Denklemler, kinetik hız sabitleri, reaksiyonların başlangıç hızları ve metabolitlerin başlangıç konsantrasyonları gibi kinetik parametreleri içermektedir. Enzimatik reaksiyonların doğrusal olmayan doğası ve fazla sayıda parametre içermesi doğru simule edilebilmelerinde sorun yaratmaktadır. Meta-

bolik mühendislik, parametre sayısını azaltarak model denklemlerini basitleştirmeye çalışmaktadır. Bu çalışmada üç enzimli kreatin kinaz, hekzokinaz ve glukoz altı fosfat dehidrogenaz sistemi cebirsel diferensiyel denklemler yardımıyla modellenmiş, sayısal yöntemlerle çözülmüş ve parametre tahmini yapılmıştır. Sayısal sonuçların literatürde yer alan çalışmalarla karşılaştırılması yapılmıştır. Bu karşılaştırmada CK-HK-G6PDH sisteminin doğrudan çözülebilmesine dayanan çözüm metodumuzun basitleştirilmiş çözümlerden oldukça farklı sonuçlar verdiği gösterilmiştir. Bunların yanısıra genetik algoritma (GA) ile parametre kestiriminin deneysel verilere özellikle NADPH metabolitinede çok daha yakın sonuçlar sağladığı gösterilmiştir.

Anahtar Kelimeler: metabolik mühendislik, kinetik modelleme biyokimyasal modelleme, enzimatik reaksiyonlar, cebirsel diferensiyel denklemler, parametre tahmini, optimizasyon, genetik algoritmalar.

To my family

Acknowledgments

I would like to thank all those people who have helped in the preparation of this study. I am grateful to Prof.Dr. Bülent KARASÖZEN, Prof.Dr. Gerhard Wilhelm WEBER and Gülçin Sağdıçoğlu Çelep for patiently guiding, motivating, and encouraging me throughout this study.

Special thanks to Dr. Necmettin Yıldırım for all his efforts in every phase of my study. I would like to thank him for his help and patience when writing an answer to all of my questions from North Carolina.

I would like to thank Prof.Dr. Bülent KARASÖZEN who has always encouraged me for my study. I would like to thank him for all his efforts, energy and support. I am very glad that I have met Prof.Dr. Gerhard-Wilhelm Weber who encourages me throughout this study.

I would like to thank Dr. Hakan Öktem and Gülçin Sağdıçoğlu Çelep for their valuable suggestions and helps whenever I need.

I want to thank my parents without whose encouragement this thesis would not be possible.

I am also grateful to Akın Akyüz, Yeliz Yolcu, Turgut Hanoymak, Mesut Taştan and Öznur Yaşar for being with me all the way. I am very glad that I have met Akın Akyüz who supports me in every respects and thanks to him for his endless patience throughout this study.

Finally, I am indebted to Oktay Sürücü, Ayşegül İşcanoğlu, Derya Altıntan, Osman Özgür, Selime Gürol, Didem Akçay, Emre Dağlı and Başak Akteke for their support.

Table of Contents

Abstract iii

Öz v

Acknowledgments viii

Table of Contents ix

List of Figures xii

List of Tables xiv

CHAPTER

1 Introduction 1

1.1 Brief History of Metabolic Engineering 3

1.2 Modelling of Biochemical Pathways 4

1.2.1 Aims and Scope of Metabolic Models 5

1.2.2 Metabolic Flux Analysis 7

1.2.3 Metabolic Control Analysis 10

1.2.4 Non-Stationary Mechanistic Models 11

2 Modelling Approaches 12
2.1 A Model of Enzyme Kinetics 12
2.1.1 Steady State Kinetics 12
2.1.2 Enzyme Catalyzed Reactions 15
2.1.3 Kinetics of Multi-substrate Enzymes 17
2.1.4 The Stoichiometric Matrix 25
2.2 Deriving Conservation Relations 27
2.2.1 Deriving Conservation Relations by Inspection 27
2.2.2 Deriving Conservation Relations by Gaussian Elimination 30
2.2.3 Deriving Conservation Relations by Partioning the Stoichiometric Matrix into Submatrices 32
2.3 Creatine Kinase, Hexokinase and Glucose-6Phosphate Dehydrogenase System 35
2.3.1 Creatine Kinase System 36
2.3.2 Hexokinase System 37
2.3.3 Glucose 6-Phosphate Dehydrogenase System 39
2.3.4 Mathematical Modeling of CK-HK-G6PDH System . . . 40

3 Numerical Solution Methods For Solving Differential Algebraic Equations 48
3.1 Introduction 48
3.2 Differential Index 51
3.3 Numerical Solution Methods to Solve DAE's 52
3.3.1 Backward Euler 53
3.3.2 Runge-Kutta Method 53

3.3.3 Backward Differentiation Formula 54

4 Parameter Estimation 56

4.1 Genetic Algorithm 57

4.1.1 Coding 57

4.1.2 Reproduction 58

4.1.3 Crossover 58

4.1.4 Mutation 60

4.2 GA Adapted to CK-HK-G6PDH Model 61

5 Results and Discussion 63

5.1 Results of Simulation 63

5.2 Conclusion 68

References 71

List of Figures

1.1 Example of fluxes for a simple reaction network where stoichiometric relations of fluxes are $u = x + y$, $v = x - z$, $w = y + z$ [31]. 8

1.2 Typical situations in which stoichiometric MFA fails [32]. 9

2.1 Schematic explanation of different types of reaction mechanisms (taken from [38]). 20

2.2 Schematic explanation of different types of reaction mechanisms (taken from [38]). 21

2.3 Schematic explanation of different types of reaction mechanisms (taken from [38]). 22

2.4 Explicit reaction scheme for ordered Bi Bi mechanism [38]. 23

2.5 *A*. skeleton model of glycolysis, *B*. This simplified representation includes only metabolites (taken from [13]). 27

2.6 Creatine Kinase, Hexokinase and Glucose-6Phosphate Dehydrogenase System [38]. 35

4.1 Illustration of a chromosome where each box represents one gene of chromosome. 58

4.2 A binary-valued, randomly generated pool. 59

4.3 A crossover 1. 59

4.4 A crossover 2. 60

5.1 Simulation results of ATP, ADP, NADP, 6PGL. 64

5.2 Simulation results of ATP, ADP, NADP, 6PGL [38] (zaman: time, dakika: minute). 65

5.3 Simulation results of DGlu6P, Cr. 65

5.4 Simulation results of DGlu6P, Cr [38] (zaman: time, dakika: minute). 66

5.5 Simulation results of CrP, DGlu. 66

5.6 Simulation results of CrP, DGlu [38] (zaman: time, dakika: minute). 67

5.7 Experimental values of NADPH and simulation results of NADPH by *ode15s*. 67

5.8 Experimental values of NADPH and simulation results of NADPH [38] (zaman: time, dakika: minute). 68

5.9 NADPH results after parameter estimation and experimental values of NADPH. 69

List of Tables

2.1 Parameters of CK system (Morrison and James, 1965). 46

2.2 Parameters of HK system (Viola et al., 1982). 46

2.3 Parameters of G6PDH system (Gordon et al., 1995). 47

Chapter 1

Introduction

Metabolic engineering is a challenging research field on nowadays incorporating of modern applied mathematics into bioetechnology, engineering science and pharmacy. Moreover, in medical studies, scientists work on human metabolism to improve the capabilities of some metabolites or enzymes in a metabolic pathways. In industrial applications, metabolic engineering methods are also widely used to develop certain methods for improving functionality of some molecules in a cell. To achieve these goals, a mathematical model of such metabolic systems must be constructed and simulated. Most of the dynamical systems can be approximated by various types of differential algebraic and integral equations involving finite number of variables and parameters. Thus, the future behaviour of the system can be predicted if model parameters and initial states of the variables are available. In particular, ordinary differential equations (ODE's) and differential algebraic equations (DAE's) are popular in modelling of the metabolic pathways. Because of difficulies in solving DAE's, in studies given in literature, the systems are generally reduced to ODE's. In this thesis, three enzymatic system, Creatin Kinase, Hexokinase and Glucose 6 Phosphate Dehydrogenase is modeled by DAE's with 3 differential equations and 6 constraints. These enzymatic systems play important roles in glycolysis, energy metabolism of tissues such as skeletal and cardiac muscle, and neural tissues. Since G6PDH is a rate limiting enzyme in pentose phosphate pathway, predicting or improving the behaviour of the enzymes or metabolites in the pathway is a crucial task

in bioinformatics. Modelling by differential algebraic equations considered in this work is also an important issue because of the fact that it is a different, a relatively new type of modelling. Solving the system of differential algebraic equations is generally avoided in most of the published works [26, 27, 38] because of its difficulty. Thus, our work introduces a new approach to the solution of our model problem. Furthermore, we assume more general cases in model and give a more accurate and general solution than the existing one in the literature. We assume that the pathway has three different rate equations rather than same rate equations assumed in steady state. Instead of approximating nonlinear rate equations as a polynomial functions with Gröbner basis theory, like in [38], they are taken completely as nonlinear and solved numerically. The experimental data are based on Bergmeyer (1982) and it is carried out in Laboratory of Biotechnology Applied and Research Center in Atatürk University in Erzurum [38]. The experimental conditions and the concentration values are given explicitly in [38]. For the estimation problem, a different method is used for our case. Since the experimental data and simulation results are available, one must minimize the difference of these two quantities to find the parameters that fit to the model best. This can be done by several methods, for example, minimizing an approximated objective function which is an easy and common way to solve this problem. Instead of this, a genetic algorithm is used for the optimization part of the problem which is a popular method in most of the estimation problems of metabolic pathways. By this approach, we achieved more accurate and closer results to the empiricial data.

In Chapter 1, methods of metabolic engineering and modelling of metabolic pathways are presented. Types of enzymatic reactions, fundamental rules and formulas in biochemistry are discussed in Chapter 2. In modelling of such systems, we must have some conservation relations in a biological manner. We explained this theory of deriving conservation relations in this chapter and at the end of the Chapter 2, CK-HK-G6PDH is introduced. For solving our DAE's we have studied DAE theory and numerical methods for them in Chapter 3. For this numerical solution, Backward Differentiation Formula (BDF) is selected

and it is implemented in the software MATLAB 6.1 by using its DAE solver **ode15s**. Especially, for complex nonlinear dynamical systems like metabolic pathways all parameters of a first principles model might not to be a priori available or the available information might no to be sufficiently reliable. In such cases, estimating the most likely parameters from the empirical observations can be substituted into an optimization problem by utilizing estimation theory. One of the most widely used method is Genetic Algorithms (GAs) for parameter estimation problems. It is introduced in Chapter 4. Finally, the simulation results, parameter estimation results and comparisons with experimental data are given in Chapter 5.

1.1 Brief History of Metabolic Engineering

Metabolic capabilities of different microorganisms are widely investigated with a motivation of producing desired compounds in pharmaceutical and chemical industry [23]. From these investigations we get a broad survey on cell structure, substance and enzymes, which gives us an insight into these microorganisms. Knowledge on the behaviour of those substances and enzymes in a cell provides some basic information, which is the foundation for modifying a pathway to improve it according to our benefit. This process is nowadays often referred to as *metabolic engineering*, a term introduced by Bailey in 1991. This dynamical interface between biology, medicine, pharmacy and mathematics aims at a directed improvement of product formation or cellular properties through the modification of specific biochemical reaction(s) or using DNA technology [14].

Metabolic engineering focusses on developping targeted methods to improve the metabolic capabilities of microorganisms. To achieve this goal, genetic manipulations of cell metabolism are required [31]. Metabolic engineering does not only deal with a modification of genes or pathways in an organism but also offers a deeper understanding of metabolic pathways in a systemic and analytical way [30]. In fact, metabolic engineering is an interdisciplinary area where

biochemists, molecular biologists, chemical engineers and mathematicians corporate to handle and solve the problem in both a way based on real nature and by a language and methodology of modern mathematics. It mainly studies metabolic pathways and gene networks in organisms to optimize some desired metabolites, especially, in industrial applications [23].

In metabolic engineering, *a metabolic design problem* can be formulated by constructing a mathematical model of a metabolic network. This model can be used for anticipating modifications or simplifications in the genotype of microorganisms.

The use of these mathematical models of metabolic networks plays a significant role in medical and life science applications also. There may be difficulties with the complex nature of cellular metabolism and regulation. Therefore, it is often necessary to analyze metabolism as a whole system where the most important biological parts are included and represented with their interrelations [23].

1.2 Modelling of Biochemical Pathways

Mathematical modelling is one of the most important clues for metabolic engineering. Here, metabolic models, several computational and simulation tools have been developped with the aim of system analysis, prediction, design and optimization. However, the model depends on the assumptions, simplifications and data used for constructing it. The goal of metabolic engineering is to find a solution in an easy, efficient and profitable way while improving the capabilities of industrially relevant microorganisms. Modelling tools and simulation software are already an important task.

The main issue in our modelling part is to find the best way of representing the biological information and experimental data in the sense of mathematics. There remains an open question about which information and data are applicable, since there are great and strong differences between biology and

computational mathematics. Due to the solvability problem, not every detail about a functional knowledge on cellular systems or all physical conditions of the experimental environment can be included in a mathematical model. In particular, biological systems can not be easily decomposed into their unit components. However, a developing number of studies and growing data bases give us an increasing functional knowledge about cellular systems.

While metabolic engineering has initially aimed at specific biochemical reactions which, however, seemed to limit the microbial production process, it soon became realized that a more global approach is required, such as integrated pathways or, going one step further, networks of pathways. Since any change in a pathway affects other pathways directly or indirectly, how one should use available kinetic information to manipulate any pathway without affecting other pathways is an essential question [30].

For all the insights we obtain, e.g., on enzyme-catalyzed reactions, individual pathways and integrated metabolic networks, let us construct mathematical methods and use computational methods. Only a few decades ago, mathematical analysis of metabolic systems would have been very difficult or even impossible because of unsolvable nonlinear equations. Today, the increasing computer power available allows simulating systems. However, the existing of a good mathematical model is a must for such a simulation to represent and analyze complex systems like metabolic pathways. This points out the importance of choosing a good mathematical representation and an efficient method of simulation. For closer information we refer to [30].

1.2.1 Aims and Scope of Metabolic Models

Both the aim of modelling of a given complex metabolic system and which modelling will be used should be specified. According to the increasing demand for the precision and quality of the model, the following requirements are listed in [31]:

1. *Structural understanding*: mathematical models are the most exact way of representing the metabolic system, while having unique and objective application. Therefore, they should not allow any ambigious statements. As a conclusion, modelling is a methodology used for giving a structure to a metabolic system or defining our system in a more precise statement. Even if the model is not used for simulation, it provides having a knowledge about important parts of the system. Nowadays, this methodology is used for designing complex software systems [31].

2. *Simulation for exploring the system*: In model applications, simulation is used for exploring the mechanism or behaviour of the system. A simulation procedure based on rather crude models is not strong enough to understand the system behaviour explicitly. Thus, such studies cannot directly produce right ideas. However, they provide most probable explanations for the behaviour of the metabolic system.

3. *Measured data analysis and evaluation*: Reproduction of measured data by a constructed mathematical model is a well established tool in scientific disciplines. For instance, growth nutrient uptake and product formation or concentration of metabolites in the biochemical reactions, which are characterized by kinetic models, are a common procedure in bioprocess development. It should be pointed out that it is only a reproduction of measured data and does not involve in any prediction. So, without further experimental data we cannot say anything about the predictive power of the model.

4. *Metabolic system analysis*: After constructing a model, mathematical methods can give sensible results for understanding the system's structure and its qualitative behaviour. The most popular methods are *parameter estimation* and *sensitivity analysis*. In fact, they are used to understand, which parameters are sensitive to small changes, for understanding the dynamical behaviour of the system and investigation of its oscillating or chaotic behaviour. Nevertheless, to obtain meaningful results, more exact

models are required.

5. *Predicting future results*: The outcome of future experiments can be predicted based on a validated model. However, the validity and predictive power is often restrictive, i.e., *obtaining an expected target feature might not be possible.*

6. *Optimization*: After having a model structure with a predictive character, when estimating unknown or desired parameters, minimizing or maximizing, namely, of fluxes or some metabolites, is required. Thus, it is an *optimization* problem.

In general, the following has to be defined before starting to solve our mathematical model:

- parameter and concentration range,
- external conditions and input, and
- physiological modes for which it holds.

These definitions construct our model in a systemic way. As a rule, it is better to have the *simplest* model with reasonable accuracy [31]. Several modelling tools are developed to achieve this, e.g., the *mechanistic approach*, metabolic control analysis (*MCA*) and metabolic flux analysis (*MFA*). In this study after giving basic definitions of these types, we will focus on nonstationary mechanistic models, which we construct our mathematical model by.

1.2.2 Metabolic Flux Analysis

MFA is a powerful tool in metabolic engineering and has an important place in bioengineering sciences. If only extracellular flux data are available and intracellular fluxes are to be estimated, metabolic flux analysis is a helpful tool for evaluating these unknown fluxes. Thus, this data analysis provides to extend the

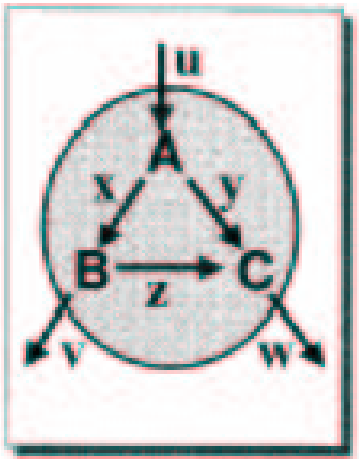


Figure 1.1: Example of fluxes for a simple reaction network where stoichiometric relations of fluxes are $u = x + y$, $v = x - z$, $w = y + z$ [31].

information about other pathways, which are connected in cellular metabolism and, moreover, it helps to obtain a detailed description of the metabolic state of the cell. Furthermore, MFA helps to understand the effect of genetic manipulations and suggests any other modifications. Here, a basic approach to MFA is established on stoichiometry coefficients of biochemical reactions in the pathway. It obeys the fundemental law of mass conservation and the application of optimization principles. By the measured fluxes, MFA defines the stoichiometric relations and if the measurements are not redundant, then all intracellular metabolic fluxes can be estimated from the experimental data. This rises to solve a classical linear estimation problem. When one applies stoichiometric relations, one should consider forward and backward fluxes [31]. In Figure 1.1, there is an example of fluxes for a simple reaction.

The aim of MFA is to identify detailed fluxes and quantify all intracellular and extracellular fluxes over the metabolic network. According to this flux map, possible identifications of genetic manipulations and comparisons of different flux maps can be performed. Furthermore, it is possible to judge or draw conclusion about already manipulated genes in a cellular mechanism [32].

This method is simply based on the given stoichiometry of a system and

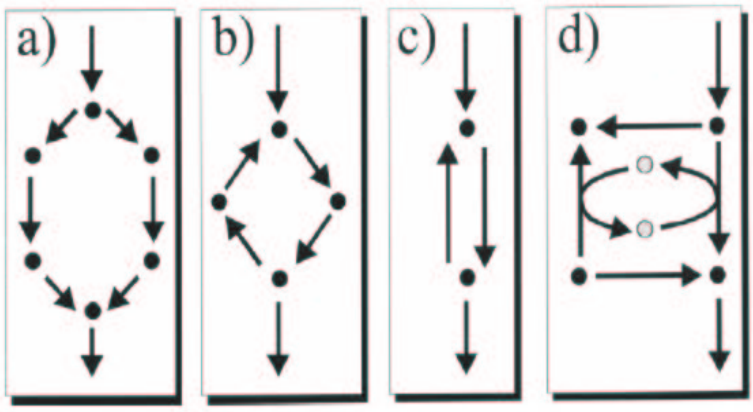


Figure 1.2: Typical situations in which stoichiometric MFA fails [32].

the only necessary assumption is that the biological system is in a *stationary* or *quasistationary* state, which means that the concentrations of metabolites in the pathways do not change over the time. Under these assumptions, intracellular fluxes balance the extracellular fluxes which result from a linear system of equations [32].

Unfortunately, stoichiometric MFA is strongly limited and it fails. In Figure 1.2 taken from [32], a typical situation in which stoichiometric MFA fails. These are:

1. *Parallel pathways without any related flux measurement.* It is impossible to find a value of any flux from two branch fluxes where none of the branches is coupled to a measurable variable in a parallel metabolic pathway (Figure 1.2a)).

2. *Certain metabolic cycles.* "'*Metabolic cycles which are not coupled to measurable fluxes can not be resolved*"' [32]. In Figure 1.2b), flux is not sufficient to determine the fluxes in the metabolic cycle.

3. *Bidirectional reaction steps.* Bidirectional reactions are the special case of metabolic cycles where reactions occur in both directions at the same time (Figure 1.2c)).

4. *Split pathways when cofactors are not balanced.* When some metabolites are not in balance, pathway splits up (Figure 1.2d)). For example, in the glycolysis, PEP, the citric acid cycle, it is only possible to determine all the fluxes if ATP, NADPH, NADH, ... are balanced together with the other metabolites. Thus, energy producing and consuming reactions and all the conversion reaction with the energy metabolites must be exactly known [32].

1.2.3 Metabolic Control Analysis

MCA requires smaller a number of experiments and parameter values than the other methods. It introduces some control coefficients and elasticitices when getting responses to perturbation or manipulation of metabolic parameters and defining the kinetic properties of metabolic pathways.

The mathematical definition of biochemical systems allows us to understand the unexpected experimental results and provides to see the responses of modifications in genes. MCA describes these relations in the form of some indices (elasticities and control coefficients). These indices can be computed from experiments in the steady state. However, the prediction made by the steady state condition is limited to small changes in the parameters [11].

In [9], MCA is defined as the analysis of the sensitivity of metabolites. It is a local linear approach that does not give accurate results if large changes are made in the parameters. It is focussed on the steady state level of the system. However, true steady states never occur in real experiments and the irrelevance of matching of the quasi-steady state to the true steady state always remains. It is mentioned in [9]: *"This problem and its consequences for interpreting measures control coefficients or elasticities are largely ignored by modellers and experimenters. For this and other reasons, it is not yet clear what the experimental side of metabolic control analysis will contribute to our understanding of metabolism and its regulation in the long run."'*

1.2.4 Non-Stationary Mechanistic Models

The last method which plays a fundamental role in this thesis is based on solving a time-dependent model. If the model is formed by the short-time behaviour of microorganism under rapidly changing external conditions, *time-dependent* models are required. The general non-stationary model extends the stationary model into a differential equation:

$$\dot{x} = N \cdot v(x, k, e),$$

where x is the vector of all metabolites, N is the stoichiometric matrix and v is the rate vector of reactions, which depends on metabolites x, the vectors k and e denote the parameters and external metabolites, respectively. In this study, external metabolites are assumed to be constant and are neglected.

Mechanistic models have been used for the analysis of metabolic systems to predict future results for a given present data or to define the kinetic behaviour of the pathway with respect to time. This kind of model requires the knowledge about the kinetic parameters like rate constants and rate expressions dependent on the variables such as metabolite concentrations and metaboic fluxes. In order to achieve these parameters, numerous experimental data are required. Moreover, validity of approximated kinetic parameters from experiments is important for the robustness of the model. If one tries to model the metabolic pathway depending on parameter values, one can be faced with a large number of parameters. The concentration of metabolites over time can be computed by solving either DAE's or after eliminating the constraints solving a system of ODE's. It is not an easy task to obtain solutions when the number of equations is quite large, and stiff problems occur in many metabolic systems. In this work, for the simulation of the metabolites we have used the DAE solver *ode15s* based on DASSL (backward differentiation methods).

Chapter 2

Modelling Approaches

2.1 A Model of Enzyme Kinetics

Before constructing our kinetic model, some biological properties and biochemical laws and formulas should be known for the robustness of the model. In this section, basic concepts of biochemistry are discussed such as steady state kinetics, Michaelis Menten equation and its derivation, enzyme catalyzed reactions and kinetics of multi substrate enzymes.

2.1.1 Steady State Kinetics

In a cell, if the production and degradation of metabolite rates are very rapid and approximately at the same level, then the concentration of a metabolite is at *steady state level*. This shows that the catalytic activity of an enzyme under steady state conditions in the cell is very important for the understanding of metabolism.

Measurement of steady-state reactions are easy, because at a steady state level the rate of the reaction is constant. Steady state levels show more similar properties to metabolic levels. So it is possible to understand the behaviour or function of the enzyme in the reaction by using parameters. Moreover, it is predictable to learn how an enzyme works under different conditions by changing

substrate concentrations [35].

In the following, we will describe the *Michaelis Menten equation* which is fundamental for enzyme kinetics. It can be descibed by the following kinetic mechanism:

$$E + S \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} ES \xrightarrow{k_2} P + E.$$

In the above reaction, E denotes *enzyme*, S denotes *substrate*, P denotes *product* and ES denotes *enzyme substrate complex*. The k_i 's are the rate constants to be determined by the experiments or from literature.
Michaelis-Menten equation assumes the following relations:

- The concentration of an enzyme $[E]$ is much smaller than the substrate $[S]$; i.e., $[E] << [S]$.
- The rate of the reaction is proportional with the initial enzyme concentration; i.e., $v \propto [E]_0$.
- At sufficiently low substrate concentration $[S]$, v increases linearly with $[S]$.
- Basic equation of enzyme kinetics is given by [8]:

$$v = \frac{v_{max}[S]}{K_M + [S]}, \tag{2.1.1}$$

  where K_M is the Michaelis constant:

$$K_M = \frac{k_{-1} + k_2}{k_1}. \tag{2.1.2}$$

We will derive the Michaelis Menten equation by describing the kinetic behaviour of the enzymes.

Let the rate of formation of the product P be equal to the initial velocity of

the reaction, i.e.:

$$\frac{d[P]}{dt} = v_0,$$

where

$$v_0 = k_2[ES]. \tag{2.1.3}$$

Here, the above relations hold. Then, the *mass conservations* for enzyme and substrate are:

$$[E]_{total} = [ES] + [E]_{free}, \tag{2.1.4}$$

$$[S]_0 = [S]_{free} + [ES] + [P], \approx [S]_0 = [S]_{free}. \tag{2.1.5}$$

The change of enzyme substrate complex $[ES]$ concentration with respect to time $\frac{d[ES]}{dt}$ is equal to the difference of rate of formation $k_1[E][S]$ and rate of consumption $k_{-1}[ES] - k_2[ES]$. The velocity of the reaction remains constant during the initial state of the reaction; thus, we have

$$\frac{d[ES]}{dt} = 0. \tag{2.1.6}$$

From (2.1.6) we get following relation

$$k_{-1}[ES] + k_2[ES] = k_1[E][S]. \tag{2.1.7}$$

If we gather the terms with $[ES]$, we obtain

$$(k_{-1} + k_2)[ES] = k_1[E][S].$$

Then, dividing both sides to the k_1 and $[ES]$,

$$\frac{(k_{-1} + k_2)}{k_1} = \frac{[E][S]}{[ES]}.$$

Let call the left hand side term K_M and rewrite the previous equation as

$$K_M = \frac{[E][S]}{[ES]}. \tag{2.1.8}$$

Multiplying both sides by $[ES]$ and substituting $[E]$ defined in (2.1.4) to (2.1.8) with collecting $[ES]$ terms together, gives us following relation:

$$(K_M + [S])[ES] = [E][S].$$

Finally, dividing both sides by $K_M + [S]$, we get the following relation:

$$[ES] = \frac{[E][S]}{K_M + [S]}. \tag{2.1.9}$$

In addition to this, v_{max} occurs when $[ES] = [E]_{total}$.
Substituting (2.1.9) to (2.1.3) to get final rate equation

$$v_0 = \frac{k_2[E][S]}{K_M + [S]}. \tag{2.1.10}$$

Let us call $k_2[E] = v_{max}$ and, herewith, we rewrite (2.1.3) by

$$v_0 = \frac{v_{max}[S]}{K_M + [S]}. \tag{2.1.11}$$

2.1.2 Enzyme Catalyzed Reactions

If a small amount of enzyme is used and all but one substrate is kept constant, then the rate of the enzymatically catalyzed reaction depends on the substrate concentration and initial rate like in the equation (2.1.1).

The typical notation of the enzyme catalyzed reaction with one substrate can be given as

$$A + E \underset{k_2}{\overset{k_1}{\longleftrightarrow}} X \underset{k_4}{\overset{k_3}{\longleftrightarrow}} P + E,$$

where A is *substrate*, E is *enzyme*, X is *enzyme-substrate complex* and P is product.

The kinetic equations consist of

$$\begin{aligned} \frac{d[A]}{dt} &= k_2[X] - k_1[A][E], \\ \frac{d[E]}{dt} &= (k_2 + k_3)[X] - (k_1[A] + k_4[P])[E], \\ \frac{d[P]}{dt} &= k_3[X] - k_4[P][E], \end{aligned} \tag{2.1.12}$$

with a conservation relation given in [25]:

$$[E] + [X] = [E]_{total}. \tag{2.1.13}$$

It is obvious that the derivative of a substrate with respect to time gives the rate. Thus, the rate is a function of compounds c, s (intracellular and extracellular), enzyme concentrations E and kinetic parameters k. However, the enzyme concentration is hidden in the kinetic constants in the parameter vector k; herewith, we can write v as a function of c, s and P, i.e., $v = v(c, s, k)$ [14]. The more general form of (2.1.12) can be written in the form of

$$\begin{aligned} \frac{d[A]}{dt} &= v_2 - v_1, \\ \frac{d[E]}{dt} &= v_2 + v_3 - v_1 - v_4, \\ \frac{d[X]}{dt} &= v_1 + v_4 - v_2 - v_3, \\ \frac{d[P]}{dt} &= v_3 - v_4, \end{aligned}$$

with a conservation equation $[E] + [X] = [E]_{total}$.
The form of rate equations is as follows:

$$\begin{aligned} v_1 &= k_1[A][E], \\ v_2 &= k_2[X], \\ v_3 &= k_3[X], \\ v_4 &= k_4[P][E]. \end{aligned}$$

2.1.3 Kinetics of Multi-substrate Enzymes

Introduction

In previous subsections, only one substrate enzyme kinetics has been discussed. Now, in this subsection, *more* than one substrate enzyme kinetics will be studied, which possesses more complex mechanisms and reveals a large number of rate equations. Typical enzymes play a role as catalysts in such systems to convert two substrates into product. Two or three substrates are common, but not four. The usual approach while studying an enzyme with more than one substrate is that all substrates are varying during a reaction but one of them is kept constant. By this way it is observed that most reactions obey *Michaelis-Menten kinetics* [8]. So, the remaining substrate is varied in kinetic assays and valuable information about the relationship between enzyme, substrate and inhibitors can be gained [36]. There are some types of reactions involving their variability to bind enzymes and producing a certain product. In the following subsection, the schematic descriptions and rate equations of each of these multi-substrate enzyme reactions, which are mainly divided in sequential and ping pong mechanisms, will be given [8].

Ping Pong Mechanism

Definition 2.1.1. If at least one product is released before all of the substrates have been bound, then it is called ***ping pong mechanism*** [36]. The process usually starts by binding of the enzyme to the first substrate:

$$E + A \rightleftharpoons (EA).$$

The next reaction is the key of the whole process:

$$(EA) \rightleftharpoons (FP).$$

In this reaction, a part of the substrate has been removed from substrate A, converting it to product P. The removed section has become covalently bound to the enzyme to create a new form of the enzyme, enzyme F. The first product of the reaction is now released and the second substrate binds:

$$(FP) \rightleftharpoons F + P,$$

$$F + B \rightleftharpoons (FB).$$

Now, the stored section of the first substrate is transferred to the second substrate to create the second product, which is then released:

$$(FB) \rightleftharpoons (EQ),$$

$$(EQ) \rightleftharpoons E + Q.$$

A plotting of these reactions is shown in Figure 2.1.

Sequential Mechanisms

Definition 2.1.2. If all the substrates bind to the enzyme before the first product is formed, this is called a ***sequential reaction***.

Definition 2.1.3. If all substrates bind to enzymes and products are dissociated in an obligatory order, the sequential reaction is called ***ordered***.

Definition 2.1.4. If a sequential mechanism has no any obligatory order in binding and releasing, it is called ***random sequential mechanism***.

The terminolgy for sequential mechanisms was developed by Cleland (1963) for multisubstrate enzymes according to their number of substrates, products, binding type of substrates to enzymes and dissociation of products from enzymes [38]. According to this terminology, substrate and product numbers are called by *Uni, Bi, Ter* and *Quad*. For example, for two substrates and two products, our mechanism is called a *Bi Bi* mechanism, for one substrate and two products

it is called a *Uni Bi* mechanism. In this study, enzymes in our model follow Ordered Bi Bi and Rapid Equilibrium Random Bi Bi Mechanism. So, in the following sections, these two types of mechanisms will be introduced. Schematic representations of other mechanisms will be given in Figures 2.1, 2.2 and 2.3 [38]. The rate equations are in detail presented in [16].

Ordered Bi Bi Mechanism

In this kind of reaction mechanism, substrates bind to enzymes in order and, then, an enzyme substrate complex is formed. In addition to this, products dissociates in order again. As you will see in Figure 2.1, firstly substrate A binds to enzyme E and forms enzyme substrate complex EA. Afterwards, second substrate B binds to EA and forms second enzyme substrate complex EAB. The products leave the enzyme complex *in order*, namely, first product P is produced and, then, Q.

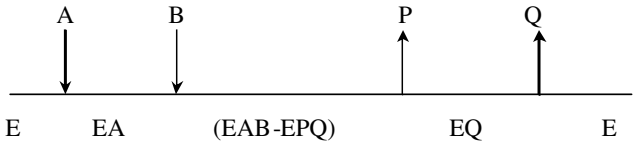
According to the steady state assumption and conservation mass law, differential equations of the reaction mechanism can be written in the form by the explicit reaction scheme given in Figure 2.4 [38]:

$$\begin{aligned} \frac{d[EA]}{dt} &= k_1[E][A] + k_4[EAB] - [EA](k_2 + k_3[B]) = 0, \\ \frac{d[EAB]}{dt} &= k_3[EA][B] + k_6[EQ][B] - [EAB](k_4 + k_5) = 0, \\ \frac{d[EQ]}{dt} &= k_8[E][Q] + k_5[EPQ] - [EQ](k_7 + k_6[P]) = 0, \end{aligned} \tag{2.1.14}$$

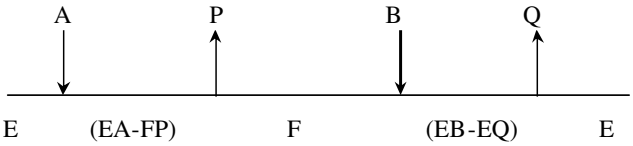
and

$$\begin{aligned} [E_0] - ([E] + [EA] + [EAB] + [EQ]) &= 0, \\ v = \frac{d[Q]}{dt} &= k_7[EQ] - k_8[E][Q]. \end{aligned} \tag{2.1.15}$$

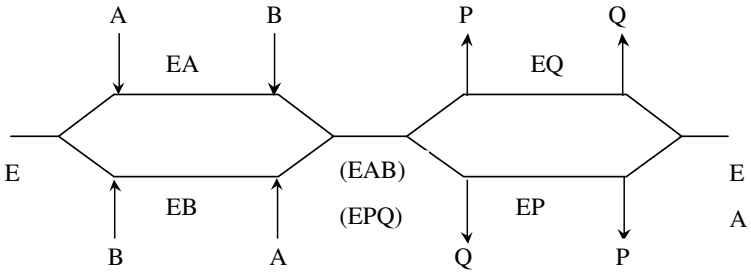
The rate equation can be written as a function of internal and external metabolites, and parameters can be determined by some improved methods such as King Altman and Cleland method. In [25], formulas and proofs are given in a



Ordered Bi Bi Mechanism.



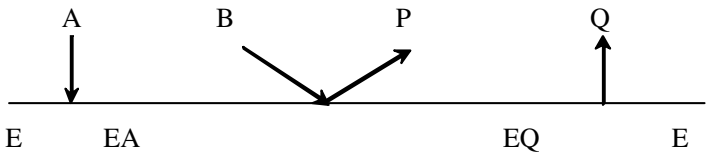
Ping Pong Mechanism.



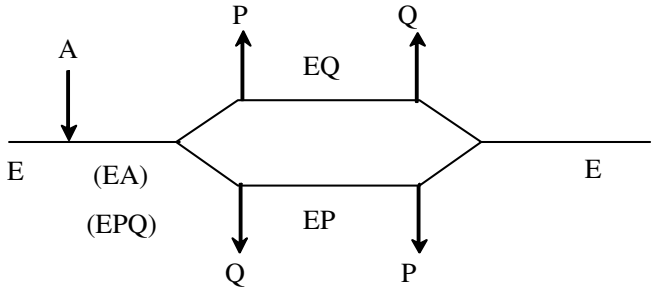
Random ordered Bi Bi Mechanism.

Ordered Uni Bi Mechanism.

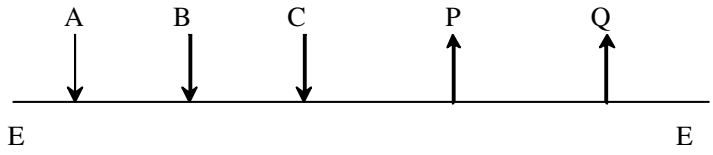
Figure 2.1: Schematic explanation of different types of reaction mechanisms (taken from [38]).



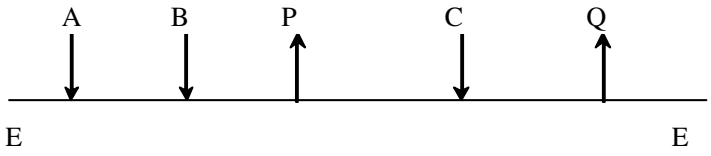
Theorell Chance Mechanism.



Random Ordered Uni Bi Mechanism.

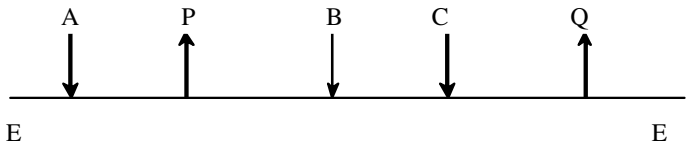


Ordered Ter Bi Mechanism.

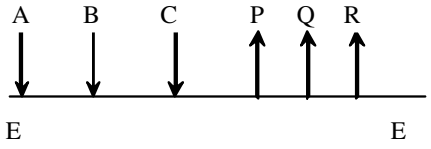


Bi Uni Uni Uni Ping Pong Mechanism.

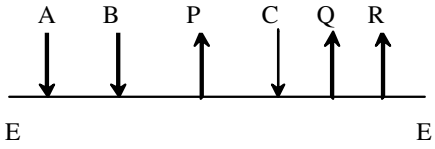
Figure 2.2: Schematic explanation of different types of reaction mechanisms (taken from [38]).



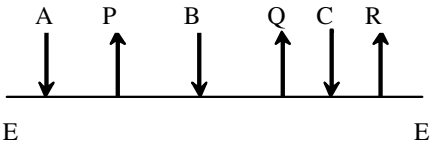
Uni Uni Bi Uni Ping Pong Mechanism.



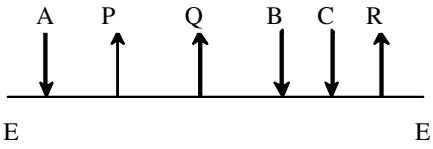
Ordered Ter Ter Mechanism .



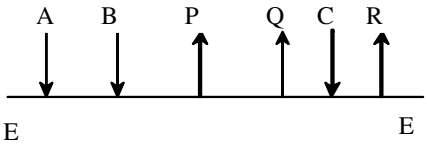
Bi Uni Uni Bi Ping Pong Me chanism .



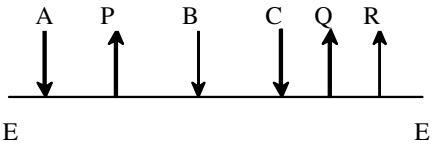
Uni Uni Uni Uni Uni Uni Me chanism .



Uni Bi Bi Uni Ping Pong Me chanism .



Bi Bi Uni Uni Me chanism .



Uni Uni Bi Bi Ping Pong Me chanism .

Figure 2.3: Schematic explanation of different types of reaction mechanisms (taken from [38]).

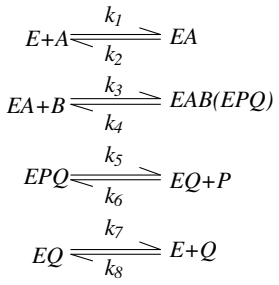


Figure 2.4: Explicit reaction scheme for ordered Bi Bi mechanism [38].

more detailed way. In [38], they are stated explicitly as follows:

$$X = 1 + \frac{A}{K_I^A} + \frac{K_M^A B}{K_I^A K_M^B} + \frac{K_M^Q P}{K_M^P K_I^Q} + \frac{Q}{K_I^Q} + \frac{AB}{K_I^A K_M^B} + \frac{K_M^Q BP}{K_I^A K_M^P K_I^Q},$$

$$Y = \frac{K_M^A BQ}{K_I^A K_M^B K_I^Q} + \frac{PQ}{K_M^P K_I^Q} + \frac{ABP}{K_I^A K_M^B K_I^P} + \frac{BPQ}{K_I^B K_M^P K_I^Q},$$

$$v = \frac{\frac{V_{max}^f AB}{K_I^A K_M^B} - \frac{V_{max}^r PQ}{K_M^P K_I^Q}}{X + Y},$$

where K_I's are inhibition constants and K_M's are Michaelis Menten constants. The explicit form of the parameters is given by Bowden (1979) and stated below [38].

$$
\begin{aligned}
V_{max}^{f} &= k_5 k_7 E_0/(k_5+k_7),\\
V_{max}^{r} &= k_2 k_4 E_0/(k_2+k_4),\\
K_M^A &= k_5 k_7/(k_1(k_5+k_7)),\\
K_M^B &= (k_4+k_5)k_7/(k_3(k_5+k_7)),\\
K_M^P &= (k_4+k_5)k_2/(k_2+k_4)k_6,\\
K_M^Q &= k_2 k_4/(k_8(k_2+k_4)),\\
K_I^A &= k_2/k_1,\\
K_I^B &= (k_2+k_4)/k_3,\\
K_I^P &= (k_5+k_7)/k_6,\\
K_I^Q &= k_7/k_8.
\end{aligned}
$$

Rapid Equilibrium Random Bi Bi Mechanism

In this mechanism, the enzyme binds to substrates and dissociates from them randomly. A free enzyme can bind either to the first substrate or to the second substrate. Since it is difficult to obtain rate equations with the steady state assumption, a rate equation is derived from equilibrium assumption. According to this assumption, except $EAB \longleftrightarrow EPQ$, all reactions are assumed to be in the equilibrium state [38].

Definition 2.1.5. In 1913, Lenor Michaelis and Maude Menten, advancing earlier work of the chemist Victor Henri, assumed that $k_{-1} >> k_2$. If this is true, then the reversible step in the mechanism does achieve an equilibrium and we can write the law of chemical equilibrium for this reversible step and, hence, equate the ratio between the forward (k_1) and the reverse (k_{-1}) rate constants with the equilibrium expression [37]:

$$K_s = \frac{k_{-1}}{k_1} = \frac{[E][S]}{[ES]},$$

where all variables are the same as in the mechanism fron Section 2.1.2 and K_s is the equilibrium constant. In [37], it is found that rate equation is

$$v = (\frac{k_2}{K_s})[E][S],$$

where $k_{-1} >> k_2$. The rate equation and kinetic parameters are given in [38]:

$$v = \frac{\frac{V^f_{max}AB}{K^A_I K^B_M} - \frac{V^r_{max}PQ}{K^P_M K^Q_I}}{1 + \frac{A}{K^A_I} + \frac{B}{K^B_I} + \frac{P}{K^P_I} + \frac{Q}{K^Q_I} + \frac{AB}{K^A_I K^B_M} + \frac{PQ}{K^P_M K^Q_I}},$$

where

$$\begin{aligned}
V^f_{max} &= C_{E_0} k_1, \\
V^r_{max} &= C_{E_0} k_2, \\
K^A_I &= [A][E]/[EA], \\
K^B_I &= [B][E]/[EB], \\
K^A_M &= [EB][A]/[EAB], \\
K^P_I &= [P][E]/[EP], \\
K^Q_I &= [Q][E]/[EQ], \\
K^P_M &= [EQ][P]/[EPQ],
\end{aligned}$$

and

$$[E_0] = [E] + [EA] + [EB] + [EAB] + [EP] + [EQ] + [EPQ], \quad (2.1.16)$$

$$\frac{dP}{dt} = v = k_1[EAB] - k_2[EPQ]. \quad (2.1.17)$$

2.1.4 The Stoichiometric Matrix

The chemical reaction model can be described with the matrix representation which is called a *stoichiometric matrix* N. Each chemical reaction in the model network has a stoichiometric coefficient which comes before the reactants. The

entries of our $m \times n$ matrix N are constructed by the following rule:

$$N_{i,j} = \begin{cases} +c, \text{ if the reaction produces metabolite } X_i \\ -c, \text{ if the reaction consumes metabolite } X_i \\ 0, \text{ if the reaction neither produces nor consumes metabolite } X_i, \end{cases}$$

where m denotes the number of metabolite, n denotes the number of reactions and c is the stoichiometric constant.

The stoichiometric matrix N is a tool for deriving conservation relations of a chemical system. It can be reduced to a simple matrix formed by identitiy elements and zero elements to get conservation relations. With the help of the stoichiometric matrix N, the rate vector v and general kinetic equations discussed above, the following equation can be derived:

$$\dot{x} = N \cdot v(c, k). \tag{2.1.18}$$

Here, x is the vector of all metabolites, N is the stoichiometric matrix and v is the vector of all rates.

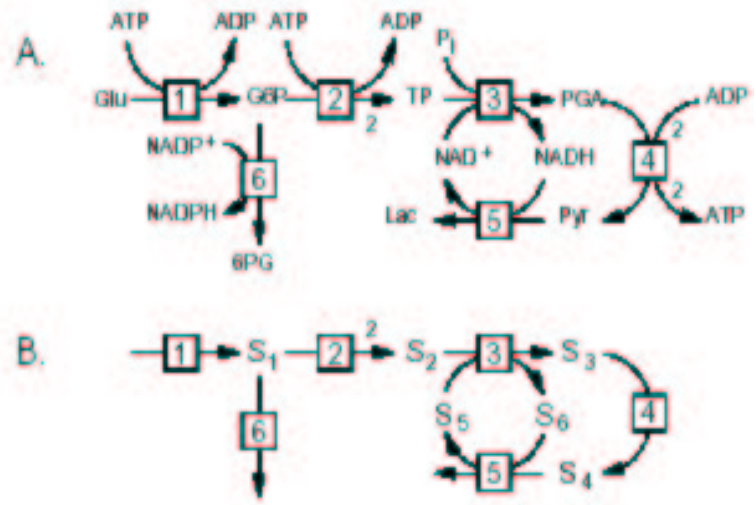


Figure 2.5: *A.* skeleton model of glycolysis, *B.* This simplified representation includes only metabolites (taken from [13]).

2.2 Deriving Conservation Relations

The previous chapter gives an introductory idea of modeling tools and rate equations. In this chapter, conservation relations, stoichiometric matrix analysis with conservation relations will be introduced by an example, namely, a skeleton model of mammalian glycolysis which is represented in Figure 2.5 [13].

2.2.1 Deriving Conservation Relations by Inspection

We assume that the external metabolites glucose, lactate, ADP, ATP, Pi, NADP+ and NADPH are fixed parameters and, therefore, are not included in the kinetic model. The model equation of the system in Figure 2.5 is of the

following form:

$$\begin{aligned}
\frac{ds_1}{dt} &= v_1 - v_2 - v_6, && (2.2.19)\\
\frac{ds_2}{dt} &= 2v_2 - v_3, && (2.2.20)\\
\frac{ds_3}{dt} &= v_3 - v_4, && (2.2.21)\\
\frac{ds_4}{dt} &= v_4 - v_5, && (2.2.22)\\
\frac{ds_5}{dt} &= v_5 - v_3, && (2.2.23)\\
\frac{ds_6}{dt} &= v_3 - v_5. && (2.2.24)
\end{aligned}$$

Matrix Formulation by Stoichiometricity

A matrix formulation of the kinetic model is derived by the above differential equations. As it is mentioned before, the kinetic model can be written in the form with the stoichiometric matrix and the vector of metabolites, namely,

$$\dot{x} = N \cdot v(C, K).$$

As we noted, the numbers on the right-hand side matrix are stoichiometric coefficients which come before the reactant in the chemical reaction and their signs are decided by the rule mentioned before:

$$\begin{bmatrix} \frac{ds_1}{dt} \\ \frac{ds_2}{dt} \\ \frac{ds_3}{dt} \\ \frac{ds_4}{dt} \\ \frac{ds_5}{dt} \\ \frac{ds_6}{dt} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}. \qquad (2.2.25)$$

Let us consider the fifth and sixth equations

$$\frac{ds_5}{dt} = v_5 - v_3, \qquad (2.2.26)$$

$$\frac{ds_6}{dt} = v_3 - v_5. \tag{2.2.27}$$

It is obvious that

$$\frac{ds_5}{dt} + \frac{ds_6}{dt} = 0. \tag{2.2.28}$$

Equation (2.2.28) can be also written as

$$\frac{d}{dt}(s_5 + s_6) = 0. \tag{2.2.29}$$

From Calculus it is obvious that if the derivative of a smooth function over an interval is zero, then it must be constant.
Thus,

$$s_5 + s_6 = T_1, \tag{2.2.30}$$

where T_1 is a constant. We call equation (2.2.30) a *conservation equation*. In physical manner, this equation states that although the concentrations of s_5 and s_6 change with the state of the system, the sum of their concentrations must remain constant.
The second constraint is not so obvious as the first:

$$\frac{ds_3}{dt} + \frac{ds_4}{dt} + \frac{ds_5}{dt} = v_3 - v_4 + v_4 - v_5 + v_5 - v_3 = 0, \tag{2.2.31}$$

so that

$$s_3 + s_4 + s_5 = T_2. \tag{2.2.32}$$

This procedure is not obvious for all enzymatic reactions, but it gives complex and big systems. So, we should generalize this for big systems. Note that differential equations which turned into conservation relations are linearly dependent. Namely, one can be written in terms of the others, e.g., by (2.2.28) and (2.2.31),

$$\frac{ds_6}{dt} = -\frac{ds_5}{dt} \tag{2.2.33}$$

$$\frac{ds_4}{dt} = -\frac{ds_3}{dt} - \frac{ds_5}{dt}. \tag{2.2.34}$$

2.2.2 Deriving Conservation Relations by Gaussian Elimination

For simple models, conservation relations can be found just by inspection, but it is not so easy for complex systems. A more systematic procedure is needed for big systems. In this section, the way of getting conservation relations provides to partition the fluxes and concentrations into dependent and independent sets.

The procedure is formally called *Gaussian elimination* which we know from Linear Algebra. Gaussian elimination is applied to our stoichiometric matrix; then, (2.2.25) can be rewritten as:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{ds_1}{dt} \\ \frac{ds_2}{dt} \\ \frac{ds_3}{dt} \\ \frac{ds_4}{dt} \\ \frac{ds_5}{dt} \\ \frac{ds_6}{dt} \end{bmatrix}. \tag{2.2.35}$$

While Gaussian elimination is applied to the stoichiometric matrix, same elementary row operations must simultaneously be applied on the identity matrix. By performing this we have recorded which multiple of which row was added to other ones. Hence, we get linearly dependent or independent metabolites. This means that we only have to handle the following matrices [13]:

	R_1	R_2	R_3	R_4	R_5	R_6	$\dot{s_1}$	$\dot{s_2}$	$\dot{s_3}$	$\dot{s_4}$	$\dot{s_5}$	$\dot{s_6}$
S_1	1	-1	0	0	0	-1	1	0	0	0	0	0
S_2	0	2	-1	0	0	-1	0	1	0	0	0	0
S_3	0	0	1	-1	0	0	0	0	1	0	0	0
S_4	0	0	0	1	-1	0	0	0	0	1	0	0
S_5	0	0	-1	0	1	0	0	0	0	0	1	0
S_6	0	0	1	0	-1	0	0	0	0	0	0	1

(2.2.36)

In terms of this matrix formulation, Gaussian elimination has to perform row manipulations on the whole matrix until a row reduced echelon form is obtained. In our model, a row reduced echelon form is performed by matlab command *rref*. However, this command does not pay attention to row substraction. In fact, if row substraction is done during Gaussian process, it is not going to ensure the added elements of the identity matrix being positive. In biological terms, concentrations in the conservation relationships must be added, not substracted. But with *rref* command of Matlab, we can see which variables are dependent or independent. We should stop the program if concentrations would be substracted. Although subtracting is not a problem for computational purposes, it is not at all equivalent in terms of physcial and biological interpretation.

If we apply the above procedure without substracting the rows, then we obtain the following matrices:

	R_1	R_2	R_3	R_4	R_5	R_6	$\dot{s_1}$	$\dot{s_2}$	$\dot{s_3}$	$\dot{s_4}$	$\dot{s_5}$	$\dot{s_6}$
S_1	1	-1	0	0	0	-1	1	0	0	0	0	0
S_2	0	2	-1	0	0	0	0	1	0	0	0	0
S_3	0	0	1	-1	0	0	0	0	1	0	0	0
S_4	0	0	0	1	-1	0	0	0	0	1	0	0
S_5	0	0	0	0	0	0	0	0	1	1	1	0
S_6	0	0	0	0	0	0	0	0	0	0	1	1

(2.2.37)

Thus, we have transformed (2.2.35) into the following form:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{ds_1}{dt} \\ \frac{ds_2}{dt} \\ \frac{ds_3}{dt} \\ \frac{ds_4}{dt} \\ \frac{ds_5}{dt} \\ \frac{ds_6}{dt} \end{bmatrix}. \tag{2.2.38}$$

2.2.3 Deriving Conservation Relations by Partioning the Stoichiometric Matrix into Submatrices

In this subsection, conservation relations are derived by dividing the stoichiometric matrix N from the model equation (2.1.18) into submatrices.
Let N be an $m \times n$ matrix represented by

$$N = L \cdot N_R, \tag{2.2.39}$$

where N_R is an $m_0 \times r$ matrix and r is the dimension of rate vector v. Thus, L can be written in the following form:

$$L = \begin{bmatrix} I_{m_0} \\ L_0 \end{bmatrix}, \tag{2.2.40}$$

where I_{m_0} is the $m_0 \times m_0$ identitiy matrix and L_0 is an $(m - m_0) \times m_0$ matrix. It is clear that we can divide the vector X by the following subvectors because of the relationships with N:

$$x = \begin{bmatrix} x_{R1} \\ x_{R2} \end{bmatrix}, \tag{2.2.41}$$

where x_{R1}, x_{R2} are $m_0 \times 1$ and $(m-m_0) \times 1$ column vectors, respectively. Therefore, according to these definitions and by the help of the following theorem, conservation relations can be determined by using the matrix L_0 [38, 39].

Theorem 2.2.1. *Every conservation relation in multi enzymatic systems can be written by the linear composition of* $(m - m_0)$ *linearly independent* conservation relations *defined by the following equation:*

$$\frac{d}{dt}(x_{R2} - L_0 x_{R1}) = 0. \tag{2.2.42}$$

Proof. We know that

$$L = \begin{bmatrix} I_{m_0} \\ L_0 \end{bmatrix} \tag{2.2.43}$$

and

$$x = \begin{bmatrix} x_{R_1} \\ x_{R_2} \end{bmatrix}. \tag{2.2.44}$$

If we put these into (2.1.18) we get

$$\frac{d}{dt}\begin{bmatrix} x_{R_1} \\ x_{R_2} \end{bmatrix} = \begin{bmatrix} I_{m_0} \\ L_0 \end{bmatrix} \cdot N_R v. \tag{2.2.45}$$

Thus,

$$\frac{d[x_{R_1}]}{dt} = N_R v \tag{2.2.46}$$

and

$$\frac{d[x_{R_2}]}{dt} = L_0 N_R v. \tag{2.2.47}$$

If in (2.2.47) we write the left-hand side of (2.2.46) instead of $N_R v$, we have

$$\frac{d[x_{R_2}]}{dt} - L_0 \frac{d[x_{R_1}]}{dt} = 0. \tag{2.2.48}$$

Rearranging this equation, we obtain

$$\frac{d}{dt}(x_{R2} - L_0 x_{R1}) = 0. \tag{2.2.49}$$

□

In case of logical errors, the following algorithm is proposed in [38]:

Algorithm:

Let N be an $m \times n$ matrix and its rank be m_0. The rank of N can be computed easily by the Maple command *rank*. (In this study, Matlab 6.5 is used, and to evaluate rank, the same command *rank* is used.)

- Matrices N_{R1} and N_{R2} are evaluated, where N_{R1} is composed of m_0 linearly

dependent rows of N and N_{R2} consists of the remaining rows of N. Thus, the dimension of N_{R1} and N_{R2} is $m_0 \times r$ and $(m - m_0) \times r$, respectively.

- Now, N_{R1} is divided into submatrices as N_{R11} and N_{R12}. The matrix N_{R11} consists of linearly independent columns of N_{R1}. Thus, the inverse of N_{R11} is an $m_0 \times m_0$ square matrix which can be computed, and N_{R12} is an $m_0 \times (r - m_0)$ matrix.

- Then, N_{R2} is divided into submatrices as N_{R21} and N_{R22}, where N_{R21} consists of columns of N_{R2} during construction of N_{R11} matrix, and N_{R22} is the remaining part of N_{R2} after subtracting N_{R21}. Thus, the dimension of N_{R21} is $(m - m_0) \times m_0$ and N_{R22} is of the type $(m - m_0) \times (r - m_0)$. So, our matrix N is of the following form:

$$N = \begin{bmatrix} N_{R11} & N_{R12} \\ N_{R21} & N_{R22} \end{bmatrix}. \tag{2.2.50}$$

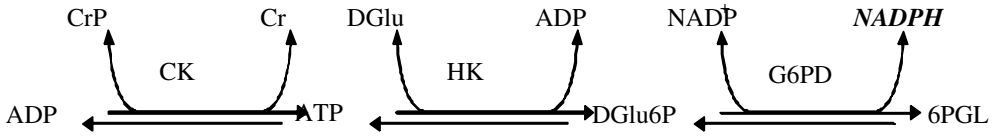
- Furthermore, L_0 can be computed by the formula :

$$L_0 = N_{R21}.(N_{R22})^{-1}. \tag{2.2.51}$$

- Finally, x can be divided as before:

$$x = \begin{bmatrix} x_{R1} \\ x_{R2} \end{bmatrix}, \tag{2.2.52}$$

and we compute conservation relations by the equation (2.2.42).



.

Figure 2.6: Creatine Kinase, Hexokinase and Glucose-6Phosphate Dehydrogenase System [38].

2.3 Creatine Kinase, Hexokinase and Glucose-6Phosphate Dehydrogenase System

Up to now, in previous sections, biochemical laws and mechanisms of some types of reactions and how to construct a model by differential equations with deriving conservation relations have been given. In this section, biological information of our main model will be introduced which consists of three enzymes: Creatin Kinase, Hexokinase and Glucose-6Phosphate Dehydrogenase. The overall reaction of this three enzymatic system is illustrated in Figure 2.7 [38] and (2.3.53) in [39]:

$$\begin{aligned} &Creatinphosphate + ADP \overset{creatinekinase}{\leftrightarrow} creatin + ATP, \\ &ATP + D-Glucose \overset{hexokinase}{\leftrightarrow} D-Glucose6-phosphate + ADP, \\ &D-Glucose6-phosphate + NADP \overset{G6PDH}{\leftrightarrow} D-Gluconate6-phoshate + NADPH. \end{aligned} \quad (2.3.53)$$

2.3.1 Creatine Kinase System

Creatine Kinase (EC 2.7.3.2) (CK) plays an important role in energy metabolism of tissues such as skeletal and cardiac muscle neural tissues like brain and retina by providing regeneration of ATP. It catalyzes the reversible transfer of the phosphoryl group from phosphocreatine to ADP, to regenerate ATP. There are types of CK isoenzymes differring according to the place of ATP production such as mitochondria and cytosol. *"'Brain cytosolic and mitochondrial isozymes are BB-CK and Mia-CK, respectively (Wallimann et al., 1992; Wallimann et al., 1998a)"'* [7]. Energy is important for development or maintaining of the functional celebral activities. Because of this reason, change in the level of CK activity may lead to neuronal loss in brain, which occurs in a neurodegenerative pathway (Tomimoto et al., 1993). Recent studies strengthen this hypothesis by showing that CK activity is severely reduced in some neurodegenerative diseases (David et al., 1998; Aksenov et al., 2000). Cytosolic CK's are divided into muscle type (M) and brain type (B). *"'CK-MM and CK-BB are expressed at high levels in the skeletal muscle, and in the brain and smooth muscle, respectively, and hybrid CK-MB is found in the cardiac muscle"'* [15]. In this study, CK is obtained from a heart of a rat. All experiments are done in the Atatürk University of Applied Biotechnology Research Center. This experimental study can be found in more detail including experimental assays and units in Necmettin Yildirim's research. CK enters the reaction following rapid equilibrium random ordered Bi Bi mechanism. With the help of schematic representation of rapid equilibrium random ordered Bi Bi mechanism presented in Subsection 2.1.3, the rate equations and parameters are introduced by Necmettin Yildirim:

$$
\begin{aligned}
V_{max}^{f} &= C_{E_0} k_1, \\
V_{max}^{r} &= C_{E_0} k_2, \\
K_I^{ADP} &= [ADP][E]/[EADP], \\
K_I^{CrP} &= [CrP][E]/[ECrP], \\
K_I^{ATP} &= [ATP][E]/[EATP], \\
K_I^{Cr} &= [Cr][E]/[ECr], \\
K_M^{ADP} &= [ECrP][ADP]/[EADPCrP], \\
K_M^{ATP} &= [ECr][ATP]/[EATPCr],
\end{aligned}
$$

$$[E_0] = [E]+[EADP]+[ECrP]+[EADPCrP]+[EATP]+[ECr]+[EATPCr],$$

$$\frac{dATP}{dt} = v = k_1[EADPCrP] - k_2[EATPCr].$$

The rate equation is rearranged by Necmettin Yildirim using Gröbner Basis method in Maple [38], which is not studied in this research [12]. It has the following form:

$$v = \frac{\frac{V_{max}^{f}[CrP][ADP]}{K_I^{CrP}K_M^{ADP}} - \frac{V_{max}^{r}[Cr][ATP]}{K_M^{Cr}K_I^{ATP}}}{1 + \frac{[CrP]}{K_I^{CrP}} + \frac{[ADP]}{K_I^{ADP}} + \frac{[Cr]}{K_I^{Cr}} + \frac{[ATP]}{K_I^{ATP}} + \frac{[CrP][ADP]}{K_I^{CrP}K_M^{ADP}} + \frac{[Cr][ATP]}{K_M^{Cr}K_I^{ATP}}}.$$

2.3.2 Hexokinase System

Hexokinase (EC 2.7.1.1) is the first enzyme of the glycolysis reaction catalyzing the conversion of glucose to glucose-6phosphate. It takes phosphate of ATP and binds to the glucose inorder to produce glucose-6phosphate. During taking phosphate from ATP, Hexokinase requires Mg2+, and ATP binds to enzyme with Mg2++ [33]. The enzyme has a low K_M for the sugar substrate (about 0.1 mM) and is inhibited by the product of its reaction, G6P [34]. There are four isoforms of this enzyme in mammalian tissue differing with respect to their function significantly. Isoform I is responsible for rate limiting step in gly-

colysis in brain and red blood cells. The reaction product Glucose-6phosphate inhibits both isoform I and II (but not IV). Inorganic phosphate Pi, however, dissociates from Glucose-6phosphate only by inhibition of isoform I. *"'Thus, among hexokinase isoforms, brain hexokinase exhibits unique regulatory properties in that physiological levels of Pi can reverse inhibition due to physiological levels of Gluc-6-P [1315] "'* [1].

In this research, hexokinase from the yeast cell is used in the experiment. It reacts as rapid equilibrium random Bi Bi mechanism. According to this mechanism, as discussed in the previous section, our mathematical model is:

$$\begin{aligned}
V_{max}^{f} &= C_{E_0} k_1, \\
V_{max}^{r} &= C_{E_0} k_2, \\
K_I^{ATP} &= [ATP][E]/[EATP], \\
K_I^{DGlu} &= [DGlu][E]/[EDGlu], \\
K_I^{ADP} &= [ADP][E]/[EADP], \\
K_I^{DGlu6P} &= [DGlu6P][E]/[EDGlu6P], \\
K_M^{ATP} &= [EDGlu][ATP]/[EATPDGlu], \\
K_M^{DGlu6P} &= [EDGlu6P][ADP]/[EADPDGlu6P],
\end{aligned}$$

$$[E_0] = [E]+[EADP]+[EDGlu6P]+[EADPDGlu6P]+[EATP]+[EDGlu]+[EATPDGlu],$$

$$\frac{dDGlu6P}{dt} = v = k_1[EATPDGlu] - k_2[EADPDGlu6P].$$

After rearranging the rate equation by Maple finding Gröbner basis (for detailed information see [38]) it is reduced to

$$v = \frac{\frac{V_{max}^{f}[DGlu][ATP]}{K_I^{DGlu}K_M^{ATP}} - \frac{V_{max}^{r}[DGlu6P][ADP]}{K_M^{DGlu6P}K_I^{ADP}}}{1 + \frac{[DGlu]}{K_I^{DGlu}} + \frac{[ATP]}{K_I^{ATP}} + \frac{[DGlu6P]}{K_I^{DGlu6P}} + \frac{[ADP]}{K_I^{ADP}} + \frac{[DGlu][ATP]}{K_I^{DGlu}K_M^{ATP}} + \frac{[DGlu6P][ADP]}{K_M^{DGlu6P}K_I^{ADP}}}.$$

2.3.3 Glucose 6-Phosphate Dehydrogenase System

Glucose-6-phosphate Dehydrogenase (G6PDH) (EC 1.1.1.49) is the rate limiting enzyme in pentose phosphate pathway (PPP) providing to control the amount of NADPH. It is important since PPP is the only source for NADPH in the erythrocytes. From the recent studies it is found that G6PDH plays a crucial role against antioxidative stress occuring in intracellular metabolic processes caused by harmful radicals in human erythrocytes. It prevents the forming of such harmful radicals by reducing NADP+ to NADPH. Furthermore, it plays a protective role against the reactive oxygen space in nucleatited eukoryatic cells that results with alternative ways to produce NADPH [40].

G6PDH is taken from human erythrocytes during the experiment, and it obeys the ordered Bi Bi mechanism. By the mechanism given in Subsection 2.1.3, the mathematical model and reduced rate equation of this system is given according to [38]:

$$\begin{aligned}
\frac{d[ENADP^+]}{dt} &= k_1[E][NADP^+] + +k_4[ENADP^+DGlu6P] \\
&\quad -[ENADP^+](k_2 + k_3[DGlu6P]) = 0, \\
\frac{d[ENADP^+DGlu6P]}{dt} &= k_3[E][NADP^+][DGlu6P] + k_6[6PGL][NADPH] \\
&\quad -[ENADP^+DGlu6P](k_4 + k_5) = 0, \\
\frac{d[ENADPH]}{dt} &= k_8[E][NADPH] + k_5[ENADP^+DGlu6P] \\
&\quad -[ENADPH](k_7 + k_6[6PGL]) = 0,
\end{aligned}$$

and

$$[E_0] = [EENADP^+] + [ENADP^+DGlu6P] + [ENADPH] + [E],$$
$$v = \frac{d[NADPH]}{dt} = k_7[ENADPH] - k_8[E][NADPH],$$

$$v = \frac{\frac{V^f_{max}[NADP^+][DGlu6P]}{K_I^{NADP^+}K_M^{DGlu6P}} - \frac{V^r_{max}[6PGL][NADPH]}{K_M^{6PGL}K_I^{NADPH}}}{1 + A + B + C + D + E}.$$

Here, A,B and C are:

$$\begin{aligned}
A &= \frac{[NADP^+]}{K_I^{NADP^+}} + \frac{K_M^{NADP^+} DGlu6P}{K_I^{NADP^+} K_M^{DGlu6P}} + \frac{K_M^{NADPH}[6PGL]}{K_M^{6PGL} K_I^{NADPH}},\\
B &= \frac{[NADPH]}{K_I^{NADPH}} + \frac{[NADP^+][DGlu6P]}{K_I^{NADP^+} K_M^{DGlu6P}},\\
C &= \frac{K_M^{NADPH}[NADP^+][6PGL]}{K_I^{NADP^+} K_M^{6PGL} K_I^{NADPH}},\\
D &= \frac{K_M^{NADP^+}[DGlu6P][NADPH]}{K_I^{NADP^+} K_M^{DGlu6P} K_I^{NADPH}} + \frac{[6PGL][NADPH]}{K_M^{6PGL} K_I^{NADPH}},\\
E &= \frac{[NADP^+][DGlu6P][6PGL]}{K_I^{NADP^+} K_M^{DGlu6P} K_I^{6PGL}} + \frac{[DGlu6P][6PGL][NADPH]}{K_I^{DGlu6P} K_M^{6PGL} K_I^{Q}},
\end{aligned}$$

and the included kinetic parameters are

$$\begin{aligned}
V_{max}^f &= k_5 k_7 E_0/(k_5 + k_7),\\
V_{max}^r &= k_2 k_4 E_0/(k_2 + k_4),\\
K_M^{NADP^+} &= k_5 k_7/(k_1(k_5 + k_7)),\\
K_M^{DGlu6P} &= (k_4 + k_5) k_7/(k_3(k_5 + k_7)),\\
K_M^{6PGL} &= (k_4 + k_5) k_2/(k_2 + k_4) k_6,\\
K_M^{NADPH} &= k_2 k_4/(k_8(k_2 + k_4)),\\
K_I^{NADP^+} &= k_2/k_1,\\
K_I^{DGlu6P} &= (k_2 + k_4)/k_3,\\
K_I^{6PGL} &= (k_5 + k_7)/k_6,\\
K_I^{Q} &= k_7/k_8.
\end{aligned}$$

2.3.4 Mathematical Modeling of CK-HK-G6PDH System

In this subsection, all DAE's will be introduced with given rate equations depending on concentrations and parameters discussed in previous subsections. Since chemical reactions occur with some definite rate changing with parameters

and concentrations over the time, the kinetic behaviour of the reaction can be defined by some ODE's with initial conditions taken during an experiment. As discussed in Chapter 2, differential equations and the stoichiometric matrix of our model are given in the following:
Let x be the vector of metabolites in Figure 2.7:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} = \begin{bmatrix} CrP \\ ADP \\ DGlu6P \\ ATP \\ DGlu \\ Cr \\ NADP^{+} \\ 6PGL \\ NADPH \end{bmatrix},$$

and v be the rate vector of dimension 3, since we have 3 overall reactions:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

The stoichiometric matrix of this system can be written according to the definition given in Chapter 2:

$$N = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.3.54}$$

By using (2.1.18), all differential equations can be denoted in the vector form:

$$\frac{dx}{dt} = \begin{bmatrix} -v_1 \\ -v_1 + v_2 \\ v_2 - v_3 \\ v_1 - v_2 \\ -v_2 \\ v_1 \\ -v_3 \\ v_3 \\ v_3 \end{bmatrix}. \tag{2.3.55}$$

Conservation Relations

Looking for the rank of matrix N in Matlab 6.5 by using the command *rank(N)*, it gives us the value 3. Clearly, when N is subdivided into matrices as discussed by "'Conservation Relations"' in Chapter 2, we will have 6 conservation relations since N is a (9×3) matrix and $9 - rank(N) = 6$. Thus, we can divide N in

the following form:

$$N_{R1} = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \tag{2.3.56}$$

$$N_{R2} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.3.57}$$

It is obvious that $N_{R1} = N_{R11}$, since the rank of N is equal to the number of columns. Remember that N_{R11} consists of the linearly independent columns of N_{R1}. Obviously, as a conseqence of the previous observation, $N_{R2} = N_{R21}$. Thus, L_0 can be computed by equation (2.2.51):

$$L_0 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & -1 \end{bmatrix}.$$

In Subsection 2.2.3, the vector x is divided into two parts as x_{R1} and x_{R2}, where the dimension of x_{R1} is the rank of N (maximal number of linearly independent columns of N) and x_{R2} is the remaining vector in x. Thus, we can write x according to our model by two parts:

$$x_{R1} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{2.3.58}$$

and

$$x_{R2} = \begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} . \tag{2.3.59}$$

By Theorem 2.2.1, it is clear that conservation relations can be stated in the following form:

$$\frac{d}{dt}(x_{R2} - L_0 x_{R1}) = 0 \Longrightarrow \frac{d}{dt}\left(\begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} - \begin{bmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = 0$$

$\Longrightarrow$

$$\frac{d}{dt} \begin{bmatrix} x_4 + x_2 \\ x_5 - x_1 + x_2 \\ x_6 + x_1 \\ x_7 - x_1 + x_2 - x_3 \\ x_8 + x_1 - x_2 + x_3 \\ x_9 + x_1 - x_2 + x_3 \end{bmatrix} = 0 . \tag{2.3.60}$$

Consequently, the function which we differentiated must be a constant and, to be biologically meaningful, these constants must be the initial concentrations

of these metabolites because of the law of mass conservation. Thus, we have:

$$\begin{bmatrix} x_4 + x_2 \\ x_5 - x_1 + x_2 \\ x_6 + x_1 \\ x_7 - x_1 + x_2 - x_3 \\ x_8 + x_1 - x_2 + x_3 \\ x_9 + x_1 - x_2 + x_3 \end{bmatrix} = \begin{bmatrix} x_{4_0} + x_{2_0} \\ x_{5_0} - x_{1_0} + x_{2_0} \\ x_{6_0} + x_{1_0} \\ x_{7_0} - x_{1_0} + x_{2_0} - x_{3_0} \\ x_{8_0} + x_{1_0} - x_{2_0} + x_{3_0} \\ x_{9_0} + x_{1_0} - x_{2_0} + x_{3_0} \end{bmatrix} \tag{2.3.61}$$

$\Longrightarrow$

$$\begin{bmatrix} x_4 + x_2 - x_{4_0} - x_{2_0} \\ x_5 - x_1 + x_2 - x_{5_0} + x_{1_0} - x_{2_0} \\ x_6 + x_1 - x_{6_0} - x_{1_0} \\ x_7 - x_1 + x_2 - x_3 - x_{7_0} + x_{1_0} - x_{2_0} + x_{3_0} \\ x_8 + x_1 - x_2 + x_3 - x_{8_0} - x_{1_0} + x_{2_0} - x_{3_0} \\ x_9 + x_1 - x_2 + x_3 - x_{9_0} - x_{1_0} + x_{2_0} - x_{3_0} \end{bmatrix} = 0 \ . \tag{2.3.62}$$

Thus, our model consists of three differential equations given in (2.3.63) with constraints (2.3.62):

$$\begin{aligned} \frac{dx_1}{dt} &= v_1 \ , \\ \frac{dx_2}{dt} &= v_2 \ , \\ \frac{dx_3}{dt} &= v_3 \ , \end{aligned} \tag{2.3.63}$$

where v_1, v_2, v_3 are the rate equations of the CK, HK and G6PDH system, respectively, given in Subsections 2.3.1, 2.3.2 and 2.3.3, respectively.

Parameters Used in the Mathematical Model

It is known that rate equations depend on both metabolites concentrations and parameters. In the CK-HK-G6PDH model, we have 3 differential equations with 6 linear constraints and 26 parameters such as K_M, K_I and V_{max} values.

	Parameters of CK System	
K_M^{ADP}	$5,00x10^{-2}mM$	Literature
K_I^{ADP}	$1,70x10^{-1}mM$	Literature
K_M^{ATP}	$4,80x10^{-1}mM$	Literature
K_I^{ATP}	$1,20x10^{0}mM$	Literature
K_M^{Cr}	$6,10x10^{0}mM$	Literature
K_I^{Cr}	$1,56x10^{1}mM$	Literature
K_M^{CrP}	$2,90x10^{0}mM$	Literature
K_I^{CrP}	$8,60x10^{0}mM$	Literature
V_{cat}^{f}	$200min^{-1}$	Literature
V_{cat}^{r}	$100min^{-1}$	Literature

Table 2.1: Parameters of CK system (Morrison and James, 1965).

	Parameters of HK System	
K_M^{ATP}	$6,30x10^{-2}mM$	Literature
K_I^{ATP}	$6,30x10^{-2}mM$	Literature
K_M^{ADP}	$2,30x10^{-1}mM$	Literature
K_I^{ADP}	$2,30x10^{-1}mM$	Literature
K_M^{DGlu6P}	$4,00x10^{-2}mM$	Literature
K_I^{DGlu6P}	$6,70x10^{0}mM$	Literature
K_M^{Glu}	$1,00x10^{-1}mM$	Literature
K_I^{Glu}	$1,00x10^{-1}mM$	Literature
V_{cat}^{f}	$58.82min^{-1}$	Literature
V_{cat}^{r}	$11764.7min^{-1}$	Literature

Table 2.2: Parameters of HK system (Viola et al., 1982).

The initial values of metabolites and all parameters are taken from [38].

	Parameters of G6PDH System	
$K_M^{NADP^+}$	$6,10x10^{-3}mM$	Literature
$K_I^{NADP^+}$	$6,20x10^{-3}mM$	Literature
K_M^{DGlu6P}	$3,90x10^{-2}mM$	Literature
K_I^{DGlu6P}	$8,90x10^{-1}mM$	Literature
K_M^{6PGL}	$1,38x10^{-1}mM$	Literature
K_I^{6PGL}	$6,90x10^{-3}mM$	Literature
K_M^{NADPH}	$3,90x10^{-3}mM$	Literature
K_I^{NADPH}	$6,80x10^{-3}mM$	Literature
V_{cat}^f	$2836.87min^{-1}$	Literature
V_{cat}^r	$42553.19min^{-1}$	Literature

Table 2.3: Parameters of G6PDH system (Gordon et al., 1995).

The initial concentrations of the metabolites and enzyme concentrations are

$$\begin{aligned} x_{1_0} &= 2,55x10^1mM, \\ x_{2_0} &= 5,9x10^{-1}mM, \\ x_{3_0} &= 0mM, \\ x_{4_0} &= 0mM, \\ x_{5_0} &= 1,13x10^1mM, \\ x_{6_0} &= 0mM, \\ x_{7_0} &= 3,3x10^{-1}mM, \\ x_{8_0} &= 0mM, \\ x_{9_0} &= 0mM, \\ CK_0 &= 8x10^{-6}mM, \\ HK_0 &= 3,4x10^{-5}mM, \\ G6PDH_0 &= 1,4x10^{-5}mM. \end{aligned} \tag{2.3.64}$$

Chapter 3

Numerical Solution Methods For Solving Differential Algebraic Equations

The model obtained in the previous chapter consists of a system of differential equations with algebraic constraints, i.e., a system of differential algebraic equations (DAE's). In this chapter, we will summarize the existing methods for solving DAE's and apply the MATLAB's DAE solver **ode15s** to our model.

3.1 Introduction

Differential algebraic equations are much more recent than ordinary differential equations. In the recent years, many enginneering, bioinformatics and medical problems are modelled by DAE's. However, difficulties arise, especially, when nonlinear algebraic constants and a lot of parameters exist. Modern application of DAE's can be found in the field of exercise metabolism where they are studied, e.g., in *sports medicine*. Here, we refer to the investigations [26, 27]. In these studies, a DAE model is reduced to the form of ODE's in order to simplify the equations. In the last years, several numerical methods are devised for solving DAE's, see, for example, the monographs [10, 24].

The differences and similarities between ODE's and DAE's can be explained as follows:
Consider $y(t)$ and $z(t)$ be the two functions defined on some interval $[0, b]$ and related by the differential equation

$$y^{'}(t) = z(t), \quad 0 \leq t \leq b \ . \tag{3.1.1}$$

To obtain $y(t)$ from $z(t)$, an integration with additional knowledge of $y(0)$, over the interval $a \leq t \leq b$ is needed; and to get $z(t)$ from $y(t)$, only $y(t)$ should be differentiated. The differentiation is an easier process than integration. Indeed, $y(t)$ is usually much more smooth than $z(t)$. For example, $z(t)$ may be bounded but have jump discontinuities. On the one hand, integration is a *smoothing*, process. On the other hand, differentiation is an *anti-smoothing*, *a roughening* process. Solution of ODE's involves integration, thus, it is smoothing, but solution of DAE's involves both integration and differentiation.

Definition 3.1.1. A system of DAE's is called ***fully implicit*** if it is of the form

$$F(y, \dot{y}, t) = 0 \text{ with } y(0) = y_0 \ . \tag{3.1.2}$$

Definition 3.1.2. A system of DAE's is called ***linearly implicit*** if it of in the form

$$A\dot{y} + f(y, t) = 0 \text{ with } y(0) = y_0 \ , \tag{3.1.3}$$

with a singular matrix A.

Definition 3.1.3. A system of DAE's is called ***semi explicit*** if it is of the form

$$\dot{x} = f(x, z, t), \tag{3.1.4}$$

$$g(x, z, t) = 0. \tag{3.1.5}$$

It is also assumed that $\frac{\partial g}{\partial z}$ has a bounded inverse in a neighbourhood of the solution.

In our model CK-HK-G6PDH, equation (2.1.18), the differential part, cor-

responds to (3.1.4), and equation (2.3.62) is related to the algebraic part (3.1.5). Here, the vector of differential variables is denoted by $x = (x_1, x_2, x_3)^T$ and the vector of algebraic variables is called $z = (x_4, x_5, x_6, x_7, x_8, x_9)^T$. Then, we can write our model as a semi-explicit system of DAE's in the form

$$M\dot{y} = f(t, y, p), \tag{3.1.6}$$

where $y = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9)^T$, and p is the vector of parameter. Furthermore, a singular matrix M is given by

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3.1.7}$$

and $f(t, y, p)$ is of the following form:

$$f(t, y, p) = \begin{bmatrix} v_1(t, x, p) \\ v_2(t, x, p) \\ v_3(t, x, p) \\ x_4 + x_2 - x_{4_0} - x_{2_0} \\ x_5 - x_1 + x_2 - x_{5_0} + x_{1_0} - x_{2_0} \\ x_6 + x_1 - x_{6_0} - x_{1_0} \\ x_7 - x_1 + x_2 - x_3 - x_{7_0} + x_{1_0} - x_{2_0} + x_{3_0} \\ x_8 + x_1 - x_2 + x_3 - x_{8_0} - x_{1_0} + x_{2_0} - x_{3_0} \\ x_9 + x_1 - x_2 + x_3 - x_{9_0} - x_{1_0} + x_{2_0} - x_{3_0} \end{bmatrix}. \tag{3.1.8}$$

Here, $x_{1_0}, ..., x_{9_0}$ denote the initial values of differential and algebraic variables.

3.2 Differential Index

The differential index of a DAE is defined to be the number of differentiations needed to transform the DAE into an explicit ODE. For the *semi-explicit* case, the index is the minimum number of differentiation of (3.1.5) with respect to time t, to obtain the following system of ODE's:

$$\dot{x} = f(x, z, t) \ , \tag{3.2.9}$$

$$\dot{z} = s(x, z, t) \ . \tag{3.2.10}$$

Definition 3.2.1. Let f be the once continuously differentiable function. DAE (3.1.2) has the differential ***index one*** if and only if a system of variables t, x, $\dot{x}$, $\ddot{x}$ uniquely determines the variable $\dot{x}$ as a function of (x, t):

$$\begin{aligned} f(x, \dot{x}, t) &= 0 \ , \\ \frac{d}{dt} f(x, \dot{x}, t) &= \frac{\partial}{\partial x} f(x, \dot{x}, t)\ddot{x} + \frac{\partial}{\partial \dot{x}} f(x, \dot{x}, t)\dot{x} + \frac{\partial}{\partial t} f(x, \dot{x}, t) = 0 \ . \end{aligned}$$

Example 3.2.2. Consider the following DAE problem

$$\begin{aligned} \dot{x} &= 3x + 4y \ , \\ y &= 5x + 3 \ . \end{aligned}$$

Taking the derivative of second equation and substituting $\dot{x}$ from the first equation in the second one, we obtain a system of pure ODE's.
Because the number of differentiations to obtain an ODE was one, the ODE is of index one, too.

Definition 3.2.3. Let f be twice continuously differentiable. The DAE (3.1.2) has differential ***index two*** if and only if it is not of index one and the equation system of the variables t, x, $\dot{x}$, $\ddot{x}$, $\dddot{x}$ uniquely determines the variable $\dot{x}$ as a

function of (x,t):

$$\begin{aligned} f(x,\dot{x},t) &= 0, \\ \frac{d}{dt}f(x,\dot{x},t) &= \frac{\partial}{\partial x}f(x,\dot{x},t)\ddot{x} + \frac{\partial}{\partial \dot{x}}f(x,\dot{x},t)\dot{x} + \frac{\partial}{\partial t}f(x,\dot{x},t) = 0, \\ \frac{d^2}{dt^2}f(x,\dot{x},t) &= \frac{\partial}{\partial x}f(x,\dot{x},t)\dddot{x} + \dots \ . \end{aligned}$$

Example 3.2.4. Consider the system of DAE

$$\begin{aligned} \dot{x_2} &= x_1 + z_1(t), \\ x_2 + z_2(t) &= 0. \end{aligned}$$

Differentiating the second equation and substituting it into the first one gives:

$$\dot{x}_2 = -\dot{z}_2(t) = x_1 + z_1(t) \Longrightarrow x_1 = -z_1(t) - \dot{z}_2(t) \ .$$

Again differentiaing we get an ODE:

$$\dot{x}_1 = -\dot{z}_1(t) - \ddot{z}_2(t) \ .$$

Thus, our system has *index two*.

DAE's with higher index are much more complex, because the number of differentiations needed in order to reduce this system to an ODE increases. Therfore, the numerical methods are devised for DAE's up to index three. Higher order index problems should be reduced to lower index ones. This procedure is called ***index reduction***.

3.3 Numerical Solution Methods to Solve DAE's

Numerical methods for DAE's are based either on backward differentiation methods or on Runge-Kutta methods. For an extensive study of these methods see the monographs [10] and [24].

3.3.1 Backward Euler

Consider the fully implicit DAE

$$F(t, y, y') = 0 \ .$$

The idea here is to approximate y and y' by a discretization formula. Discretize the state variable y into n equal pieces, i.e., y_n. We approximate y' by

$$\frac{y_n - y_{n-1}}{h_n} \ ,$$

where h_n is the step size. So, our DAE turns into

$$F(t_n, y_n, \frac{y_n - y_{n-1}}{h_n}) = 0. \tag{3.3.11}$$

The system (3.3.11) consists of m nonlinear equations for y_n at each time step n [24].

3.3.2 Runge-Kutta Method

In general, the s-stage Runge-Kutta method applied to *fully implicit DAE* is defined by (cf. [24]):

$$\begin{aligned} 0 &= F(t_i, Y_i, K_i), \\ t_i &= t_{n-1} + c_i h, \qquad i = 1, 2, ..., s, \\ Y_i &= y_{n-1} + h\sum_{j=1}^{s} a_{ij} K_j, \end{aligned} \tag{3.3.12}$$

and

$$y_n = y_{n-1} + h\sum_{i=1}^{s} b_i K_i. \tag{3.3.13}$$

We assume here that the coefficient matrix $A = (a_{ij})$ is nonsingular and the method is specified by this matrix A and the vector b having the entries b_i. For the case of a *semi-explicit DAE* formulated in (3.1.4) and (3.1.5), the formula

(3.3.12) is converted in [24]:

$$\begin{aligned} K_i &= f(t_i, X_i, Z_i), \\ X_i &= x_{n-1} + h\sum_{j=1}^{s} a_{ij} K_j, \\ 0 &= g(t_i, X_i, Z_i). \end{aligned}$$

3.3.3 Backward Differentiation Formula

Backward differentiation formula (BDF) is a *multistep method* that uses information from previous steps to construct higher order approximations. Compared to the Runge-Kutta method, multistep methods require fewer function evaluations per step. *Backward Differentiation Formula (BDF)* is the most popular multistep method. The difference from the Runge-Kutta method is that $f(t, y)$ is evaluated only at the right end of the current step, (t_n, y_n).
The general BDF method for solving ODE's is given in the form of

$$\sum_{i=0}^{k} \alpha_i y_{n-1} = h\beta_0 f(t_n, y_n), \tag{3.3.14}$$

where h stands for the step-size, α_i and β_0 are coefficients. Here, k is the order of the method and denotes how many backward steps are used. The method is an implicit one if $\beta_0 \neq 0$. If it is implicit, then it is necessary to calculate the function value at (t_n, y_n) which requires the treatment of a nonlinear system of equations solved usually by a *modified Newton method*.

Application of BDF method to DAE (3.1.2) gives

$$F\left(t_n, y_n, \frac{1}{\beta_0 h}\sum_{j=0}^{k} \alpha_j y_{n-j}\right) = 0. \tag{3.3.15}$$

For $\alpha_0 = \beta_0 = 1$ this is simply *Euler's method*.

MATLAB's DAE solver ode15s

BDF methods are implemented in MATLAB by the function **ode15s** which works for DAE index up to 2. The important advantage of this case is finding consistent initial values for the DAE system. The user supplies initial values for the system and the code evaluates the initial values again, and finds appropriate and consistent initial conditions. Because of these properties ***ode15s*** is chosen to simulate our model problem in this study. The MATLAB DAE solver *ode15s* can be applied to any semi-explicit system (3.1.6) with a singular matrix M. Matlab ***ode15s*** automatically recognizes that the problem is a DAE and computes consistent initial conditions close to given y_0. Some parameters used in *ode15s* are given in the following:

$[t, Y]$ = ode15s(odefun,tspan,y0), or $[t, Y]$= ode15s(odefun,tspan,y0,options)

odefun: A function that evaluates the right-hand side of the differential equation of the form $y' = f(t, y)$ or $My' = f(t, y)$.

tspan: This is the time interval of integration $[t_0 \;\; t_f]$. The solver takes initial condition at t_0 and integrates from t_0 to t_f.

y_0: denotes the vector of initial conditions.

options: The set of optional parameters that change the default integration properties like error tolerances. Options can be changed using **odeset** of MATLAB [18].

Chapter 4

Parameter Estimation

In most of the real world problems such as considered and tested in engineering, computational biology and medicine, or in bioinformatics, mathematical models, contain a number of parameters. Usually, these parameters are not known exactly. Changing some parameter may affect the solutions strongly, which means that the model is sensitive. Sometimes, it may be hard to evaluate the all parameters by an experiment due to its difficulty of measuring the values. In such cases, several parameter estimation methods like *Simplex method, Levenberg-Marquardt (LM) method*, or *quasi-Newton method* are used. All of them are a kind of gradient-based method, or mixture between gradient and Newton's method. From these algortihms, LM is the most popular gradient-based method that converges quadratically to the *local minimum*.

Apart from these methods, there are some parameter estimation algorithms that do not require specific functions to be optimized or evaluated. These are *genetic algorithms* and *derivative free algorithms*. In many cases, the function which represents the difference between the simulation and experimental results is not given in explicit form. For a closer information about such *model free algorithms*, we can refer to statistical learning, e.g., to the valuable book [29]. In bioinformatics studies, one of the most popular optimization method is genetic algorithm ***(GA)*** for reducing the error with respect to model parameters. We refer to [19] for the heuristic modelling and optimization algorithms like GA. In

this work, our nonlinear continuous parameter estimation problem is:

$$\min_k ||x_{simulation}(k) - x_{experimental}(k)||_2 \ , \quad k_l \leq k \leq k_u. \tag{4.0.1}$$

Here, $x_{simulation}$, $x_{experimental}$ are the vectors of simulation and experimental results, respectively. Furthermore, k denotes the parameter, k_l and k_u are the upper and lower bound for parameter k, component-wise regarded, and $||.||_2$ denotes Euclidean norm.

4.1 Genetic Algorithm

Genetic algorithms were firstly introduced by Holland in 1975 [2]. They are based on *fittest* and *natural selection* of the given population or mechanism which are simulated by these principles. GA is a random search algorithm which is good for nonlinear problems and its efficiency relies on the fact that it does not require to fix the initial value for the parameter estimation [20]. It is widely applied in chemical engineering, biology and and biochemistry, in short: in *biotechnology*. Contrary to LM, it finds the *global* minimum of the error function.

In a classical GA, the set of candidate solutions are created randomly; because of this property, it is known as a *stochastic algorithm*. These randomly created solutions are saved in a pool and all individual ones are in the vector form called *chromosomes*. After a fitness criterion is applied to the solution space, the new generation is constructed from actual population according to *reproduction, mutation* and *crossover* [21].

4.1.1 Coding

One of the fundamental issue of GA is *coding* the solution space. Typically, each individual solution is coded in strings, sometimes called *chromosomes*. One of the usual form of the coding is *binary coding* which consists of the elements

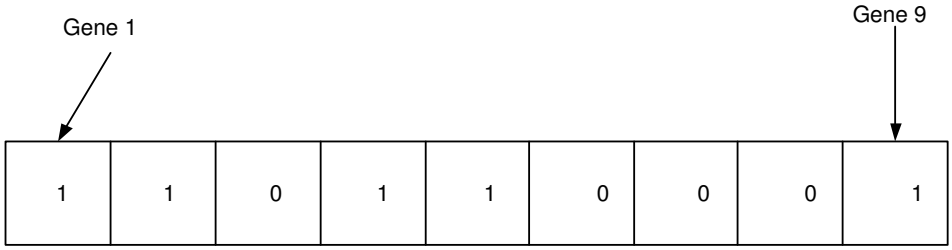


Figure 4.1: Illustration of a chromosome where each box represents one gene of chromosome.

of the genes. Binary values are given to each gene indicating, for example, genes' functions in the metabolic pathways. In Figure 4.1, we have illustrated the coding of a chromosome.

4.1.2 Reproduction

This operator selects the pairs of chromosomes to produce *offsprings* which are the children of mated pairs. *Reproduction* occurs randomly where the pairs that obey the fitness criteria higher than the others, have more chance to be selected [21]. Each chromosome is ordered with respect to its fitness in the population where the one with minimum error is going to be the first candidate to be chosen. If this individual fulfills the given conditions, e.g., a tolerance value, then the algorithm chooses this vector of genes. If not, then GA follows the mutation or crossover operators. We illustrate the randomly generated pairs of chromosomes in a pool shown Figure 4.2, where each string represents one chromosome.

4.1.3 Crossover

The *crossover* operator redistributes the genetic characteristics between the individuals [21]. Modification occurs within the two pairs of strings which are cut off from randomly chosen points. Then, these cutted pairs *alternate* each other. These alterations can be done in two ways:

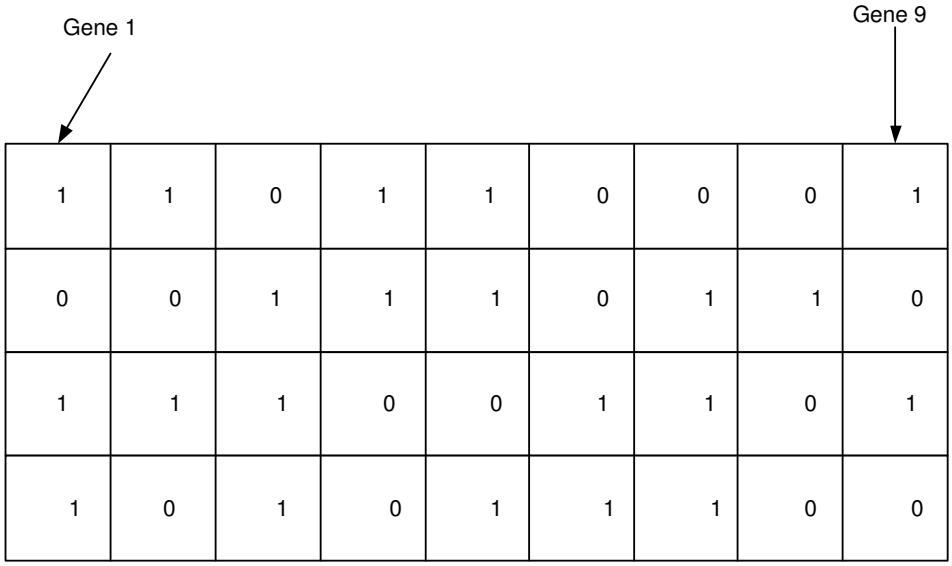


Figure 4.2: A binary-valued, randomly generated pool.

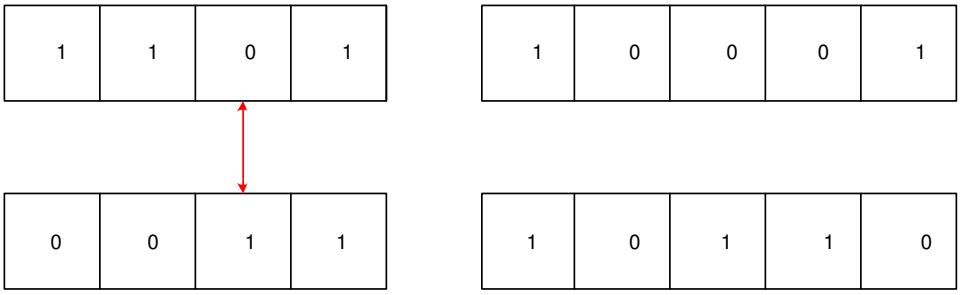


Figure 4.3: A crossover 1.

1. We can replace the first part of the first string by the first part of the second, or the first parts may remain the same but the tails can be replaced by each other. We represent this in Figure 4.3.

2. We can *crossover* the selected parts of the cut strings. We have illustrated this type in Figure 4.4.

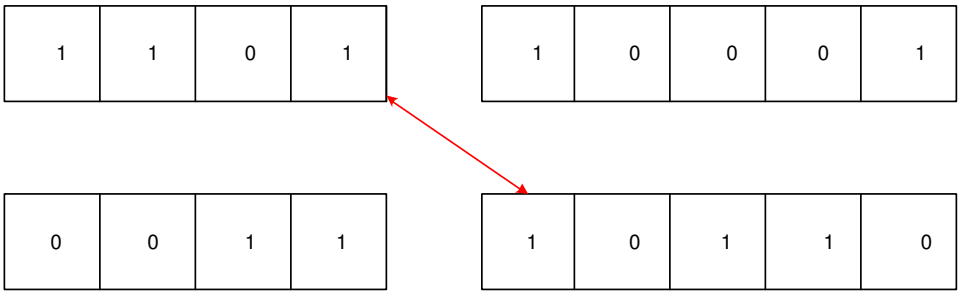


Figure 4.4: A crossover 2.

4.1.4 Mutation

In nature, some genes of offsprings may have totally different characteristics from their parents because of various factors [21]. Mutation is a process to simulate this natural phenomenon. It selects a random gene in a string and changes its value again randomly. Like in case of the other operators, after chromosomes are mutated, they are going to be tested whether they fit to solution or not, and the algorithm continues until the stopping criterion holds.

The following algorithm summarizes the whole steps in a GA.

1. Construct your population by randomly generated vectors of solution candidates (*reproduction*).
2. Test if your candidates in the pool fit your model.
3. Order them according to their fitness by putting solution with minimum error as a first candidate.
4. Apply *mutation* and *crossover*.
5. Test a newly generated *pool* in step 4, whether it fits or not.
6. Go to step 3.
7. If there is a solution which satisfies your stopping criterion: STOP; if not, go to step 4.

4.2 GA Adapted to CK-HK-G6PDH Model

In our parameter estimation problem, we coded the genes differently from the previous section, because our parameter values changes on some intervals $[a, b]$. In Section 4.1, we have discussed that genes are coded by binary values as opposed to their functionality in pathways. However, here, we are looking for the parameter values in the kinetic model of CK-HK-G6PDH system. These parameters are v_{max} and the K_M, K_I values of the chemical reactions in the model pathway. We totally have 26 parameters and 9 unknowns for the metabolite concentrations in the model. So, our chromosomes consist of 26 genes, i.e., they are represented by a vector of dimension 26. Now, GA is adapted to our model in the following form:

Reproduction

We generate our population pool randomly with the size of a 200×26 matrix by the Matlab command *rand(200,26)*.

Also, we specify the *number of iterations* to put some limit on the algorithm. After constructing candidates of solutions, the next step is to assess them according to their fitness to model, by looking at the error function:

$$Error = ||x_{simulation}(k) - x_{experiment}(k)||_2. \tag{4.2.2}$$

Here, $x_{simulation}$ is the vector after simulation with parameter $k = (v_{max}, K_M, K_I)$ within the randomly generated pool, and $x_{experiment}$ is the experimental value of a metabolite over the time. After finding all error vectors, the errors are ordered in ascending form. If the first chromosome satisfies our stopping criterion, e.g., $Error < Tolerance$, then the algorithm stops. If not, it will continue by the mutation and crossover operations.

In this thesis, we could correct the *prior* simulated NADPH values by a GA, since we have only the empirical results of NADPH.

Mutation

The chromosomes and genes that mutate them are chosen randomly. Here, we

should be careful by giving random values to genes in chromosomes since we have constraints in our parameters. We must restrict our interval of randomization process. Another restriction in this step is that mutation is not repeated in every iteration, but every second iteration, i.e., $mod(2)$. This criterion is chosen to prevent big errors in parameter values. In fact, we may be far from the exact solution.

Crossover

In this step, we have randomly chosen the cut off point of every pair of individuals and make the crossover operation to the whole pool. As indicated in Subsection 4.1.3, there are two kinds of crossovering. We have chosen the first type, i.e., interchanging the first parts of the cutted strings. The reason for this choice consists in the fact that big changes in structure of chromosomes may produce large errors.

After mutation and crossover, we must look at the fitness of the new candidates modified by GA. As we have mentioned before, we should evaluate the error functions, order them and look if the first string, i.e., first row in the pool satisfies our error tolerance. The algorithm continues up to the maximum iteration.

Chapter 5

Results and Discussion

In this section, the results of simulation of our DAE system given in Chapter 3 will be discussed. The numerical results obtained using the MATLAB DAE solver *ode15s* are compared with the experimental and simulation results existing in the literature. After parameter estimation, the model is simulated again and the solutions are improved.

5.1 Results of Simulation

In Figures 5.1, 5.3, 5.5, our simulation results are illustrated and in Figure 5.2, 5.4, 5.6, the results of [38] are given, where the reaction rates were assumed to be equal: $v_1 = v_2 = v_3$. In Subsections 2.3.1, 2.3.2, 2.3.3, nonlinear rate equations are approximated using Gröbner basis in MAPLE, and the resulting equations are solved numerically in FORTRAN. In fact, we have used different values for reaction rates v_1, v_2, v_3 and solved our model as a DAE system. Our results differ from those given in [38]. However, our simulation results are qualitatively correct and give the tendency of the concentrations over time. Since only the concentration values of NADPH over time are measured in Laboratory of Biotechnology Applied and Research Center in Atatürk University in Erzurum [38], our simulation results are compared with experimental values in Figure 5.1. A bit later, in Figure 5.8, we have the simulation results from [38].

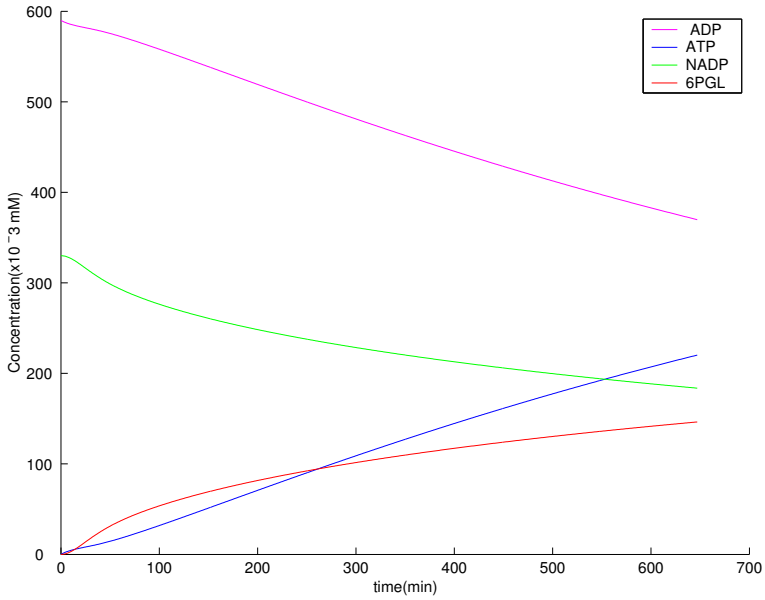


Figure 5.1: Simulation results of ATP, ADP, NADP, 6PGL.

The nonlinearity of the rate equations may cause the roundoff errors during numerical solutions with MATLAB. Furthermore, MATLAB works with sixteen digits, so there may be truncation errors during the numerical computations. Also the assumptions can cause some differences with the values in literature, since the steady state assumption is ignored. It turns out that our results are different from the ones in [38] before parameter estimation. After parameter estimation, however, our computed results fit very well to the experimental data, as it will be shown in Figure 5.9.

In the previous chapter, we have introduced GA's which are widely used in parameter estimation problems. We have estimated our 26 parameters by using experimental data of NADPH and GA. By using estimated parameters, we have simulated our model again to get better and more accurate results. Our objective function during estimation was

$$\min ||x_{9,resimulated}(k) - x_{9,experimental}(k)||_2. \tag{5.1.1}$$

Here, $x_{9,resimulated}(k)$ is the vector of NADPH concentration after GA, $x_{9,experimental}(k)$

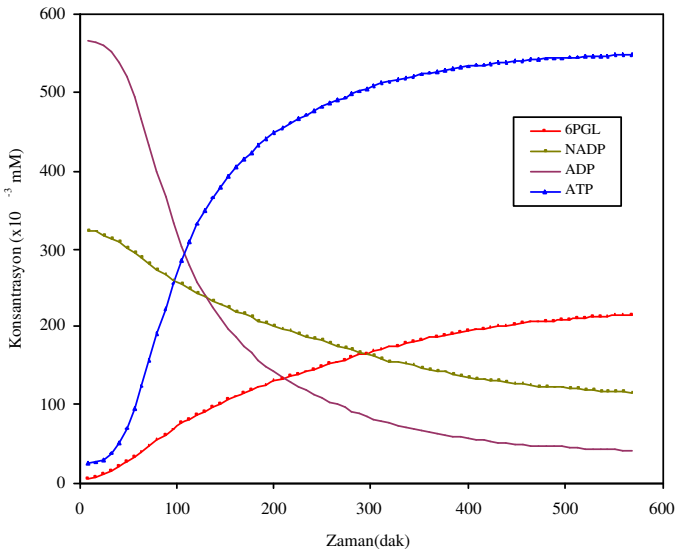


Figure 5.2: Simulation results of ATP, ADP, NADP, 6PGL [38] (zaman: time, dakika: minute).

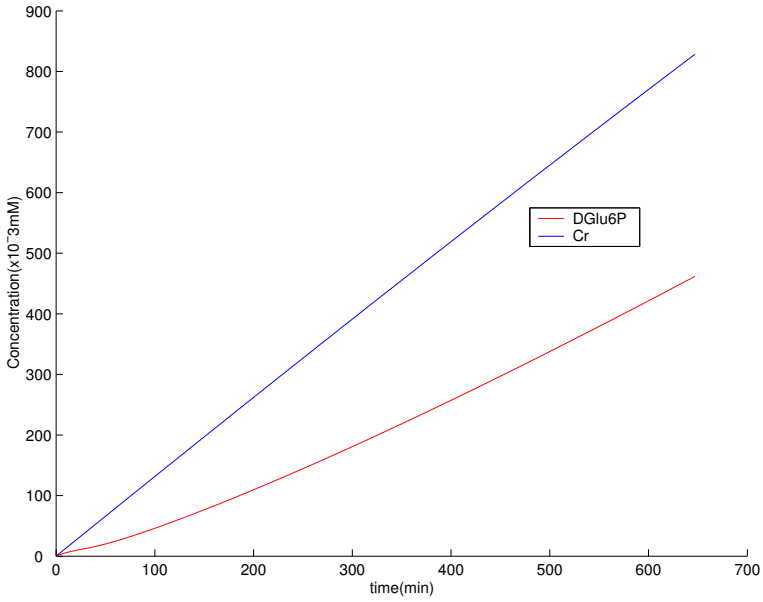


Figure 5.3: Simulation results of DGlu6P, Cr.

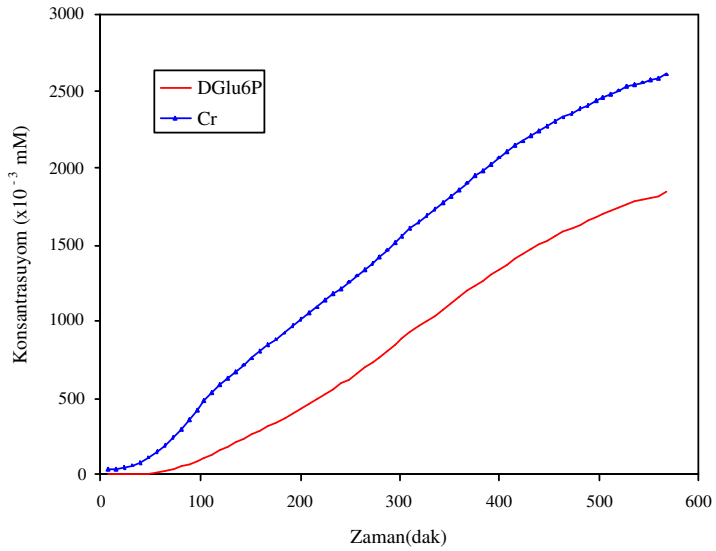


Figure 5.4: Simulation results of DGlu6P, Cr [38] (zaman: time, dakika: minute).

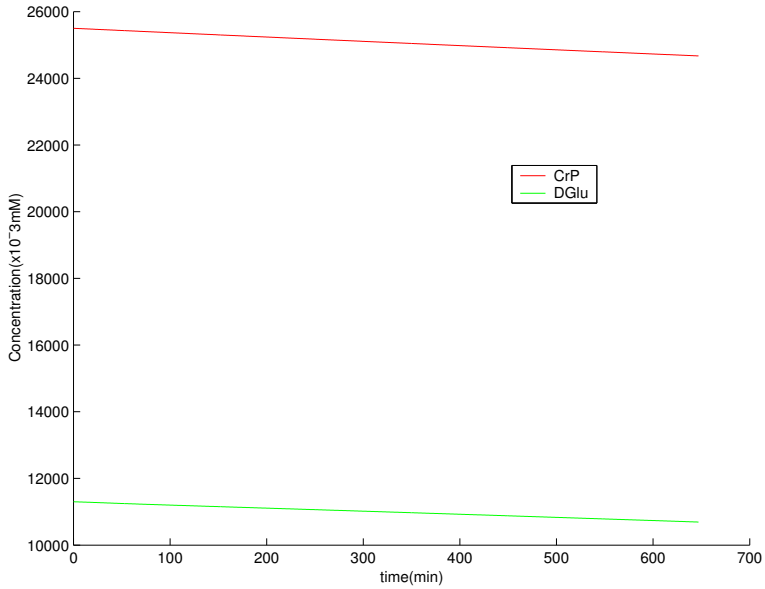


Figure 5.5: Simulation results of CrP, DGlu.

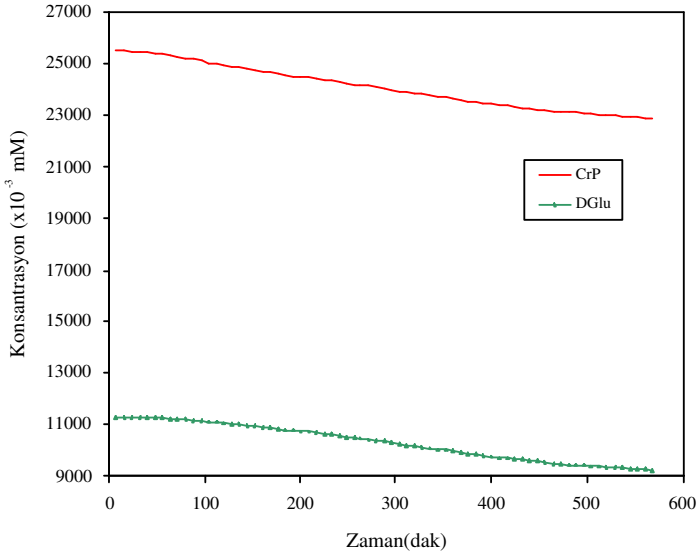


Figure 5.6: Simulation results of CrP, DGlu [38] (zaman: time, dakika: minute).

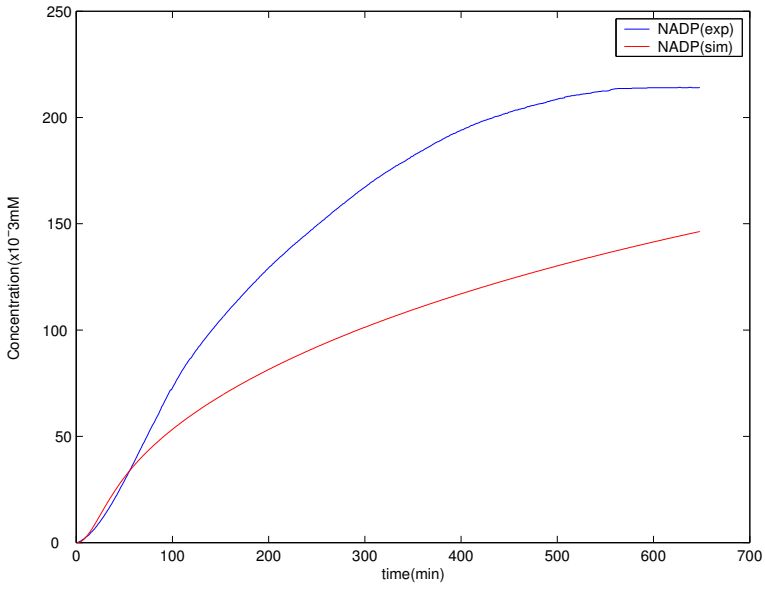


Figure 5.7: Experimental values of NADPH and simulation results of NADPH by *ode15s*.

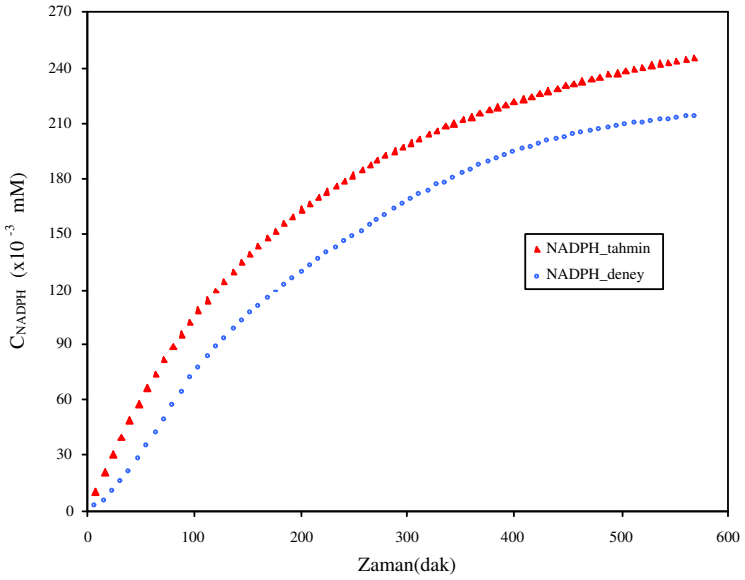


Figure 5.8: Experimental values of NADPH and simulation results of NADPH [38] (zaman: time, dakika: minute).

consists of the experimental values of NADPH over the time interval of $[0,\ 647]$, and k is the vector of parameters.

The NADPH concentration after parameter estimation by GA is given in Figure 5.9. It is obvious that our results *better* fits the experimental values.

5.2 Conclusion

We have modelled an enzymatic reaction of a system CK-HK-G6PDH as a system of differential algebraic equations. The algebraic parts consisted of linear constraints. The numerical results obtained using MATLAB's ODE solver are comparable with the results given in the literature. After applying parameter estimation using a genetic algorithm, the new results fit very well the experimental results.

An important point of this work consists in the application of *DAE* approach for the analysis of metabolic pathways. In literature, only in few cases the

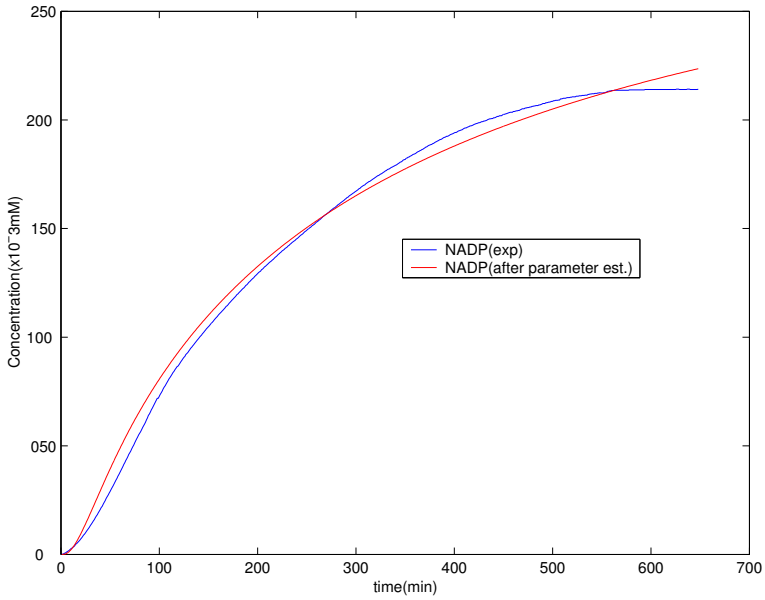


Figure 5.9: NADPH results after parameter estimation and experimental values of NADPH.

applicability of DAE's to metabolic engineering are mentioned [26, 27, 38]. In these researches, DAE's are applied to human exercise metabolism, especially, in sports medicine where the concentration of lactate and other metabolites are measured for the prediction of future values of these reactants. Since lactate plays an important role in exercise physiology, the simulation of this product is analyzed by the model form of DAE's, and to reduce the system of equations it is simplified to a system of ODE's. We mention that in sports medicine one aim consists in the scientific development and design of training programs. The insights obtained here, e.g., by metabolic engineering which we are dealing with in this thesis, should also serve for *general medicine*, for instance, for patients in recovering.

Indeed, we have modelled our system by means of DAE's and the experimental data are in vitro, i.e, not in living cell. Thus, in future work, this thesis can hopefully be applied to large systems or human metabolism, like those given in [26, 27], or in further and more general contexts. In this work, we have considered linear constraints only, but the application of a DAE solver for the analysis

of metabolic pathways with *nonlinear* constraints will be useful, too. Another important aspect and subject of this thesis's attention is *sensitivity analysis* of parameters. The sensitivity analysis for DAE methods [5, 6, 17] can be applied to metabolic pathways in future.

Moreover, we have emphasized the role of *optimization* theory, e.g., by using the methods of *genetic algorithms* which we approached and used from the viewpoints of statistical learning and heuristics. In fact, our entire fruitful approach should be applied to large systems of those metabolic pathways. GA's can be scientifically located in *model free* approach, since the parameters are estimated by empirical data and stochastic random numbers. In contrast to classical optimization methods, GA's are applied by us since the objective function is not given explicitly. In fact, this function consists of a random and experimental, i.e., of a discrete set of variables. Further progress in optimization theory and statistical learning is expected by us, and it will be utilized in metabolic engineering for many important fields of medicine, health care and other areas of modern life.

References

[1] Aleshin, Alexander E., Zeng, Chenbo, Bourenkov Gleb P., Bartunik, Hans D., Fromm, Herbert J., and Honzatko, Richard B , *The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose and glucose-6-phosphate*, Structure, 6, pp. 39–50, 1998.

[2] Altunkaynak, Bülent, and Esin, Alptekin, *The genetic algorithm method for parameter estimation in non linear regression*, Gazi Üniversitesi Journal of Science 17, 43-51, 2004.

[3] Bendtsen, Claus, and Thomsen, Per Grove, *Numerical solution of differential algebraic equations*, TECHNICAL REPORT IMM-REP-1999-8.

[4] Biegler, L.T., *Differential Algebraic Equations (DAEs)*, http://dynopt.cheme.cmu.edu, May 11, 2000.

[5] Cao, Y., Shengtai Li, Petzold, Linda, *Adjoint sensitivity analysis for differential-algebraic equations: algorithms and software*, Journal of Computational and Applied Mathematics 149, 171-191, 2002.

[6] Cao, Yang, Li, S., Petzold, L., and Serban, R., *Adjoint sensitivity analysis for differential-algebraic equations: The adjoint system and its numerical solution*, SIAM J. SCI. COMPUT. Society for Industrial and Applied Mathematics Vol. 24, No. 3, pp. 1076-1089, 2003.

[7] Cornelio, Andrea R., Rodrigues, Valnés, Jr, Terezinha de Souza, Wyse, Angela, Dutra-Filho, Carlos Severo, Wajner, M., and Wannmacher, Clo-

vis Milton Duval, *Tryptophan reduces creatine kinase activity in the brain cortex of rats,* Int. J. Development Neuroscience 22, pp.95–101, 2004.

[8] Fersht, Alan, *Enzyme Structure Mechanism.* Second Edition, W.H. Freeman and Company, New York, 1985.

[9] Giersch, Christoph, *Mathematical Modelling of Metabolism*, Current Opinion in Plant Biology 3, pp. 249-253, 2000.

[10] Hairer, E., Lubich, C., and Roche, M., *The numerical solution of differential-algebraic systems by Runga-Kutta methods*, Berlin; New York, Springer-Verlag, 1989.

[11] Hatzimanikatis, Vassily, *Nonlinear metabolic control analysis*, Metabolic Engineering 1, pp.75-87, 1999.

[12] Heyworth, Anne *Grbner basis theory*, University of Leicester, Department of Computer Science,
http://www.cs.le.ac.uk/ aheyworth/grobner/ (2001).

[13] Hofmeyr, Jan-Hendrik S., Snoep, Jacky L., Westerhoff, and Hans, V., *Kinetics, Control and Regulation of Metabolic Systems*, J.-H.S 3-77, 2002.

[14] Hurlebaus, Jochen, *A Pathway Modeling Tool for Metabolic Engineering*, Institut für Biotechnologie Jülich-3912 Bundesrepublik Deutschland, 2001.

[15] Kanemitsu, F., Kageoka, T., and Kira, S., *Heterogeneity of mitochondrial creatine kinase*, Journal of Chromatography B, 806, pp.95-100, 2004.

[16] Laidler, Keith J., and Peter S. Bunting's, *The Chemical Kinetics of Enzyme Action* Second Edition, Clarendon Press, Oxford, 1973.

[17] Li, S., and Petzold, L., *Software and algorithms for sensitivity analysis of large-scale differential-algebraic systems*, J. Comput. Appl. Math. 125, 131-145, 2000.

[18] Matlab User Manual,
http://www.mathworks.com/access/helpdesk/help/techdoc/ref/ode45.html.

[19] Michalewicz, Z., and Fogel, D.B., *How to Solve It: Modern Heuristics*, Springer, 1998.

[20] Morbiduccia, Umberto, Turab, Andrea, and Grigionia, Mauro, *Genetic algorithms for parameter estimation in mathematical modeling of glucose metabolism*, to be published in "Computers in Biology and Medicine".

[21] Nougues, José Maria, Grau, M. Dolors, and Puigjaner, Luis, *Parameter estimation with genetic algorithm in control of fed-batch reactors*, Chemical Engineering and Processing 41, pp. 303-309, 2002.

[22] Shampine, Lawrence F., Reichelt, Mark W., Kierzenka, and Jacek A, *Solving index-I DAESs in MATLAB and simulink*, SIAM Review Vol 41, No.3, 538-552, 1999.

[23] Patil, Kiran Raosaheb, Akesson, Mats, and Nielsen, Jens, *Use of genome-scale microbial models for metabolic engineering*, Current Opinion in Biotechnology 5, pp. 64-69, 2004.

[24] Petzold, Linda R., and Ascher, Uri M., *Computer Methods for Ordinary Differential Equations and Differential Algebraic Equations*, Philadelphia, SIAM, 1998.

[25] Plowman, Kent M., *Enzyme Kinetics*, New York, Mc Graw-Hill, 1972.

[26] Schulte-Thomas, and Dörrscheidt, F., *Simplification of a mathematical model for the exercise metabolism. Proceedings 4th Mathmod Vienna - International IMACS Symposium on Mathematical Modelling*, Volume 1 and Volume 2. ARGESIM Report No. 24, 2003.

[27] Schulte, A., Kracht, P., Dörrscheidt, F., and Liesen, H., *Modeling and simulation of the human exercise metabolism*, conference volume "Software and Hardware Engineering for the 21st Century", Military Institutions of University Education, Hellenic Naval Academy, Greece, 377-382, 1999.

[28] Tischendorf, Caren, *Solution of Index-2 Differential Algebraic Equations and Its Applications in Circuit Simulation*, Dissertation, Fachbereich Mathematik der Humboldt-Universität zu Berlin, 1996.

[29] Trevor, Hastie, Tibshirani, Robert, and Friedman, Jerome, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, Springer, 2001

[30] Torres, V. Nestor, Voit, O. Eberhard, *Pathway Analysis and Optimization in Metabolic Engineering*, New York, Cambridge University Press, 2002.

[31] Wolfgang, Wiechert, *Modeling and simulation: tools for metabolic engineering*, Journal of Biotechnology 94, pp.37-63, 2001.

[32] Wolfgang, Wiechert, *C Metabolic Flux Analysis*, Metabolic Engineering 3, pp.195-206, 2001.

[33] Voet, Donald, *Fundamentals of Biochemistry*, New York, Wiley, 2002.

[34] Diwan, Joyce J., *Glycolysis and Fermentation*, www.rpi.edu/dept/bcbp/molbiochem/MBWeb/mb1/part2/glycolysis.htm.

[35] Rule, Gordon S., *Enzyme Kinetics*, http://stingray.bio.cmu.edu/web/bc/Lec/Lec17/lec17.PDF.

[36] Birch, Peter, *Kinetics of multisubstrate enzymes* copyright, University of Paisley, http://orion1.paisley.ac.uk/kinetics/Chapter-4/contents-chap4.html.

[37] Adrian, Brown, *More Mechanisms: A Model of Enzyme Kinetics*, http://dept.physics.upenn.edu/courses/gladney/mathphys/subsection4-1-6.html.

[38] Yıldırım, Necmettin, *Sembolik ve Nümerik metotlarla Enzim Kinetiği Problemlerinin İncelenmesi*, PhD. Thesis, Atatürk Universitesi Fen Bilimleri Enstitüsü, Erzurum, 2000.

[39] Yıldırım, Necmettin, and Bayram, Mustafa, *Derivation of Conservation Relationships for Metabolic Networks Using MAPLE*, Applied Mathematics and Computation 112, pp. 255–263, 2000.

[40] Yu Long-Jiang, Lan Wen Zhi, Chen Chao, and Yang Ying, *Glutathione levels control glucose-6-phosphate dehydrogenase activity during elicitor-induced oxidative stress in cell suspension cultures of Taxus chinensis*, Plant Science 167, pp. 329–335, 2004.