



Faculteit Bio-ingenieurswetenschappen

Academiejaar 2012-2013

Experimental design for multi-enzyme processes

Niels Nicolai

Promotors: Prof. Dr. ir. Ingmar Nopens

Prof. Dr. ir. Bernard De Baets

Tutor: Ir. Katrijn Cierkens

Masterproef voorgedragen tot het behalen van de graad van
Master in de bio-ingenieurswetenschappen: Chemie en bioprocestechnologie



Faculteit Bio-ingenieurswetenschappen

Academiejaar 2012-2013

Experimental design for multi-enzyme processes

Niels Nicolai

Promotors: Prof. Dr. ir. Ingmar Nopens

Prof. Dr. ir. Bernard De Baets

Tutor: Ir. Katrijn Cierkens

Masterproef voorgedragen tot het behalen van de graad van
Master in de bio-ingenieurswetenschappen: Chemie en bioprocestechnologie

De auteur en promotor geven de toelating deze scriptie voor consultatie beschikbaar te stellen en delen ervan te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting uitdrukkelijk de bron te vermelden bij het aanhalen van resultaten uit deze scriptie.

The author and promotor give the permission to use this thesis for consultation and to copy parts of it for personal use. Every other use is subject to the copyright laws, more specifically the source must be extensively specified when using results from this thesis.

Gent, 6 Juni 2013

De promotor,

De promotor,

De auteur,

Prof. dr. ir. Ingmar Nopens

Prof. dr. Bernard De Baets

Niels Nicolai

Woord van Dank

*“As the circle of light increases,
so does the circumference of darkness around it.”*

— Albert Einstein

Eindelijk is het zover, mijn eigen “meesterwerk” op papier. Ik moet zeggen, zo een thesis schrijven is toch niet van de poes! Maar toch, het was een fantastische ervaring die me bovendien een enorme kennis heeft opgeleverd. Ik herinner ze mij nog goed, die eerste dagen van mijn thesis. Juist terug na een maand rondtrekken doorheen de Canadese wouden, geen computer gezien in weken, dan vraag je jezelf al snel af waarmee je eigenlijk bezig bent. Gelukkig werd en word ik nog steeds gesteund door een hoop fantastische mensen die mij overigens altijd zullen bijblijven.

In de eerste plaats wil ik natuurlijk mijn moeder, Jens, Thor en vader bedanken voor alles wat jullie tot nu toe gedaan hebben voor mij. Het is dankzij hen dat ik zo een geweldige tijd heb beleefd en al zeker de laatste jaren. Zij lieten al mijn grote wensen toe: een jaar op kot, Las Vegas & de Grand Canyon bezoeken, rondtrekken in Canada, citytrippen in Rome, binnenkort een maand naar Indonesië, . . . Kortom zij hebben ervoor gezorgd dat ik dit alles kon bereiken.

Diegene die ik waarschijnlijk het meest dank verschuldigd ben voor de voorbije twee jaar is Lise. Zij heeft er voor gezorgd dat elke dag een goede dag was.

De voorbije jaren heb ik ook heel wat fantastische momenten beleefd met oude vrienden zoals Stijn, Frédéric en Pepijn, alsook met een hele boel nieuwe vrienden zoals Laurens, Tommy, Stef, Margo, Dieter, Valerie, Harm, Thomas vdv, Louise, Ine, . . . Zij verdienen dan ook enige dank!

Ik zou beide promotoren willen bedanken voor al hun hulp, het vertrouwen dat ze mij hebben gegeven om dit intellectueel uitdagend onderwerp tot een goed einde te brengen. Ingmar zal ik altijd herinneren als een persoon die oor heeft naar alles en openstaat voor iedereen. Ik moet toegeven dat ik mij het afgelopen jaar een beetje meer BIOMATHer gevoeld heb dan KERMITer maar toch wil ik ook Bernard bedanken om mij dit jaar tot zijn fantastische vakgroep te rekenen. Vervolgens ben ik ook heel wat dank verschuldigd aan mijn altijd gemotiveerde begeleidster Katrijn. Afgelopen jaar vormden wij een top team waarbij Katrijn instond voor het nodige stuurwerk, waarvoor dank! Daarnaast is ook Timothy met zijn hilarische lach een goede vriend geworden. Soms maakte hij ons wel jaloers met de luxehotels waar hij in het kader van de wetenschap mocht verblijven maar dat maakte hij ruimschoots goed met al zijn hulp. Timothy, het ga je goed! Ook Stijn verdient enige bedanking voor al zijn handige tips & tricks, voor het delen van zijn open source visies en vooral ook om me op te nemen in het sfeervolle simulatie kot. Ook al heeft ze het laatste jaar weinig te maken gehad met mijn thesis, toch wil ik ook Elena bedanken, al zij het voor de drie jaar dat ik haar als assistente had. Het moet gezegd worden, de oefeningenlessen van Elena behoren tot de betere die ik gedurende mijn opleiding heb gekregen en waren een belangrijke reden om een thesis bij BIOMATH aan te vangen. De vele activiteiten van deze vakgroep, i.e. BIOMATH fuif, BIOMATH weekend, BIOMATCH voetbal, . . . zijn natuurlijk ook wel mooi meegenomen. Tinne, ook jij bedankt dat ik het afgelopen jaar van al jouw baktalenten mocht genieten. Jouw rol binnen deze onderzoeksgroep valt niet te onderschatten! Daarnaast zijn er nog heel wat personen op deze aangename vakgroep die ook een bedanking verdienen: Joris, Giacomo, Wouter, Yuri, Lieven, Thomas, Michiel, . . . alsook mijn thesislotgenoten: Pieter, Niels, Aaron en Ayla.

Niels Nicolai
Gent, 6 Juni 2013

Contents

List of Tables	xi
List of Figures	xiii
Acronyms	xv
I Introduction	1
II Literature review	7
1 Biocatalysis	9
1.1 Introduction to the field of enzyme technology	9
1.2 Enzyme cofactors	9
1.3 Multi-enzyme processes	10
1.4 (S)-1-Phenylethylamine (PEA) for optical pure amines	11
2 Modelling enzyme kinetics	13
2.1 Introduction to mathematical modelling	13
2.2 Single-substrate reactions	13
2.3 Multi-substrate reactions	15
2.4 Constraints on the kinetic parameters	18
3 Calibration of enzyme kinetic models	21
3.1 Calibration methods	22
3.1.1 Initial rate analysis	22
3.1.2 Progress curve analysis	22
3.1.3 Incremental methods	23
3.2 Calibration of the (S)-1-phenylethylamine (PEA) system	23
3.3 Identifiability	24

4	Optimal experimental design for enzyme processes	25
4.1	Optimal experimental design for parameter estimation	25
4.2	Optimal experimental design for model discrimination	26
III	Materials and methods	27
5	Mathematical modelling and tools	29
5.1	Model implementation and simulation	29
5.2	Mathematical parameter optimisation	29
5.2.1	Defining the objective function	29
5.2.2	Modifications of the objective function	30
5.2.3	Metaheuristics	31
5.3	Identifiability analysis	34
5.4	Sensitivity analysis	35
5.5	Subset selection	37
6	Software	41
6.1	Model and experimental design implementation	41
6.2	Mathematical parameter optimisation	41
6.3	Identifiability analysis	43
7	Experimental data	45
7.1	Real experiments	45
7.1.1	Experimental set-up	45
7.2	Virtual experiments	45
IV	Results and discussion	47
8	Progress curve analysis	49
8.1	Introduction	49
8.2	One single enzyme batch reaction	50
8.2.1	Standard approach	50
8.2.2	Improvements of parameter estimation	54
8.3	Impact of the experimental design	59
8.3.1	Optimal experimental design for parameter estimation	59
8.3.2	Multiple experimental designs	59
8.4	Structural identifiability analysis	60

9	Initial rate analysis	63
9.1	A systematic approach for estimating the kinetic parameters of biocatalytic reactions	63
9.2	Structural identifiability of the initial rate model	64
9.3	Practical identifiability of the initial rate model	69
9.4	Subset selection for experimental design	71
10	The coupled tri-enzyme system	75
10.1	Proof of concept	75
V	Conclusions and perspectives	77
VI	Appendices	83
A	Relation kinetic parameters and rate constants	85
B	Experimental designs	89
C	Results and discussion	91
VII	Bibliography	99
	Bibliography	101

List of Tables

5.1	Different parameters of the simplex-simulated annealing (SIMPSA) algorithm itself	34
6.1	Information contained in the data structure representing the mathematical model.	42
6.2	Information contained in the data structure representing an experimental design.	42
8.1	Lower (lower bound (LB)) and upper bounds (upper bound (UB)) for the different kinetic parameters of the lactate dehydrogenase (LDH) subsystem.	50
9.1	Settings used for the SIMPSA and simplex algorithm in order to accelerate estimation of the initial rate parameters.	66
A.1	The following characters representing either substrate or product, will be used throughout this thesis.	85
A.2	Definitions of the kinetic parameters for both mechanisms. Although K_{iB} is not in Eq. (2.10), it can be derived using the Haldane relation given in Section 2.4 .	86
A.3	The inverse relations between the rate constants and the kinetic parameters for the compulsory-order ternary-complex mechanism	87
B.1	Experimental design of the different LDH experiments.	89
B.2	Experimental design of the different LDH initial rate experiments. All experiments were measured for 1 minute with a measurement frequency of 20 Hz. . .	90
B.3	Experimental design of the different GDH experiments.	90
C.1	Parameter values after optimisation.	91
C.2	Summarizing result after 20 iterations of progress curve analysis. Standard settings; starting from the same initial guess.	92
C.3	Summarizing result after 20 iterations of progress curve analysis. Standard settings; starting from the different initial guesses.	92
C.4	Summarizing result after 20 iterations of progress curve analysis. No initial temperature; starting from the same initial guess.	93
C.5	Summarizing result after 20 iterations of progress curve analysis. Freezing temperature of 0.1; starting from the same initial guess.	93
C.6	Summarizing result after 20 iterations of progress curve analysis. Multiplying the value of the objective function with a factor; starting from the same initial guess.	94

C.7	Summarizing result after 20 iterations of progress curve analysis. Taking into account the additional constraints; starting from the same initial guess.	94
C.8	Summarizing result after 20 iterations of progress curve analysis. Considering the literature parameters as constants; starting from the same initial guess.	95
C.9	Summarizing result after 20 iterations of progress curve analysis. Weighted by the maximum of the total relative sensitivity functions; starting from the same initial guess.	95
C.10	Summarizing result after 20 iterations of progress curve analysis. Weighted by the maximum of the total relative sensitivity functions; starting from the same initial guess.	96
C.11	Summarizing result after 20 iterations of progress curve analysis. Considering data from multiple experiments; starting from the same initial guess.	96
C.12	Parameter values for the LDH-1 isoenzyme taken from the literature (Borgmann et al., 1975). Note that these are not valid for the experimental data which was available to us.	97
C.13	Remaining parameter values needed to generate <i>in silico</i> data.	97
C.14	Estimated parameter values using progress curve analysis. The values in bold were obtained from the literature.	98

List of Figures

1.1	An enzymatic cofactor regeneration cycle.	10
1.2	Tri-enzyme process for the production of (S)-1-phenylethylamine (PEA)	12
2.1	(a) Dynamic behaviour of the substrate and product concentrations during a batch experiment. (b) Reaction rate at different substrate concentrations.	15
2.2	Classification of multi-substrate enzyme mechanisms.	16
2.3	The compulsory-order ternary-complex mechanism for the LDH and glucose dehydrogenase (GDH) conversions (2.3a) and the substituted enzyme mechanism for the transaminase (TA) conversion (2.3b)	17
3.1	(a) Initial rate analysis (r_{init} : initial reaction rate). (b) Progress curve analysis.	23
5.1	Elementary operations in a two-dimensional space used by the Nelder and Mead simplex optimisation algorithm. (a) Simplex at start (lower vertex has highest value for the objective function), (b) reflection, (c) reflection and expansion, (d) contraction, (e) multiple contraction	32
5.2	The following grid represents the total set of experimental designs. Note that A and B are two degrees of freedom which vary for each experiment in the set. The reference experiment is represented by one of the black boxes, while the neighbouring experiments are shown in grey.	40
8.1	Progress curve of NADH concentration during an LDH batch reaction (experimental design no. 1). Experimental data vs. model prediction, note the systematic error of the residuals.	51
8.2	Using the parameters from Tabel C.1 the following progress curves and total relative sensitivity functions are obtained for both experiment no. 1 and no. 10.	53
8.4	The temperature course for different settings of the cooling rate (δ) and minimal cooling factor (f_{min}).	54
8.3	In order to estimate the parameters of the LDH model, different approaches were examined. In this figure an overview of the standard deviation between the estimates of different iterations is given. This is done for each of the proposed improvements.	55

8.5	The estimated value of V_1 for different iterations taking into account multiple <i>in silico</i> generated datasets. Note that the result of different iterations is scattered over the whole parameter space.	61
9.1	The initial reaction rate ($r_{init,f}$) as a function of the initial concentration of pyruvic acid (Pva) and β -nicotinamide adenine dinucleotide, reduced (NADH).	65
9.2	Comparison between the estimates of both the tuned simplex algorithm and the tuned SIMPSA algorithm. Note that the boundaries of the last 25 iterations are a thousandfold smaller than the first 25 iterations.	67
9.3	Visualisation of the weighted sum of squared errors (WSSE) objective function near the optimal parameter values. Each mark represents a point in the parameter space for which the value of the objective function was calculated.	68
9.4	The local relative sensitivity functions (SFs) of the initial rate model. Note that the model has two independent variables ($NADH_0$ and Pva_0), which explains the necessity for multiple plots.	68
9.5	The influence of measurement noise (a) and the number of consecutive data points used for deriving the initial reaction rate (b). Note that the lines connecting the marks are only drawn to simplify the use of this figure. They should thus not be interpreted as continuous data.	70
9.6	The result after subset selection using either the simplex (a) or the SIMPSA (b) optimisation algorithm for evaluating the goodness of a subset.	72
10.1	Simulation results for the tri-enzyme system. The initial concentrations are shown in the upper left corner.	76
10.2	Simulation results for the tri-enzyme system if the limiting rate (V_1) of the TA reaction is multiplied with a factor 200 (Santacoloma, 2012).	76

Acronyms

ω -TA	ω -transaminase
AE	algebraic equation
APH	acetophenone
API	active pharmaceutical ingredient
AT	aminotransferase
ATA	amine transaminase
ATP	adenosine-5'-triphosphate
DAISY	differential algebra for identifiability of systems
EC	Enzyme Commission number
FADH	flavin adenine dinucleotide
FDH	formate dehydrogenase
FIM	Fisher Information Matrix
GDH	glucose dehydrogenase
GSA	global sensitivity analysis
IS(C)PR	<i>in situ</i> (co)product removal
IUBMB	International Union of Biochemistry and Molecular Biology
IVP	initial value problem
L-Ala	L-alanine
LB	lower bound
LDH	lactate dehydrogenase
LSA	local sensitivity analysis

MBE	(S)- α -methylbenzylamine
NAD	β -nicotinamide adenine dinucleotide, oxidized
NADH	β -nicotinamide adenine dinucleotide, reduced
NLR	nonlinear regression
ODE	ordinary differential equation
OED	optimal experimental design
OED/MD	optimal experimental design for model discrimination
OED/PE	optimal experimental design for parameter estimation
PEA	(S)-1-phenylethylamine
Pva	pyruvic acid
QSSA	quasi-steady-state assumption
SF	sensitivity function
SIMPSA	simplex-simulated annealing
TA	transaminase
TRS	total relative sensitivity
UB	upper bound
WSSE	weighted sum of squared errors
WSSRE	weighted sum of squared relative errors

Abstract

Multi-enzyme processes are considered the next-generation of biocatalytic applications. In order to get a thorough understanding of their complex dynamic behaviour, mathematical modelling can be used as a valuable tool. In this work, a theoretic study of a deterministic kinetic model based on the bi bi compulsory-order ternary-complex mechanism of lactate dehydrogenase is conducted. Nowadays, calibration of enzyme kinetic models is no longer restricted to laborious, reagent demanding and error-prone graphical analysis because of the availability of high-speed computers which allow for nonlinear dynamic regression of the entire progress curve. In this respect, practical identifiability of batch reaction experiments was investigated using the global SIMPSA optimisation algorithm. Analysis revealed that no unique set of optimal parameters could be found even when multiple improvements to the objective function and experimental design were implemented. This clearly indicates that the complete model based on progress curves of batch reactions lacks practical identifiability. In addition, structural identifiability analysis, using the DAISY software package, was unsuccessful since computational problems occurred due to the highly complex model structure. Therefore, an alternative approach based on a decomposition of the full model was investigated. This method, proposed by Al-Haque et al. (2012), shows clear advantages since both structural and practical identifiability, using noise-corrupted *in silico* data, of the initial rate models are guaranteed. However, such stepwise incremental methods require lots of different data sets to calibrate the models. In this respect, a case-specific subset selection algorithm was used to select the most informative experimental designs. From the analysis it can be concluded that only three well-selected experiments are needed for accurate parameter estimation, when using a global optimisation algorithm and noise-free data. Finally, as a proof of concept, virtual simulations of a tri-enzyme process for the production of (S)-1-phenylethylamine were conducted to show the potential of multi-enzyme systems.

Samenvatting

Multi-enzym processen worden aanzien als de nieuwe generatie van biokatalytische toepassingen. Om inzicht te krijgen in hun complexe dynamische gedrag kunnen wiskundige modellen gebruikt worden. In deze thesis werd een theoretische modelleerstudie uitgevoerd met betrekking tot het enzym lactaat dehydrogenase, dat het bi bi compulsory-order ternary-complex mechanisme volgt. Kalibratie van kinetische enzymmodellen is tegenwoordig niet langer beperkt tot een grafische analyse, die zeer arbeidsintensief en foutgevoelig is. Computers met hun enorm rekenvermogen kunnen immers vandaag de dag ingezet worden en deze laten niet-lineaire dynamische regressie van het volledige concentratieverloop toe. Dit liet toe de praktische identificeerbaarheid van batchgewijze experimenten na te gaan door gebruik te maken van het globale SIMPSA optimalisatie algoritme. Uit dit onderzoek is echter gebleken dat er geen unieke set van optimale parameters kan bepaald worden. Zelfs na het aanbrengen van verscheidene aanpassingen aan zowel objectief-functie als aan het experimentele ontwerp werd geen verbetering waargenomen. Dit wijst er duidelijk op dat het volledige model, gebaseerd op het totale concentratie verloop van batch reacties, niet praktische identificeerbaar is. Tevens werden er computationele problemen vastgesteld wanneer de structurele identificeerbaarheid werd bepaald door middel van het DAISY software pakket, wat te wijten is aan de hoge complexiteit van het model. Om deze redenen werd geopteerd voor een alternatieve aanpak waarbij het volledige model opgesplitst wordt. De stapsgewijze incrementele methode voorgesteld door Al-Haque et al. (2012) heeft het voordeel dat zowel structurele en praktische identificeerbaarheid gegarandeerd zijn voor de gebruikte initiële reactie modellen wanneer met ruis verstoorde *in silico* data gebruikt wordt. Echter, dergelijke methode vereist heel wat experimentele data voor het kalibreren van de verschillende modellen. Daarom werd een specifiek subset selectie algoritme voorgesteld, om zo de meest informatieve experimenten te selecteren. Uit deze analyse kon besloten worden dat er slechts drie goed gekozen experimenten nodig zijn om de parameters accuraat te schatten, gebruik makende van een globaal optimalisatie algoritme en perfecte data. Tot slot werden ook virtuele simulaties van een tri-enzym process dat kan instaan voor de productie van (S)-1-phenylethylamine, uitgevoerd. Deze simulaties gaven naast proof of concept ook het potentieel van multi-enzym systemen weer.

PART I

INTRODUCTION

Background

The urge for environmentally friendly products and production processes recently gained tremendous attention. Current industries are pushed to find innovative green alternatives while maintaining product quality, or even improving it. In this respect, biocatalysis has become a key component in many industrial production processes. The mild operational conditions and the effective utilisation of raw materials are often cited as the most important positive features of nature's catalysts. However, most industrial applications are limited to enzymes that perform relatively simple chemistry since these biocatalysts do not require expensive chemical auxiliaries, e.g. pyridine nucleotide cofactors. This is one of the reasons why, to date, the production of fine chemicals such as pharmaceuticals is still limited to processes comprised of relatively simple steps. Clearly, such multi-step procedures are laborious, expensive and time consuming. Nonetheless multi-enzyme processes in which the catalytic activity of multiple enzymes working together simultaneously, may offer clear solutions. Indeed, such processes can be used as an innovative alternative for effective recycling of the cofactors needed in more complex biocatalytic conversions. However, few cases have been implemented at pilot or industrial scale because of the increased complexity of these next-generation processes. Therefore, detailed knowledge of the process technology is essential for effective implementation and operation. Modelling and simulation may assist in exploiting the potential benefit of these processes. Nowadays, reliable mathematical models are increasingly used for design, optimisation and control of industrial processes. However, few mathematical models have been developed for describing the dynamic behaviour of multi-enzyme processes. Yet, the individual enzymatic conversions should first be understood. Numerous techniques exist to calibrate and validate single enzyme reaction models. As long as no powerful computers were available, scientists relied on analytic and graphical tools for fine-tuning their models. Although those techniques have shown to be very useful for simple cases, highly complex systems which are common in systems biology require more advanced numerical techniques. Despite the ever increasing availability of qualitative data, calibration of complex models using such computationally demanding techniques still remains a hard nut to crack.

Objectives

In this MSc. thesis a tri-enzyme process which can be used for the production of an important synthon in the pharmaceutical industry will be investigated. In particular, the research will focus on one of the three reactions in this system, i.e. the lactate dehydrogenase conversion.

In a first step, identifiability of the complete kinetic model will be evaluated. Structural identifiability will be scrutinised using an existing software toolbox specifically designed for nonlinear

dynamic models. Furthermore, practical identifiability of the model will be assessed using different approaches.

The second part of this thesis deals with experimental design. In this respect, a recent study of Al-Haque et al. (2012) which claims to provide an attractive alternative for model calibration based on decomposition of the complete model, will be thoroughly investigated w.r.t. experimental design. Regarding this approach, a straightforward case-specific method for selecting informative subsets of experiments will be developed and evaluated.

Finally, the usefulness of the tri-enzyme system for the production of (S)-1-phenylethylamine will be demonstrated using virtual simulations.

Outline

This thesis is comprised of four major parts: “Literature review”, “Materials and methods”, “Results and discussion” and “Conclusion and perspectives”.

In the first part an extensive review of the available literature is provided in order to gain insight into the subject of enzyme kinetics modelling. First, an introduction to general principles within the field of biocatalysis is given. The case study, which laid the foundation for this research, is also discussed in the first chapter. Next, a brief yet profound overview on the construction and use of enzyme kinetic models is given. Chapter 3 of the literature review concerns the different methods for calibrating these models. At the end of this part, the use of optimal experimental design for enzyme kinetic models is described.

The second part deals with the different methods and software packages used in this thesis. To get the reader acquainted with the subject of mathematical modelling and mathematical tools such as parameter optimisation, identifiability analysis, sensitivity analysis and subset selection, a clear introduction is given first. Subsequently, the different software packages, as well as their use, are explained. The last chapter of this part concludes with the used experimental data and how they were obtained.

The “Results and discussion” part in this thesis is divided in three major chapters. The first chapter deals with the different methods that were used to evaluate practical identifiability of the complete model as well as structural identifiability. For practical identifiability a division is made between the standard approach and approaches using different improvements. In the next chapter, a robust method for calibrating enzyme kinetic models is evaluated w.r.t. experimental design. Since this method decomposes the complete “progress curve” model into less complex “initial rate” models, the focus in this chapter is shifted towards the latter type of models. Different analyses such as practical identifiability and sensitivity analysis, were conducted in order to get a better understanding of the used initial rate model. It is also in this part of the

thesis that the results of a self-implemented subset selection algorithm are discussed. The third chapter of this part is dedicated to simulation results of the coupled tri-enzyme system.

Finally, in the last part of this thesis the main conclusions that can be drawn from the obtained results are summarised and recommendations for future research are given.

PART II

LITERATURE REVIEW

CHAPTER 1

Biocatalysis

1.1 Introduction to the field of enzyme technology

Catalysis can be defined as a process in which the reaction rate of chemical reactions is increased by the use of a substance which is not consumed during the reaction, i.e. the catalyst. Not only modern industrial chemical processes but also nature uses catalysts for speeding up most of its biochemical conversions. Nature's catalysts are known as biocatalysts and were discovered only two centuries ago. Nevertheless, they have already been used for thousands of years in brewing and baking processes (Thompson, 1986). Biocatalysts play a key role in all living organisms on Earth as they regulate and control all metabolic reactions.

The first fully enzymatic industrial process dates from the mid-1960s and was developed for the conversion of starch into glucose syrup using a bacterial α -amylase and a fungal γ -amylase (Illanes, 2008). Nowadays, state-of-the-art enzyme technologies such as asymmetric synthesis have become well-established manufacturing processes (Nestl et al., 2011). It is forecasted that the global market for industrial enzymes will reach up to US\$ 3.74 billion by the year 2015. The main driving forces for market growth include new enzyme technologies that enhance cost efficiency and productivity, but also the growing interest among consumers for greener products (GIA, 2012). This argument follows from the fact that these biomolecules are able to work under rather mild conditions, i.e. gentle temperature, pH and pressure which saves on energy and chemical addition.

1.2 Enzyme cofactors

In a general biocatalytic conversion one (or more) substrate(s) is converted into the desired product(s) using the enzyme as a catalyst. However, some enzymes need an additional helper molecule in order to be active, i.e. a cofactor. Cofactors are either inorganic metal ions like Mg^{2+} , Fe^{2+} and Zn^{2+} or small organic molecules known as coenzymes such as NADH, flavin adenine dinucleotide (FADH) and adenosine-5'-triphosphate (ATP). These helpers can either be

converted or recycled during the catalytic reaction. In the former case, an additional regeneration reaction is needed to reobtain the active cofactor (Willner and Mandler, 1989).

1.3 Multi-enzyme processes

Classically, industrial synthesis of fine chemicals such as bioactive molecules is carried out in multiple steps. Here, each step comprises a single chemical reaction. Although this approach allows for excellent control, it requires a lot of time, energy, labour and space. Moreover, large amounts of waste are produced, resulting in relatively low yields caused by the purification step usually needed after each reaction in the sequence (Garcia-Junceda, 2008). Recent research is aiming at the reduction of the economic and environmental impact of such processes. A possible way to achieve this is by combining multiple reactions in one step. In this way no intermediate purification steps are needed. The idea of combining multiple steps is not only applicable for pure chemical conversions, but also enzyme-catalysed reactions can be combined, giving rise to a multi-enzyme process in which two or more enzymes are present in the same biochemical reactor (Cornish-Bowden, 2004; Santacoloma, 2012). Multi-enzyme conversions are merely a scale expansion of what is happening in the cell metabolism. However, the aim is not to reproduce the whole metabolic network but only a desired part of it.

Various configurations for multi-enzyme conversions exist. A first type is a cascade in which the product of a first reaction is used as a substrate for a second reaction. Other possibilities are parallel reactions and the network configuration of reactions. In the former multiple enzymes compete for the same substrate whereas in a network configuration certain components are used as connections between multiple reactions of the overall configuration. An example of such a multi-enzyme network configuration is an enzymatic cofactor regeneration cycle. Some cofactors, such as NADH or FADH, have to be reactivated after their intervention in the biocatalytic conversion (Section 1.2). Continuous addition of active cofactor seems an obvious solution. However, this approach is impracticable because cofactors are still very expensive (Serrano Briega, 2011). A more feasible solution is the implementation of an additional regeneration reaction whose sole function is to reactivate the inactivated cofactors (see Figure 1.1) (Santacoloma, 2012).

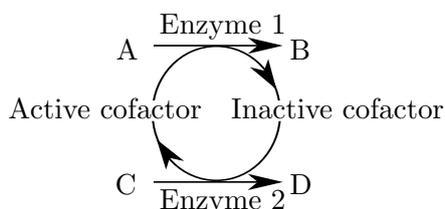


Figure 1.1: An enzymatic cofactor regeneration cycle.

Full-scale examples of multi-enzyme processes are still exceptional. This is mainly due to the

higher degree of complexity of these systems compared to single-enzyme systems. Therefore, it is necessary to gain insight in the behaviour of enzymes and substrates in such complex environments. Note that, although the conditions can be very favourable for a certain step in the conversion, it could be that this is not the case for the other steps. In such cases enzyme engineering may offer solutions (Santacoloma, 2012).

1.4 (S)-1-Phenylethylamine (PEA) for optical pure amines

The previous section dealt with the general characteristics of multi-enzyme processes. In this section a detailed description of the tri-enzyme process for the production of enantiopure (S)-1-phenylethylamine (PEA) is given.

Optically active chiral amines like PEA are important building blocks in the pharmaceutical sector for the synthesis of active pharmaceutical ingredients (APIs) applied in antihypertensives, antihyperglycemics, HIV-protease inhibitors. . . (Drauz et al., 1990; Tufvesson et al., 2011). Enantiopure amines can be produced both by biocatalytic and chemical synthesis. However, chemical synthesis is still very inefficient (Nugent and El-Shazly, 2010). Regarding biocatalysis, there are several enzymatic routes for synthesising optically active chiral amines. Aminotransferases (ATs) or TAs classified as EC 2.6.1. are the most ubiquitous enzymes for this process in nature. The reaction carried out by these catalysts is the transfer of an amino group (R-NH_2) from the amino donor to the carbonyl carbon atom (R-CO-R) of an amino acceptor under mild conditions. TAs are very diverse, hence they can be subdivided in smaller groups. Based on the used substrates one of the groups is called the ω -transaminases (ω -TAs) or amine transaminases (ATAs) and consists of enzymes that are able to transfer an amino group between two substrates of which at least one is not an α -amino acid or an α -keto acid. In general, most of the TAs accept L-alanine (L-Ala) as the amine donor which is subsequently converted into pyruvate (Koszelewski et al., 2010).

The production of (S)-1-phenylethylamine (also called (S)- α -methylbenzylamine (MBE)) starting from acetophenone (APH) can be done by asymmetric amine synthesis (see Figure 1.2). Biocatalytic asymmetric amination is done by using an (S)-specific ω -transaminase and L-alanine as amino donor. Due to the unfavourable equilibrium constant ($K_{eq} = 8.81 \times 10^{-4}$ or 4.03×10^{-5}), an equilibration time of approximately 48 h and severe product inhibition by pyruvate and PEA this transamination reaction seems at first sight not applicable at an industrial scale (Shin and Kim, 1998; Tufvesson et al., 2012). However, there are several methods for shifting amination reactions towards the desired product side. One of the possibilities is biocatalytic degradation of the coproduct pyruvate (so called *in situ* (co)product removal (IS(C)PR)), e.g. the reduction of pyruvate to (S)-lactate using an oxidoreductase like lactate dehydrogenase (LDH) (Shin and Kim, 1999). This conversion requires a pyridine nucleotide cofactor (NADH) as a source of reducing equivalents. In order to have a practically and economically feasible production of PEA,

NAD recycling is needed (Section 1.3). Reconverting NAD back to NADH can be done by using an additional enzyme such as formate dehydrogenase (FDH) or GDH combined with a suitable reducing agent, resp. formate or glucose (Shin and Kim, 1999; Koszelewski et al., 2010).

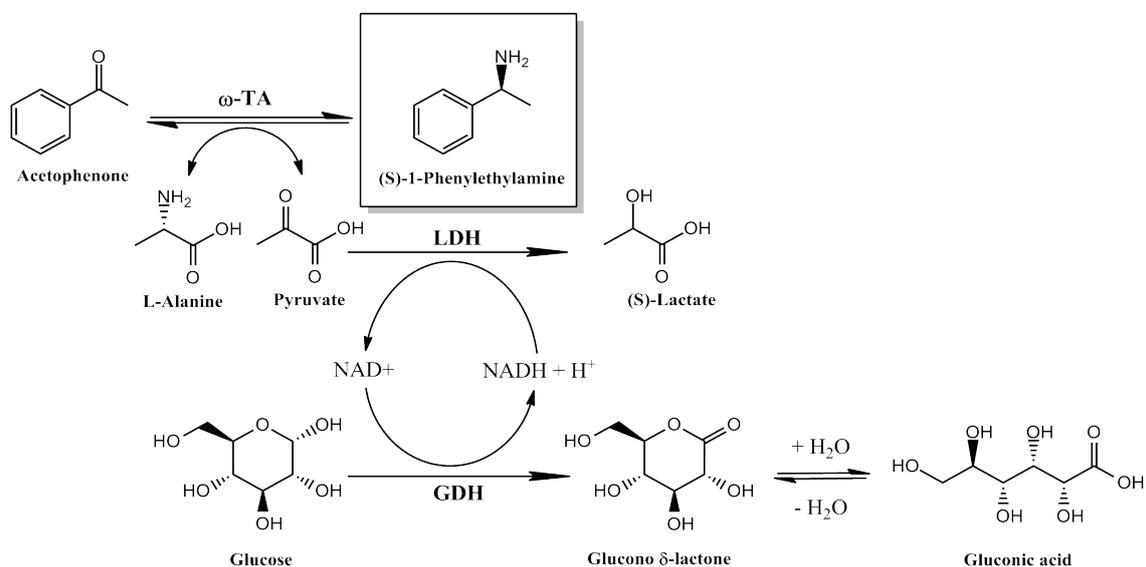


Figure 1.2: Tri-enzyme process for the production of (S)-1-phenylethylamine (PEA)

CHAPTER 2

Modelling enzyme kinetics

2.1 Introduction to mathematical modelling

The main aim of building a mechanistic mathematical model is to mimic the behaviour of a physical process (i.e. the system) in terms of one or more mathematical equations. If the model is valid for the system it represents, such a surrogate can be used as a tool for process simulation, design, control and optimisation. Various model attributes are used to characterise different types of models. The models used throughout this thesis are all classified as mechanistic, implying that the model structures are based on physical, chemical and biological laws. At the same time the models are also deterministic since they ignore random variation and thus always yield a single value for a given scenario (Dochain and Vanrolleghem, 2001).

The tri-enzyme system described in Section 1.4 can be divided into three separate systems, i.e. subsystems. Each of these subsystems contains one enzymatic reaction that can be described by means of a mathematical model. This model describes the kinetics of the enzymatic reaction, which amounts to describing the rate of change, or dynamics, in the concentration of the different reactants.

2.2 Single-substrate reactions

Although enzymatic conversions like reaction (2.1) are used for thousands of years by mankind (e.g. brewing, baking...), enzyme kinetics was still unknown territory until a century ago.



A: substrate, E: enzyme, P: product

The real increase of knowledge about enzyme kinetics and mechanisms came with the insights of the British chemist Adrian John Brown in 1902 (Laidler, 1997). It was Brown who found that the rate of the inversion of sucrose with purified invertase showed a non-linear relation between

the reaction rate and the substrate concentration. This phenomenon is now known as enzyme saturation and is the base concept in all modern biocatalytic models (Cornish-Bowden, 2004; Schulz, 1994).

When modelling a dynamic process such as an enzymatic reaction, one can use ordinary differential equations (ODEs) to describe the time-dependent behaviour of the system. The following mechanism, which consists of micro-kinetic reactions, was suggested by Brown and is used to derive the ODEs describing a single substrate enzymatic conversion, e.g. the inversion of sucrose (Laidler, 1997).



EA: enzyme substrate complex, k_i : rate constant of the i^{th} reaction

Using mass action kinetics, this batch system can be written in the following system of ODEs:

$$\begin{aligned} \frac{da}{dt} &= -k_1 e a + k_{-1} c \\ \frac{de}{dt} &= -k_1 e a + k_{-1} c + k_2 c \\ \frac{dc}{dt} &= k_1 e a - k_{-1} c - k_2 c \\ \frac{dp}{dt} &= k_2 c = r \end{aligned} \quad (2.3)$$

In this micro-kinetic model a , e and c stand respectively for the concentration of substrate, enzyme and enzyme-substrate complex (see Figure 2.1.a).

Because there is no explicit analytical solution for the reaction rate (r) in terms of the substrate concentration, the following quasi-steady-state assumption (QSSA) was introduced by Briggs and Haldane (1925):

$$\frac{dc}{dt} = k_1 e a - k_{-1} c - k_2 c = 0 \quad (2.4)$$

In words, this assumption states that the concentration of enzyme-substrate complex (c) is constant during the period of steady-state, i.e. the rate of formation equals the rate of consumption. This assumption is valid after the initial transition which takes no longer than a few milliseconds (Segel and Slemrod, 1989). Combined with the fact that the total catalyst concentration ($e_0 = e + c$) does not change during the reaction, this leads to:

$$r = \frac{dp}{dt} = \frac{k_1 k_2 e_0 a}{k_{-1} + k_2 + k_1 a} = \frac{k_2 e_0 a}{\frac{k_{-1} + k_2}{k_1} + a} \quad (2.5)$$

Eq. (2.5) can be rewritten in a more general form known as the fundamental Michaelis-Menten equation:

$$r = \frac{k_{cat} e_0 a}{K_m + a} = \frac{V a}{K_m + a} \quad (2.6)$$

Instead of using rate constants, the Michaelis-Menten equation uses experimentally measurable and meaningful (macroscopic) kinetic parameters (Cleland, 1963). For Eq. (2.5) and (2.6) the relations between the rate constants (k_i) and the kinetic parameters (K_m and V) are the following (see also Figure 2.1.b):

- $V = k_2 e_0 [\frac{\text{mol}}{\text{ls}}]$; limiting rate: the rate of the reaction when all the enzymes are saturated with substrate. Note that V depends on the total active enzyme concentration, e_0 , so it does not give a good indication of the intrinsic catalytic activity of the enzyme. To standardize this parameter, it should be divided by e_0 which results in k_2 or more generally $k_{cat} [\text{s}^{-1}]$, which is the maximum number of substrate molecules that one enzyme molecule can convert in a unit of time (Cornish-Bowden, 2004).
- $K_m = (k_{-1} + k_2)/k_1 [\frac{\text{mol}}{\text{l}}]$; Michaelis constant: the concentration of substrate which results in a reaction rate half of the limiting rate (if $a = K_m \rightarrow r = V/2$). This way K_m can be defined as the affinity of the enzyme for its substrate.

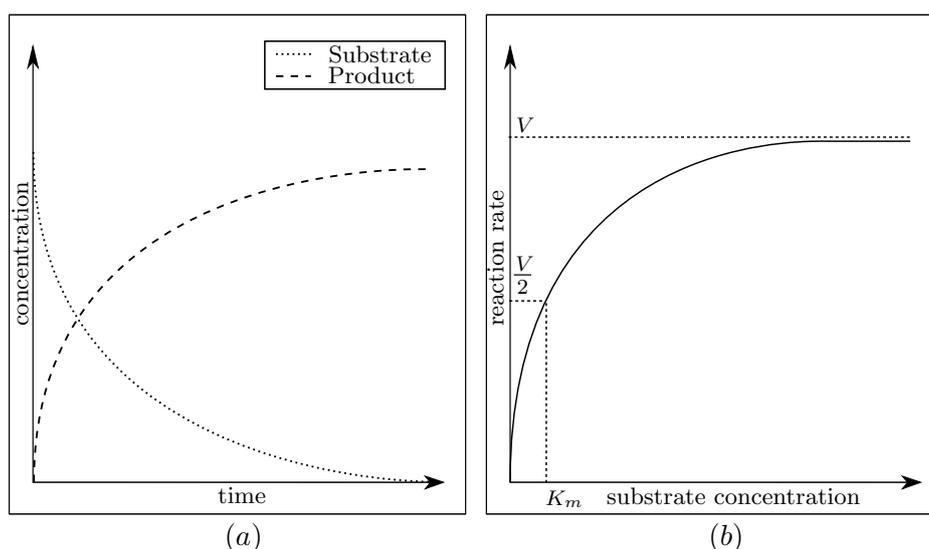


Figure 2.1: (a) Dynamic behaviour of the substrate and product concentrations during a batch experiment. (b) Reaction rate at different substrate concentrations.

2.3 Multi-substrate reactions

The mechanism used for deriving the general Michaelis-Menten equation assumes single substrate and product. Therefore, it cannot be used for enzymatic reactions with two substrates and two products. This type of reaction, known as a bi bi reaction, is by far the most common in biochemistry (Cornish-Bowden, 2004). The following equation represents a bi bi reaction in which A, B are the substrates, E is the enzyme and P, Q are the products (considering the

forward direction).



This overall reaction (2.7) does not give a good representation of what actually happens in reality. More information about the enzyme-reagent interaction is obtained when looking at the enzyme mechanism which consists of microscopic reactions. Mechanisms for multi-reactant enzymatic reactions have been classified in some general groups (Figure 2.2).

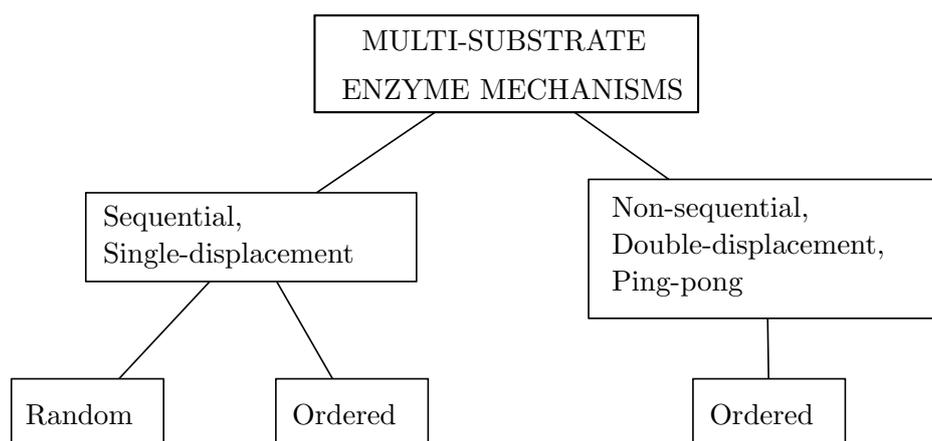


Figure 2.2: Classification of multi-substrate enzyme mechanisms.

Often a division is made between sequential mechanisms and non-sequential mechanisms. In the sequential type, both substrates must bind to the enzyme before any product is released. The sequential mechanism can be subdivided in: random-order mechanisms, in which the order of binding the substrates is not important, and compulsory-order mechanisms, when the order of binding is important (conform the induced-fit hypothesis of Koshland (1958)). For bi bi reactions the sequential mechanisms are also called ternary-complex mechanisms because of the ternary EAB/EPQ intermediate (see Figure 2.3a). Typically, kinetics of LDH and GDH can be described with a compulsory-order ternary-complex mechanism. Non-sequential mechanisms on the other hand, do not require all the substrates to bind before a product is released and thus a substituted-enzyme intermediate, E^* will be formed. In literature such mechanisms are generally referred to as ping-pong or substituted-enzyme mechanisms (see Figure 2.3b). Transaminases typically use this mechanism to transfer an amino group (Cornish-Bowden, 2004; Schulz, 1994).

In principle, the steady state rate expression of any mechanism can be derived using a similar procedure as in Section 2.2. For complex mechanisms, this will be a very laborious approach and therefore a faster graphical method developed by King and Altman is frequently used to obtain the rate equations. For the compulsory-order ternary-complex mechanism, like the LDH

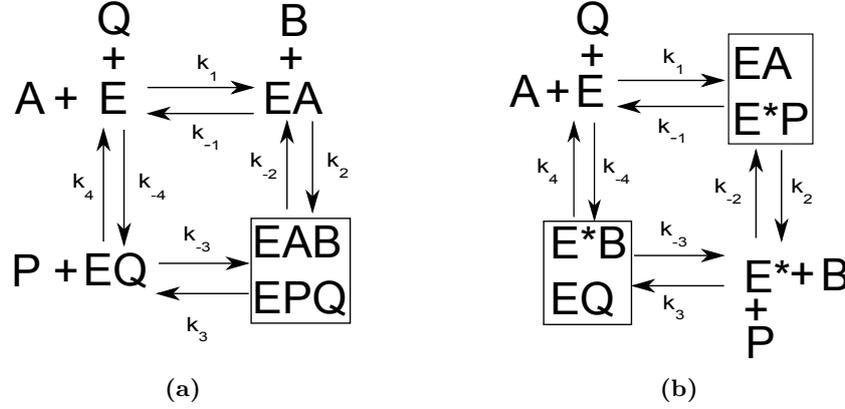


Figure 2.3: The compulsory-order ternary-complex mechanism for the LDH and GDH conversions (2.3a) and the substituted enzyme mechanism for the TA conversion (2.3b)

and GDH reactions, following rate expression is obtained (Schulz, 1994; Cleland, 1963):

$$r_{LDH/GDH} = \frac{dp}{dt} = -\frac{da}{dt} = \frac{[k_1 k_2 k_3 k_4 a b - k_{-1} k_{-2} k_{-3} k_{-4} p q] e_0}{k_{-1} k_4 (k_{-2} + k_3) + k_1 k_4 (k_{-2} + k_3) a + k_2 k_3 k_4 b + k_{-1} k_{-2} k_{-3} p + k_{-4} k_{-1} (k_{-2} + k_3) q + k_1 k_2 (k_3 + k_4) a b + k_1 k_{-2} k_{-3} a p + k_{-4} k_2 k_3 b q + k_{-4} k_{-3} (k_{-1} + k_{-2}) p q + k_1 k_2 k_{-3} a b p + k_{-4} k_2 k_{-3} b p q} \quad (2.8)$$

For the transaminase reaction a similar rate equation can be obtained.

In contrast to the single reactant enzymatic reactions, the equations described here are much more complex and cannot be expressed in only two kinetic parameters anymore (i.e. V and K_m). That is why Cleland (1963) introduced an additional kinetic parameter i.e. K_i , the core inhibition constant. When using these three kinetic parameters, rate equation (2.8) can be rewritten (recommended by the International Union of Biochemistry and Molecular Biology (IUBMB)). This leads to the following expression for the LDH and GDH conversion rates, considering the characters defined in Table A.1:

$$r_{LDH/GDH} = \frac{\frac{V_1 a b}{K_{iA} K_{mB}} - \frac{V_2 p q}{K_{mP} K_{iQ}}}{1 + \frac{a}{K_{iA}} + \frac{K_{mA} b}{K_{iA} K_{mB}} + \frac{K_{mQ} p}{K_{mP} K_{iQ}} + \frac{q}{K_{iQ}} + \frac{a b}{K_{iA} K_{mB}} + \frac{K_{mQ} a p}{K_{iA} K_{mP} K_{iQ}} + \frac{p q}{K_{mP} K_{iQ}} + \frac{K_{mA} b q}{K_{iA} K_{mB} K_{iQ}} + \frac{a b p}{K_{iA} K_{mB} K_{iP}} + \frac{b p q}{K_{iB} K_{mP} K_{iQ}}} \quad (2.9)$$

and for the TA conversion rate:

$$r_{TA} = \frac{\frac{V_1 a b}{K_{iA} K_{mB}} - \frac{V_2 p q}{K_{iP} K_{mQ}}}{\frac{a}{K_{iA}} + \frac{K_{mA} b}{K_{iA} K_{mB}} + \frac{p}{K_{iP}} + \frac{K_{mP} q}{K_{iP} K_{mQ}} + \frac{a b}{K_{iA} K_{mB}} + \frac{a p}{K_{iA} K_{iP}} + \frac{K_{mA} b q}{K_{iA} K_{mB} K_{iQ}} + \frac{p q}{K_{iP} K_{mQ}}} \quad (2.10)$$

In Table A.2 of the Appendix the relations between the kinetic parameters (K_m , K_i , V) and rate constants (k_i) are given.

The meaning of the different kinetic parameters in the multi-substrate reactions, and their relation to the kinetic parameters of the single-reactant Michaelis-Menten equation, can be derived using particular experimental conditions such as initial rate, saturation and constant concentration of one of the substrates (Cornish-Bowden, 2004).

- The limiting rate (V_1 : forward reaction, V_2 : backward reaction) represents the reaction rate in the case that all substrates in the considered direction occur at saturated concentrations, while the products are at zero concentration. In the same way as for single-substrate reactions, V for multi-substrate reactions is also dependent on the enzyme concentration.
- A clearer interpretation of the Michaelis constant (K_m) for a certain substrate X of a bi bi reaction can be achieved using a scenario in which, again, the products are at zero concentration, but only one of the two substrates is at a saturating concentration, for example substrate Y. In this case the rate equation for each of the subsystems is simplified to exactly the same expression as Eq. (2.6)

$$r_{TAm/LDH/GDH} = \frac{V_1 x}{K_{mX} + x} \quad (2.11)$$

And thus, the Michaelis constant can be defined as the substrate concentration of X when half the limiting rate is reached, taking the above scenario into account.

- The inhibition constant (K_{iX}) of a substrate X is related to the competitive inhibition constant (K_{ic}) and the uncompetitive inhibition constant (K_{iu}) obtained when the substrate X is used as a product inhibitor of the reverse reaction. Note that, besides inhibition constants such as K_{iX} , which are derived from the core mechanisms, other parameters can be included in the model if other types of inhibition are observed (Cornish-Bowden, 2004).

2.4 Constraints on the kinetic parameters

Applying the Cleland method to Eq. (2.8) (i.e. replacing the rate constants by measurable and meaningful kinetic parameters) results in a model with ten kinetic parameters, Eq.(2.9). However, the original rate model contains only eight rate constants. Consequently, there must be some redundancy among the kinetic parameters. This redundancy can be described with additional relations between the kinetic parameters.

When a reaction is at equilibrium, the reaction rate is zero and all reactants are in equilibrium. A well-known fact is that the presence of catalysts does not affect the equilibrium of the catalysed reaction. A catalyst only decreases the time needed to reach equilibrium. The thermodynamic parameter describing this steady state is called the equilibrium constant and is defined by the Haldane relationship (Schulz, 1994). As an example, the equilibrium constant for the transamination reaction can be expressed as:

$$K_{eq} = \frac{k_1 k_2 k_3 k_4}{k_{-1} k_{-2} k_{-3} k_{-4}} = \frac{K_{iP} K_{iQ}}{K_{iA} K_{iB}} = \left(\frac{V_1}{V_2} \right) \frac{K_{iP} K_{mQ}}{K_{iA} K_{mB}} = \left(\frac{V_1}{V_2} \right) \frac{K_{mP} K_{iQ}}{K_{mA} K_{iB}} = \left(\frac{V_1}{V_2} \right)^2 \frac{K_{mP} K_{mQ}}{K_{mA} K_{mB}} \quad (2.12)$$

and for both dehydrogenase reactions as:

$$K_{eq} = \frac{k_1 k_2 k_3 k_4}{k_{-1} k_{-2} k_{-3} k_{-4}} = \left(\frac{V_1}{V_2} \right) \frac{K_{mP} K_{iQ}}{K_{iA} K_{mB}} = \left(\frac{V_1}{V_2} \right)^2 \frac{K_{iP} K_{mQ}}{K_{mA} K_{iB}} \quad (2.13)$$

See Table A.2 of the Appendix for the relationship between the kinetic parameters and the rate constants.

For some mechanisms the Haldane equation explains only a part of the redundancy among the kinetic parameters. Therefore, additional non-Haldane constraints are applicable for these mechanisms. Straathof and Heijnen (1996) were the first to develop a systematic methodology for deriving such additional relations. For the compulsory-order ternary-complex mechanism the non-Haldane constraint is:

$$\frac{K_{mA}}{K_{iA}} = 1 + \frac{V_1}{V_2} \left(1 - \frac{K_{mQ}}{K_{iQ}} \right) - \frac{K_{mP} K_{iQ}}{K_{iP} K_{mQ}} \quad (2.14)$$

and for the substituted-enzyme mechanism, following constraint is valid:

$$\frac{K_{mA}}{K_{iA}} = 1 + \frac{V_1}{V_2} \left(1 - \frac{K_{mQ}}{K_{iQ}} \right) \quad (2.15)$$

CHAPTER 3

Calibration of enzyme kinetic models

As explained in the introduction of Chapter 2, mathematical models are used to simulate the behaviour of a studied process. In this way, the process behaviour can rapidly and inexpensively be investigated under various input conditions without performing the actual real-time experiment. This is referred to as model simulation or virtual experimentation. The term simulation refers to the act of solving the model, analytically or numerically, given certain experimental conditions and future model inputs. Note that different conditions and inputs will result in different model outputs. In mathematical terms an experiment is characterised by the values of the model parameters, the initial value of the different state variables and the behaviour of the process input (Dochain and Vanrolleghem, 2001).

Model parameters are those constituents of the mathematical model which do not change during the course of a single simulation, but they can differ between different simulations. Before a model can be used for any purpose whatsoever, the values of its parameters have to be determined from experimentally measured data. This inverse problem is called parameter estimation or model calibration and typically consists of minimising the deviation between the experimental data and the model prediction, by means of minimising an objective function. Once a unique set of parameters leads to the minimal deviation, the model is said to be calibrated for the used experimental data. Not only should the model predict the experimental data for calibrating, it should also be able to predict the outcome of experimental data obtained under different input conditions, this is called model validation. A model passing this is called a validated model (Donckels, 2009).

Regarding the multi-enzyme process studied throughout this thesis, the parameters are either the rate constants (k_i) or the kinetic parameters (V , K_m and K_i). There are several options to calibrate biocatalytical steady-state models. An overview is given in the subsequent sections.

3.1 Calibration methods

3.1.1 Initial rate analysis

Classically, kinetic parameters are estimated based on initial rate analysis. For this purpose, only time-concentration data at the beginning of the conversion are collected. The slope of these data points equals the initial conversion rate, i.e. the reaction rate when no products are present (see Figure 3.1(a)). Indeed, at the beginning of the reaction only reactants are present, so there will be no influence of the product on the reaction rate. Estimating parameters from the initial rate of a multi-substrate bioconversion can be done either by linear regression (with graphical linearisation) or by nonlinear regression (with algebraic parameter estimation) (Cornish-Bowden, 2004; Al-Haque et al., 2012)). Both of these approaches have the advantage of being computationally friendly and are easy to perform (Zavrel et al., 2010; Straathof, 2001). On the other hand, initial rate analysis can be very expensive and laborious because only one data point is obtained per experiment. In addition, when using graphical linearisation, the measurement errors can be distorted leading to a bias in the estimated parameters (Chen et al., 2008; Goudar et al., 1999).

3.1.2 Progress curve analysis

Instead of using only data at the beginning of the conversion, the entire time course of the conversion can be used for estimating the kinetic parameters (see Figure 3.1(b)). This approach, called progress curve analysis, is regarded as an attractive alternative to initial rate analysis. Progress curve analysis offers the ability of gathering more information from a single experiment by means of nonlinear regression (NLR). Indeed, long-term phenomena such as severe product inhibition or hysteresis due to changing enzyme activity can be derived from progress curves (Bates and Frieden, 1973). This method sounds promising but has an important limitation, namely the incompatibility of enzyme kinetic models and experimental data. Whereas kinetic models are formulated in terms of concentration dependent rates, experimental data are time-concentration data.

To solve this problem, one can differentiate the experimental data using finite difference techniques to derive the reaction rates. Direct numerical differentiation of noise-corrupted experimental data is not an option. Hence smoothing must be performed prior to the algebraic parameter estimation. This can be done by using regularisation techniques such as filtering, Fourier transformation or Tikhonov regularisation (Zavrel et al., 2010). However, such regularisation techniques inevitably lead to bias in the estimation (Marquardt, 2005). In contrast, the dynamic model can also be analytically integrated giving rise to a time-dependent function which can also be used for algebraic parameter estimation. However, no straightforward explicit analytical solution can be found even for the simplest enzyme conversion models. Therefore

numerical methods or a transcendental function have to be used (Goudar et al., 2004). After differentiation or integration, algebraic parameter estimation can be performed.

An alternative to both algebraic techniques is dynamic parameter estimation by numerical integration of the ODEs. In this method, each parameter set needs to be evaluated using a numerical solver. Although this approach requires tremendous computing power, it does not require changing the experimental data nor the model.

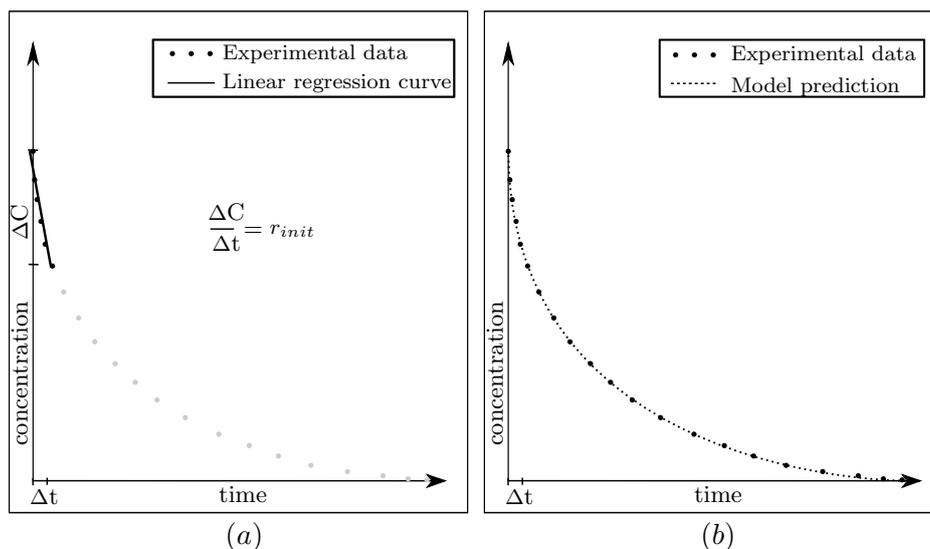


Figure 3.1: (a) Initial rate analysis (r_{init} : initial reaction rate). (b) Progress curve analysis.

3.1.3 Incremental methods

More recently, incremental hybrid methods combining preliminary algebraic parameter estimation and dynamic parameter estimation have been examined (Michalik et al., 2007; Chen et al., 2008; Al-Haque et al., 2012). In one of the first steps of this step-by-step approach some of the model parameters are estimated using initial rate analysis. While keeping the obtained values of these model constituents constant, a subsequent progress curve analysis is performed in order to determine the values of the remaining parameters. Finally, an additional progress curve analysis is carried out, but using the estimates of the previous steps as a starting point, i.e. an initial candidate solution.

3.2 Calibration of the PEA system

In the case of multi-enzyme processes such as the conversion of APH to PEA, the overall reaction can be modelled by combining the different rate expressions for each subsystem. At first glance, progress curve fitting seems to be a suitable way to estimate the parameters of a mixture of enzymes. However, according to Straathof (2001) this is not the case and thus the enzymes should be studied one at a time. The transaminase subsystem has been studied in literature. Shin and

Kim (1998) were the first to determine the kinetic parameters of a substrate-inhibited ω -TA from *Bacillus thuringiensis JS64* based on initial rate analysis. More recently, a more robust incremental method was suggested by Al-Haque et al. (2012). For the LDH and GDH subsystems, the kinetic parameters were also determined by means of initial rate analysis (Borgmann et al., 1975; Carper et al., 1983)

Fitting the predicted data to the experimental data should not be taken lightly, especially in the case of NLR problems. It is a well-known fact that there is a strong correlation between the different parameters of complex biocatalytic models. Indeed, convergence problems might occur because of multimodality, i.e. multiple local optima. This can be circumvented by (1) using a more robust optimisation algorithm, (2) doing multiple regressions at the same time, (3) starting from different initial parameter sets and (4) reducing the number of parameters to be estimated (Chen et al., 2008). For many calibrated biocatalytic models the predicted output fits the experimental data quite well. However, the parameters are not accurate at all. This can be due to an over-specified model, i.e. it is not possible to find a unique set of parameters using experimental or *in silico* data. Generally, this problem is referred to as the problem of parameter identifiability (Versyck et al., 1999; Zavrel, 2009).

3.3 Identifiability

Identifiability analysis determines whether it is possible to assign unique values to the model parameters given a data set. Generally a distinction between structural and practical identifiability is made (De Pauw, 2005). The former criterion, also called theoretical identifiability, is met when the parameters can be identified from noise-free data. Structural identifiability is only a result of the model structure, the selected output variables and the input variables. Since enzymatic models usually contain non-linearities and multiple parameters, an analysis of the structural identifiability should be performed prior to the actual model calibration. However, structural identifiability analysis is still a challenge for nonlinear dynamic models (Chis et al., 2011). Structurally unidentifiable parameters are the result of an over-specification of the model compared to the available data, i.e. an ill-posed inverse problem (Tikhonov and Arsenin, 1977). Common methods for testing the structural identifiability of nonlinear models are: the Taylor series approach, the generating series method, the similarity transformation approach and the differential algebra based method (Chis et al., 2011).

Clearly, experimental data in the field of bioprocess engineering is almost never free of noise. Hence, practical identifiability should be scrutinized. Here, noise-corrupted data is used to obtain the unique set of parameters. It is rather straightforward that structural identifiability is a prerequisite for practical identifiability. However, it is no guarantee since the quality and/or quantity of kinetic experimental data is usually restricted because of the available analytical techniques (De Pauw, 2005).

CHAPTER 4

Optimal experimental design for enzyme processes

Despite the effort of the modeller to represent the process being studied as good as possible, there will always be an inherent uncertainty about the model prediction. It is generally accepted that output uncertainty is the combined effect of uncertainty of the measured data, the parameters, the model structure, the used software and the future model inputs. When looking at one mathematical model as such, the latter two sources can be disregarded. Clearly, in order to have a useful model, the uncertainty on the prediction should be as low as possible. Since mathematical models and parameter values are derived from the observations and dedicated experiments, it is desired to design the latter in such a way that the data collected is as information rich as possible. The approach of designing experiments in order to get an accurate model using minimal time and resources is referred to as optimal experimental design (OED) (Dochain and Vanrolleghem, 2001).

4.1 Optimal experimental design for parameter estimation

OED can be used for improving the statistical reliability, i.e. reducing the variance of the estimated parameter values. Such numerical optimisation techniques, generally known as optimal experimental design for parameter estimation (OED/PE), result in well-considered choices of the experimental degrees of freedom in order to get more precise parameter estimates out of the subsequent experiment (Donckels, 2009). Note that precise parameters are generally considered to be of utmost importance for accurate predictions. However, under certain conditions, some models are rather insensitive to some of their parameters and thus imprecisions will not significantly affect the accuracy of the model. This indicates that OED/PE is not required for these parameters (Ataíde and Hitzmann, 2009).

Traditional OED/PE is based on the inverse relationship between the Fisher Information Matrix (FIM) and the parameter estimation error covariance matrix (Ljung, 1999). An overview of different OED/PE examples is given by Franceschini and Macchietto (2008). Regarding enzyme kinetics, most research has been done using fairly simple Michaelis-Menten kinetics (Duggleby

and Clarke, 1991; Murphy et al., 2002; Lindner and Hitzmann, 2006; Ataíde and Hitzmann, 2009). Different experimental degrees of freedom were investigated in these studies (measurement frequency, measurement distribution, initial concentrations, optimal input function, number of replications, ...). One of the main conclusions is that model inputs can significantly improve the precision of the estimated parameters. Since batch experiments do not have any inputs, more sophisticated fed-batch experiments are mandatory for additional precision (Lindner and Hitzmann, 2006). Feeding of the enzyme solution creates new experimental degrees of freedom (enzyme feeding, substrate feeding, optimal input function, ...) leading to more informative data. The suggested approaches by these authors can be extended to more complex enzyme kinetic models. However, this can be particularly difficult and in addition will require more computational power since complex models contain more parameters (Murphy et al., 2002; Lindner and Hitzmann, 2006).

4.2 Optimal experimental design for model discrimination

Another source of model uncertainty is when the lack of insight in the process being studied results in different model structures, i.e. rival models. Since it is desirable to predict the process behaviour in the best way, the best model structure should be used for the prediction. The problem of discriminating between rival models can also be approached by optimal experimental design, i.e. optimal experimental design for model discrimination (OED/MD). Such OED techniques were first proposed by Hunter and Reiner (1965) and are generally based on the idea of designing experiments in such a way that the different model predictions differ maximally (Michalik et al., 2010). However, uncertainty on the measurements and the model predictions can be taken into account as well (Donckels, 2009).

Discriminating between multiple rival models is not solely dependent on the model structure. It is fairly straightforward that the accuracy of the estimated parameters will have a substantial impact on the model selection process. Therefore, the objectives of both OED/PE and OED/MD should be reconciled with each other (Donckels et al., 2010; Alberton et al., 2011).

PART III

MATERIALS AND METHODS

CHAPTER 5

Mathematical modelling and tools

5.1 Model implementation and simulation

A general way to represent continuous-time dynamical systems is (Donckels, 2009):

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t); \quad (5.1)$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t); \quad (5.2)$$

Here, $\mathbf{x} \in \mathbf{S} \subseteq \mathbb{R}^{n_s}$ contains all the time-dependent state variables defining the state of the system. In the case of a bi bi enzyme catalysed reaction, \mathbf{x} represents a vector containing the concentration of all the reagents (a , b , p and q). $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{n_p}$ represents a vector of parameters, i.e. the kinetic parameters (V , K_m and K_i) and $\mathbf{y} \in \mathbf{M} \subseteq \mathbb{R}^{n_m}$ represents a vector of experimentally measurable output variables (NADH for the LDH and GDH reaction, see Section 7.1). It can be seen from Eq. (5.2) that the output variables are just an algebraic transformation of the state variables constituting the system. The experimental setup, $\boldsymbol{\xi}$, is determined by the experimental degrees of freedom such as model inputs, initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$ and the measurement frequency. The dynamic behaviour of the whole system is denoted by t .

5.2 Mathematical parameter optimisation

5.2.1 Defining the objective function

In order to find optimal solutions for discrete or continuous problems, optimisation algorithms are frequently used. Both parameter estimation and optimal experimental design can be regarded as optimisation problems (Donckels, 2009). However, this chapter will only discuss optimisation with respect to parameter estimation, although the main concept can be extended to any optimisable problem.

Parameter estimation is the determination of the optimal values of the parameters, in order to obtain the best fit between the model prediction and the measured data. For non-linear

models this is generally done by minimising a cost or objective function, e.g. the weighted sum of squared errors:

$$J(\boldsymbol{\theta}) = \text{WSSE}(\boldsymbol{\theta}) = \sum_{h=1}^{n_e} \sum_{l=1}^{n_{sp_k}} \Delta \mathbf{y}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)' \cdot \mathbf{Q}(\boldsymbol{\xi}_k, t_l) \cdot \Delta \mathbf{y}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \quad (5.3)$$

with

$$\Delta \mathbf{y}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) = \mathbf{y}_{\text{exp}}(\boldsymbol{\xi}_k, t_l) - \mathbf{y}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \quad (5.4)$$

Here, n_e represents the number of experiments from which data is used, n_{sp_k} represents the number of sampling points in experiment $\boldsymbol{\xi}_k$, \mathbf{y}_{exp} represents the experimentally measured data of experiment $\boldsymbol{\xi}_k$, and \mathbf{Q} is an $n_m \times n_m$ matrix of weighing coefficients. This matrix is not mandatory, but can be used to take measurement errors of the analytic techniques into account. Therefore, \mathbf{Q} is typically chosen as the inverse of the measurement error covariance matrix (Donckels, 2009). In more complex cases, this matrix of weigh coefficients can be time dependent, as expressed in Eq. (5.3). However, in this thesis \mathbf{Q} will usually be chosen as the unity matrix \mathbf{I}_{n_m} . Clearly, if a correct model structure is assumed and the experimental data is free of noise, the objective function will decrease to zero as soon as the values of the parameters reach the true system parameters. Also, when data contains noise, a minimum, i.e. a stationary point or extremum, is expected to be observed in the objective function at the location of the true system parameters.

5.2.2 Modifications of the objective function

Objective function relative to the measured data points

The weighted sum of squared errors (Eq. (5.3)) only takes into account absolute errors. This shortcoming can be circumvented by modifying the objective function as follows:

$$\text{WSSRE}(\boldsymbol{\theta}) = \sum_{l=0}^n \left(\frac{\Delta \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, t)}{\mathbf{y}_{\text{exp}}} \right)' \cdot \mathbf{Q} \cdot \left(\frac{\Delta \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, t)}{\mathbf{y}_{\text{exp}}} \right) \quad (5.5)$$

with WSSRE the weighted sum of squared relative errors.

Objective function relative to the sensitivity functions

As explained in Section 5.4 sensitivity functions contain information on the effect of perturbations of the different model parameters on the model output. It is clear that this information can be used to refine the search for the true system parameters. In this thesis, the following modification of the weighted sum of squared errors is used:

$$\text{WSSE}_{\text{SF}}(\boldsymbol{\theta}) = \sum_{l=0}^n (\Delta \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, t) \cdot \max \text{TRS}(t))' \cdot \mathbf{Q} \cdot (\Delta \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, t) \cdot \max \text{TRS}(t)) \quad (5.6)$$

in which $\max \text{TRS}$ is the maximum time dependent value of all the total relative sensitivity functions (Eq. (5.15)) of the model.

Handling constraints

Since the parameters of a deterministic model usually have a physical meaning, their values ought to be in a physically possible range. Different strategies to fulfill these constraints exist (Scheerlinck, 2012). However, in this thesis only the penalising strategy was considered. Here, the objective function is modified in such a way that whenever the constraint is not fulfilled, the objective function is penalised by an additional term $C(\boldsymbol{\theta})$. Note that additional constraints between enzyme kinetic parameters exist (Section 2.4). Hence these constraints can also be taken into account when evaluating the goodness of the parameter set. A constrained objective function can be expressed as:

$$J_C(\boldsymbol{\theta}) = J(\boldsymbol{\theta}) + C(\boldsymbol{\theta}) \quad (5.7)$$

5.2.3 Metaheuristics

In addition to an objective function, all optimisation algorithms use a metaheuristic, which iteratively tries to optimise the candidate solution, yet at low computational cost (Scheerlinck, 2012). In computer science, different algorithms have been, and still are being developed to do so. However, in this thesis we will only focus on two specific variants: the Nelder and Mead simplex algorithm and the SIMPSA algorithm.

The goal of an optimisation algorithm is to find the minimum of an objective function. The negative gradient of a function, i.e. $-\nabla J(\boldsymbol{\theta})$, is an obvious starting point for optimisation algorithms since this gives the direction where the decrease in the function value is the largest. However, optimisation algorithms based only on the gradient of an objective function are usually found to be sensitive to local optima, i.e. stationary points in which the objective function does not reach its overall minimum. In this thesis, the focus is on gradient-free optimisation algorithms, also known as direct search methods (Swann, 1969). Unlike gradient-based optimisation algorithms, these algorithms sample the objective function in a finite number of points at each iteration and decide which subsequent action should be taken solely based on the function values, without any explicit or implicit derivative calculation. Such metaheuristics are considered to be more robust since the probability of finding the global minimum, i.e. the overall minimum, is higher.

The simplex algorithm of Nelder and Mead

A typical gradient-free optimisation algorithm is the simplex algorithm proposed by Nelder and Mead (1964). To identify the optimum, this algorithm uses a simplex to explore the values of the objective function. A simplex in an n -dimensional space is a polytope that arises when $n + 1$ points (or vertices) are connected (Swann, 1969; Donckels, 2009). A simplex in two dimensions is represents a triangle, while in three dimensions it is a tetrahedron.

The search for the optimum, in this case the minimum of the objective function, starts with defining an initial simplex in the n -dimensional space and evaluating the objective function

at each of the $n + 1$ vertices. Subsequently, the vertices are replaced by other points in the multidimensional space based on the value of the objective function in each vertex. Elementary mathematical operations such as reflection, expansion and contraction of the simplex, allow the algorithm to find a (potentially global) optimum (Figure. 5.1).

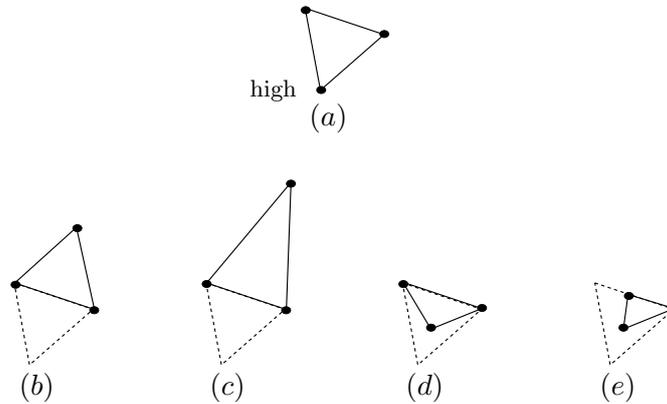


Figure 5.1: Elementary operations in a two-dimensional space used by the Nelder and Mead simplex optimisation algorithm. (a) Simplex at start (lower vertex has highest value for the objective function), (b) reflection, (c) reflection and expansion, (d) contraction, (e) multiple contraction

Although the simplex algorithm defined by Nelder and Mead was introduced as being less prone to local minima, it is still considered to be a local optimisation algorithm since the outcome is still strongly dependent on the initially proposed candidate solution (Vanhaute et al., 2012). Therefore, better alternatives are suggested.

The simplex-simulated annealing algorithm

The SIMPSA algorithm is considered to be a more robust algorithm since it attempts to find the global optimum rather than one of the local optima. As the name indicates, the SIMPSA algorithm is a hybridisation of the simplex algorithm and a method called simulated annealing. The latter is based on the process of annealing in metallurgy in which a solid material is rapidly melted by heating, followed by slow cooling and eventually it is frozen in a state of minimal energy. Hence, the atoms in the crystal lattice can wander through states of higher energy and therefore increase the chance of finding configurations with a lower free energy state. In this way, the size of the crystals in the material increases and at the same time reduces crystal imperfections creating a more ductile material (Cardoso et al., 1996; Donckels, 2009).

This concept of searching for the state of minimal energy by occasionally allowing movements to a state of higher energy, is also used in a class of optimisation algorithms called Metropolis algorithms (Cardoso et al., 1996). It can easily be implemented by allowing missteps in the outcome of some iterations and thus steer the algorithm towards a “wrong-way”, hence, increasing the chance of finding the global optimum of the objective function. In the same way as the physical

process of annealing, the probability of a misstep in a single iteration depends on a Boltzmann distribution with a scale proportional to the temperature (Eq. (5.8)) (Vanhaute et al., 2012).

$$P(A) = \begin{cases} \exp\left(\frac{-(J(\boldsymbol{\theta}_2) - J(\boldsymbol{\theta}_1))}{T}\right), & \text{if } J(\boldsymbol{\theta}_2) > J(\boldsymbol{\theta}_1) \\ 1, & \text{if } J(\boldsymbol{\theta}_2) < J(\boldsymbol{\theta}_1) \end{cases} \quad (5.8)$$

Here, A is defined as the event of accepting a candidate solution $\boldsymbol{\theta}_2$ given the prior solution $\boldsymbol{\theta}_1$. Note that a candidate solution is referred to as inferior if $\Delta J \geq 0$, i.e. an increase in the objective function value. T represents an independent variable which is referred to as the annealing temperature. In analogy with the physical process, the probability of performing a misstep in the search for the optimum decreases as the temperature decreases. Hence, this probability is very high at the beginning of the optimisation, but approaches zero near the end. As a result, the algorithm reduces exactly to the simplex method proposed by Nelder and Mead at the end of the optimisation. The decrease in temperature is determined by the cooling or annealing schedule used in the algorithm. In this thesis, the cooling schedule of Aarts and van Laarhoven (1985) was used, as discussed by Cardoso et al. (1996). An important parameter in such schedules is the cooling rate δ . Smaller values of the cooling rate ($0 < \delta < 1$) lead to slower convergence and therefore increase the chance of finding the global optimum. Note that because of the temperature-dependent probability, the optimisation algorithm is no longer completely deterministic. Hence, it contains a certain degree of randomness. Such probabilistic optimisation algorithms are therefore referred to as random probing or stochastic (Dochain, 2008).

Besides the cooling rate, the SIMPSA algorithm contains several other parameters. An overview is given in Table 5.1. Note that manual tuning for some of the parameters is needed.

Table 5.1: Different parameters of the SIMPSA algorithm itself

Parameter	Default value	Explanation
T_0	1000	Initial temperature: if none provided, an optimal one can be estimated as described by Cardoso et al. (1996).
T_{min}	1	Freezing temperature: when the temperature is below this value, the SIMPSA algorithm is reduced to the simplex method proposed by Nelder and Mead.
δ	10	The cooling rate: the smaller the value, the slower the convergence.
ζ	0.95	The acceptance ratio: needed to calculate the starting temperature. Typically a value of 95% is assigned to ζ (Cardoso et al., 1996).
f_{min}	0.9	Minimal factor for cooling
$K_{T_0,max}$	15	Maximal number of iterations in the preliminary temperature loop
$K_{T_{min},max}$	50	Maximum number of iterations in the last temperature loop
$K_{T,max}$	10	Maximum number of iterations in the remaining temperature loops
K_{max}	2500	Maximum total number of iterations
$FEVAL_{max}$	1×10^4	Maximum number of function evaluations
t_{max}	2500	Maximum CPU time (s)
TOLX	1×10^{-6}	Maximum difference between best and worst function evaluation in the simplex (convergence test 1)
TOLFUN	1×10^{-3}	Maximum difference between the coordinates of the vertices (convergence test 2)

5.3 Identifiability analysis

As described earlier in Section 3.2, identifiability analysis determines whether it is possible to assign unique values to the model parameters. According to Chis et al. (2011), the two main methods for structural identifiability analysis for systems biology models are: the Taylor series expansion and the differential algebra approach. The Taylor series expansion approach is based on the fact that the prediction is a unique observation in time of the used input and system. It is consequently a unique analytic function of time and it is possible to represent the observables by the corresponding Taylor series expansion in the vicinity of the initial state. The uniqueness of this representation will guarantee the structural identifiability of the system. In essence, a system

of nonlinear algebraic equations (AEs) in the parameters is built, based on the calculation of the Taylor series coefficients, and it is checked whether the system has a unique solution. In the differential algebra approach, the system's equations are reordered in order to eliminate the non-observable states in at least one of the system's equations. This is not as straightforward as it sounds and advanced mathematical concepts need to be used like polynomial rings and Gröbner bases. This method will not be described in detail, since only a testing in an existing software package called DAISY was done in this thesis (Saccomani and Audoly, 2010). Hattersley et al. (2011), among others, state that both the Taylor series expansion and the differential algebra approach, can be applied for very simplistic examples, however they are hard to use in more complex, real cases.

Different methods exist to investigate practical identifiability of a mathematical model. Most of these methods are based on information gained from a sensitivity analysis of the model parameters combined with information on the measurement uncertainty (De Pauw, 2005). However, a more direct way of assessing practical identifiability of a model is to create *in silico* noise corrupted data with a certain model structure, experimental design ξ_k and a set of parameters θ^* . This virtual data is afterwards used for retrieving the predefined parameters using parameter optimisation. If a parameter set is identifiable, this results in:

$$\theta^* = \arg \min_{\theta} (\mathbf{y}^* - \mathbf{y}(\theta))' \cdot \mathbf{Q} \cdot (\mathbf{y}^* - \mathbf{y}(\theta)) \quad (5.9)$$

in which θ^* represents the predefined set of parameters used to generate the *in silico* data \mathbf{y}^* (see Section 7.2).

5.4 Sensitivity analysis

Since the outcome of a mathematical model is directly determined by its degrees of freedom, i.e. model inputs, parameters and initial values, it would be very useful to determine the effect of a perturbation in the degrees of freedom on the model output. Such a modelling tool exists and is referred to as sensitivity analysis. The absolute sensitivity of the model output (\mathbf{y}) w.r.t. the model parameters (θ) can mathematically be expressed as:

$$\left. \frac{\partial \mathbf{y}(\mathbf{x}, \theta, \xi, t_l)}{\partial \theta} \right|_{\theta = \hat{\theta}} \quad (5.10)$$

Note that for dynamical models this expression needs to be evaluated at each moment t_l of the simulation, therefore one speaks of a SF. A second important issue is that although Eq. (5.10) can be calculated in every point of the multidimensional parameter space Θ , it does not give an overall picture of the model sensitivity within the whole user defined parameter range. Hence, a distinction is made between local sensitivity analysis (LSA), which determines the sensitivity of the model in a single point $\hat{\theta}$ of the parameter space, and global sensitivity analysis (GSA), which determines the sensitivity of the model in the whole parameter range. No GSA methods

were used throughout this thesis since current OED techniques generally use local SFs. Hence, they will not be explained any further.

Expression (5.10) can be computed analytically as follows (Donckels, 2009):

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial \mathbf{x}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \right) &= \frac{\partial \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} + \frac{\partial \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \mathbf{x}} \Big|_{\hat{\boldsymbol{\theta}}} \cdot \frac{\partial \mathbf{x}(\boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \\ \frac{\partial \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} &= \frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} + \frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \mathbf{x}} \Big|_{\hat{\boldsymbol{\theta}}} \cdot \frac{\partial \mathbf{x}(\boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \end{aligned} \quad (5.11)$$

Sensitivities can also be approached numerically. A very simple and often used method is finite difference analysis whereby the relative difference between the output of a simulation slightly perturbed in one of its parameters and the output of a non-perturbed simulation is calculated. Using the central difference formula, the sensitivity of the model output with respect to parameter θ_j on time t_l is approximated by:

$$\frac{\partial y_i(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \theta_j} \Big|_{\hat{\boldsymbol{\theta}}} = \lim_{\Delta \theta_j \rightarrow 0} \frac{y_i(\mathbf{x}, \boldsymbol{\theta} + \Delta \theta_j, \boldsymbol{\xi}, t_l) - y_i(\mathbf{x}, \boldsymbol{\theta} - \Delta \theta_j, \boldsymbol{\xi}, t_l)}{2 \cdot \Delta \theta_j} \quad (5.12)$$

Here $\Delta \theta_j$ represents an infinitesimally small perturbation of the parameter value θ_j . Choosing a good perturbation is not straightforward, too high values lead to inaccuracies in the approximation (a linear approximation of a non-linearity) whereas too low values jeopardize machine accuracy. Different techniques exist to determine the optimal perturbation (De Pauw and Vanrolleghem, 2006). However, most of the SFs determined in this thesis are analytically calculated, avoiding the issue of finding the optimal perturbation.

Hitherto, only absolute sensitivity functions were discussed, i.e. using the absolute value of both the parameters $\boldsymbol{\theta}$ and the model outputs \mathbf{y} . Comparing SFs derived from different parameters or model outputs requires relative values. Different relative sensitivity functions can be defined:

- Sensitivity relative to parameter θ_j :

$$\frac{\partial \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \theta_j} \cdot \theta_j \quad (5.13)$$

- Sensitivity relative to an output variable y_i :

$$\frac{\partial \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \theta_j} \cdot \frac{1}{y_i} \quad (5.14)$$

- Sensitivity relative to both parameter θ_j and output y_i , i.e. the total relative sensitivity (TRS):

$$\frac{\partial \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\xi}, t_l)}{\partial \theta_j} \cdot \frac{\theta_j}{y_i} \quad (5.15)$$

Sensitivity functions are very valuable for getting a thorough understanding of both the model and the process. For example, it gives the modeller information about correlations between certain parameters. In this context, parameters are called correlated when they have similar or opposite effect on the output variable; another term for this would be parameter interaction. Correlations are of great importance since they complicate accurate parameter estimation. On the other hand, sensitivity functions are particularly interesting when setting up optimal experiments. Thirdly, SFs can also help simplifying complex models since insensitive parameters can, in certain cases, be expelled from the model. Insensitive parameters are these parameters that do not influence variation in the model output significantly.

In the field of enzyme kinetics the term elasticity is frequently used (Cornish-Bowden, 2004). However, this should not be confused with model sensitivity. Elasticity in enzyme kinetics denotes the relative change in reaction rate, defined here as r (e.g. Eq. (2.9)), due to variation in a reagent x_j of that reaction:

$$\frac{\partial r(\mathbf{x}, \boldsymbol{\theta})}{\partial x_j} \cdot \frac{x_j}{r} \quad (5.16)$$

5.5 Subset selection

In the field of statistics and machine learning, subset selection is used as a tool for effective analysis of large data sets in an acceptable period of time. The main question tackled in these cases is simply (Scheerlinck, 2012):

“Is it possible to select a subset of samples or features from a large data set which is as informative as the entire data set?”

In this thesis, selecting optimal subsets of available data is not the main objective. However, selecting an optimal subset of experimental designs from a large set of candidate experiments is of greater importance. Hence the question tackled in our case is the following:

“Is it possible to select a subset of experiments from a large set of proposed experimental designs which is as informative as performing all the proposed experiments?”

Whether subset selection is used for data analysis or experimental design, both have the same objective, i.e. selecting a subset which is as informative as the total set. In other words, the acquired subset contains the same variability as the total set, though over a smaller number of samples or experiments. Note that the concept of subset selection for experimental design is related to practical identifiability of the model since the latter depends on the quality and/or quantity of the experimental data (Section 3.3). Hence, if the total set of experiments and corresponding data allows for a practically identifiable model, subset selection can be used.

In nearly all cases of selecting a subset from a given set, numerous candidates can be defined. Brute force methods calculate all possible subsets and decide afterwards which one is best. This

can only be performed for simple cases since it requires considerable computational power. A more attractive alternative is to reformulate the subset selection problem as a combinatorial optimisation problem, where the objective is to maximise the variance in the subset. Various algorithms are available for such problems, e.g. the Kennard and Stone algorithm, genetic algorithms, ant colony algorithms. . . However, in this thesis a fairly simple, self-implemented, subset selection algorithm was constructed (Algorithm 1).

Algorithm 1: Selecting the optimal subset of experimental designs

Input:

Set of N candidate experimental designs and the corresponding experimentally derived data \mathbf{y}_{exp} ;

True system parameters $\boldsymbol{\theta}_{\text{opt}}$;

The minimal distance d_{min} between the estimated parameters and the true system parameters;

Result:

An optimal subset of experimental designs \mathbf{E}_{opt}

Begin:

Initialise a random or user-defined subset of experimental designs \mathbf{E} , of size $N_{\text{sub}} < N$;

$\hat{\boldsymbol{\theta}}_{\mathbf{E}} = \text{argmin } J(\boldsymbol{\theta})$ using \mathbf{E} and $\mathbf{y}_{\text{exp},\mathbf{E}}$;

Calculate the relative Euclidean distance $d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}}$ between $\hat{\boldsymbol{\theta}}_{\mathbf{E}}$ and $\boldsymbol{\theta}_{\text{opt}}$;

while $d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}} > d_{\text{min}}$ **do**

Determine all the neighbouring experimental designs Nb of a randomly chosen design in the subset \mathbf{E} ;

replace = false;

for $i \leq \#\text{Nb}$ AND replace == false **do**

Replace the chosen experimental design by the i^{th} neighbouring experimental design Nb $_i$, creating a new subset \mathbf{E}_{Nb} ;

$\hat{\boldsymbol{\theta}}_{\text{Nb}} = \text{argmin } J(\boldsymbol{\theta})$ using \mathbf{E}_{Nb} and $\mathbf{y}_{\text{exp},\mathbf{E}_{\text{Nb}}}$;

Calculate the relative Euclidean distance $d_{\hat{\boldsymbol{\theta}}_{\text{Nb}}}$ between $\hat{\boldsymbol{\theta}}_{\text{Nb}}$ and $\boldsymbol{\theta}_{\text{opt}}$;

if $d_{\hat{\boldsymbol{\theta}}_{\text{Nb}}} < d_{\hat{\boldsymbol{\theta}}}$ **then**

$\mathbf{E} \leftarrow \mathbf{E}_{\text{Nb}}$;

$d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}} \leftarrow d_{\hat{\boldsymbol{\theta}}_{\text{Nb}}}$;

replace \leftarrow true;

end

end

end

$\mathbf{E}_{\text{opt}} \leftarrow \mathbf{E}$;

return \mathbf{E}_{opt}

As discussed above, the result of this algorithm is a subset of experimental designs \mathbf{E}_{opt} which are informative for the entire set of experiments.

In words, this algorithm can be explained as follows. First, the inputs need to be defined. One of the inputs is a vector containing all the possible experimental designs along with the experimental data derived from these experiments. A second input is the set of true system parameters $\boldsymbol{\theta}_{opt}$. Note that if real experiments are performed (Section 7.1), the set of true system parameters should be estimated using all the available real experimental data. This is not the case when virtual experiments are used (Section 7.2), since $\boldsymbol{\theta}_{opt}$ in such cases is just $\boldsymbol{\theta}^*$ used to generate the *in silico* data. Finally, a termination criterion for the algorithm is also needed. For this purpose, a minimal distance (d_{min}) between the estimates and the true system parameters must be specified by the user.

In the first step of the algorithm, a random or user-defined subset \mathbf{E} containing a predefined number of experiments N_{sub} is initialised. In the next step, the parameters of the model are estimated using the experiments of this subset \mathbf{E} and their corresponding data. Since the goodness of the subset needs to be defined, the relative Euclidean distance $d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}}$ between the estimated parameter set $\hat{\boldsymbol{\theta}}_{\mathbf{E}}$ and the given set of true system parameters $\boldsymbol{\theta}_{opt}$ is determined:

$$d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}} = \sqrt{\sum_{i=1}^{n_p} \frac{(\hat{\boldsymbol{\theta}}_{\mathbf{E}} - \boldsymbol{\theta}_{opt})^2}{\boldsymbol{\theta}_{opt}}} \quad (5.17)$$

If $d_{\hat{\boldsymbol{\theta}}_{\mathbf{E}}}$ is less than a predefined minimal value d_{min} , the subset is called acceptable. However if this is not the case, a new subset of experiments should be selected in the subsequent steps. In the latter case, the algorithm determines all the neighbouring experimental designs of a randomly chosen experiment in the subset. A neighbouring experiment is defined as an experiment which differs from the reference experiment with a minimal difference in only one of its degrees of freedom, similarly to the Von Neumann neighbourhood in the field of cellular automata (see Figure 5.2).

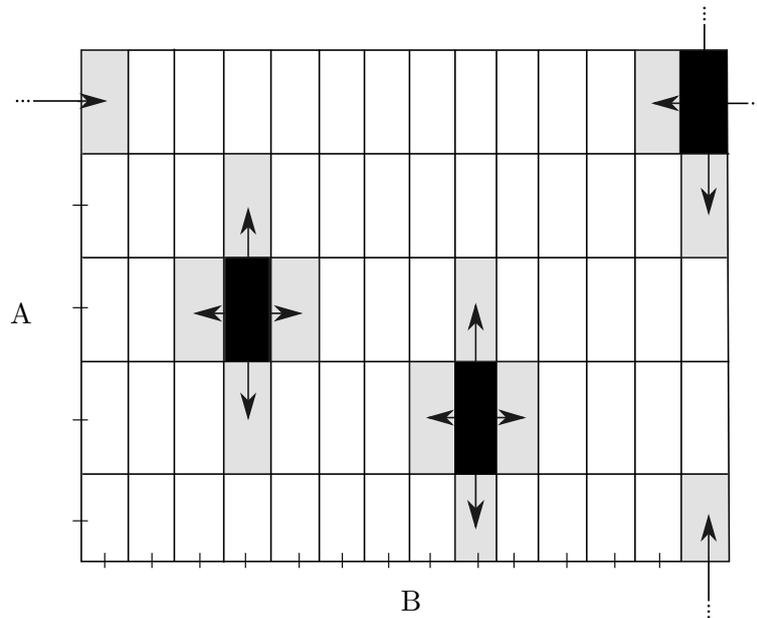


Figure 5.2: The following grid represents the total set of experimental designs. Note that A and B are two degrees of freedom which vary for each experiment in the set. The reference experiment is represented by one of the black boxes, while the neighbouring experiments are shown in grey.

Subsequently, the reference experiment in the subset is replaced by one of its neighbours and thus a new subset of experimental designs \mathbf{E}_{Nb} is defined. Again, the parameters of the model are estimated and compared with the true system parameters, although this time \mathbf{E}_{Nb} and its corresponding data set are used for the estimation. If the parameters $\hat{\theta}_{Nb}$ resulting from the estimation using the newly created subset \mathbf{E}_{Nb} are better than the parameters resulting from the previous subset, i.e. the Euclidean distance to the true system parameter is lower, then the initial subset \mathbf{E} is replaced by the more informative set \mathbf{E}_{Nb} . If this is not the case another neighbour will be selected and the process is repeated. If none of the neighbours results in an improved parameter estimation another reference experiment will be randomly chosen.

This sequence of steps which iteratively tries to optimise the candidate solution could ultimately find a subset of experiments that is as informative as the total set of experiments. However, there is a chance that the algorithm does not find a solution conform the predefined stop criterion. In the latter case the algorithm is stopped if a maximum number of iterations is exceeded.

CHAPTER 6

Software

6.1 Model and experimental design implementation

All mathematical models and most of the mathematical tools used throughout this thesis, were implemented in MATLAB® (The Mathworks, Natick, MA) using a comparable markup language as used in the open-source Systems Biology Toolbox 2 software package (SBTOOLBOX2) (Schmidt and Jirstrand, 2006). In order to simulate different experimental set-ups of the same biological system, two data structures have to be declared. It is rather obvious that one of these structures represents the model, which contains the ordinary differential equations and/or algebraic equations (e.g. Eq. (2.9) for the LDH or GDH reaction), while the other structure represents the experimental design. In Table 6.1 and Table 6.2 an overview of the information in both data structures is given.

To solve the nonlinear ODEs encountered throughout this thesis, the standard MATLAB implemented solver `ode15s` was used. This solver is built to solve stiff initial value problems (IVPs). Stiffness is an important concept in numerical analysis. According to Burden and Faires (2005) stiff differential equations are characterised by having a fast transient solution followed by the steady-state solution which causes certain numerical methods to be unstable. The ODEs considered in this thesis show a comparable dynamic behaviour, i.e. fast transient solution and steady solution.

6.2 Mathematical parameter optimisation

For the progress curve analysis, an existing MATLAB implementation of the SIMPSA algorithm was used (Donckels, 2009). This implementation is partly based on section 10.4 and 10.9 in “Numerical Recipes in C” (Press et al., 1992), and the paper of Cardoso et al. (1996).

A second freely available software package called Encora 1.2 was also used for progress curve analysis. This program, specifically designed for enzyme kinetic model calibration, uses the Nelder and Mead simplex algorithm for estimating the kinetic parameters of biocatalytic models

Table 6.1: Information contained in the data structure representing the mathematical model.

	Example	Explanation
Name	LDH	The name of the model.
Description	-	A brief explanation of the model and what it represents.
States	NADH, Pva, Lact, NAD, Eq. (2.9)	The states \mathbf{x} which represent the state of the model. This also contains the set of ODEs describing the dynamic behaviour of the model.
Variables	r_{LDH}	Variables which are not states nor model outputs and how these are defined.
Constants	-	The different constants of the model and their values.
Parameters	$V_1, V_2, K_{mA},$ $K_{mB}, K_{mP}, K_{mQ},$ $K_{iA}, K_{iB}, K_{iP}, K_{iQ}$	The different parameters of the model and their default values.
Manipulations	-	Manipulations such as spiking and sampling accompanied by the default settings.

Table 6.2: Information contained in the data structure representing an experimental design.

	Example	Explanation
Measurement times	0:1:1800 s	All the moments at which a measurement is performed.
Initial conditions	$NADH_0 = 0.100$ mM $Pva_0 = 0.100$ mM $Lact_0 = 0$ mM $NAD_0 = 0$ mM	The initial conditions of the system.
Manipulations	-	Manipulations such as spiking and sampling accompanied by the default settings.
Measured states	NADH	The states which can be measured during the experiment.

with up to two substrates and products. The differential equations are integrated using a fourth-order Runge-Kutta solver (Straathof, 2001).

For initial rate analysis the standard MATLAB embedded simplex optimisation algorithm was used. Although small modifications with regard to the objective function were made in order to deal with the encountered constrained optimisation problems (Section 5.2.2).

6.3 Identifiability analysis

For the differential algebra method, an implementation in an existing software package called DAISY was tested (Saccomani and Audoly, 2010).

CHAPTER 7

Experimental data

7.1 Real experiments

7.1.1 Experimental set-up

The lab experiments were carried out by The Center for Process Engineering and Technology (PROCESS) at the Technical University of Denmark (DTU). Single enzyme batch experiments of LDH and GDH were performed in plastic cuvettes (1 mL). Here, the change of NADH was measured online, *in situ*, using a spectrophotometer at 30°C (λ_{abs} 340 nm). To stabilise pH around 8, a phosphate buffer was used. Experiments of the coupled ω -TA/LDH/GDH tri-enzyme system were carried out at the same temperature and pH. In this case, off-line measurements of acetophenone, (S)-1-phenylethylamine and glucose were performed. An overview of the different experimental designs is given in Appendix B. Note that only progress curves at varying initial concentrations were supplied. However, by estimating the slope of a linear regression through a predefined number of initial data points, initial rates can easily be deduced from the progress curves.

7.2 Virtual experiments

Generating experimental data by means of lab experiments can be very laborious and time consuming. Hence, sometimes *in silico* data is generated by performing virtual experiments. In other words, data is generated by simulating the system using the mathematical model. The only prerequisites for this are a predefined set of parameters θ , an experimental design ξ and a model structure. Using *in silico* data, it can easily be checked if the parameter optimisation yields the correct values.

PART IV

RESULTS AND DISCUSSION

CHAPTER 8

Progress curve analysis

8.1 Introduction

In Chapter 2 of this thesis it was mentioned that micro-kinetic rate models for biocatalytic conversions, such as Eq. (2.3), had to be transformed into models comprised of macroscopic kinetic parameters. However, the use of kinetic parameters (V, K_m, K_i) instead of rate constants (k_i) is actually not mandatory, it is merely a historical artefact. The fact is that for considerable time, no advanced numerical tools were available for estimating the parameters of dynamic models within a reasonable period of time. Therefore, graphical initial rate methods were vastly used. However, graphical analysis requires the models to be expressed in the form of easily derivable kinetic parameters (i.e. V, K_m, K_i). Nowadays, high-speed computers are available for numerical analysis and thus one is no longer limited to graphical methods. Hence, progress curve analysis can be performed in a reasonable period of time. In contrast to graphical analysis all of the parameters can be estimated simultaneously when using progress curves. Therefore, estimating kinetic parameters or rate constants makes no difference when using the entire time series. However, in this thesis biocatalytic models comprised of macroscopic kinetic parameters were used since these are still the most abundant in the present literature (Zavrel, 2009).

According to Straathof (2001) different enzymatic reactions in a multi-enzyme process should be studied one at a time, i.e. single enzyme kinetics (Section 3.2). In this thesis the focus is mainly on the lactate dehydrogenase (LDH) subsystem, used for the *in situ* removal of pyruvate. This system was selected due to the availability of a large amount of accurate experimental data. However, analysis of the glucose dehydrogenase (GDH) subsystem is very similar since both enzymes have the same reaction mechanism, i.e. a bi bi compulsory-order ternary-complex mechanism. Hence, the methods and techniques applied in this part of the thesis can also be used for analysing the latter subsystem.

Calibrating the rate equation requires experimental data. In what follows a distinction is made between progress curve analysis using data derived from one single experiment (i.e. a single progress curve) and data derived from multiple experiments (i.e. multiple progress curves).

Intuitively, it is expected that calibration of a model using multiple experimental data sets leads to more precise parameters (especially if some of the experimental degrees of freedom are varied). However, more experimental data implies that more experiments have to be performed. In practice this can be quite time- and resource-consuming and might not even be necessary. Therefore, information contained in only one progress curve is examined as well. Note that one progress curve already contains a large number of data points.

8.2 One single enzyme batch reaction

This section deals with the dynamics of the LDH subsystem in a batch reactor. In such an experimental set-up the change of the product concentrations (NAD and lactate) is simply given by the reaction rate r_{LDH} (Eq. (2.9)), whereas the change in substrate concentrations (NADH and pyruvate) is given by the negative value of that reaction rate ($-r_{LDH}$). Note that enzyme deactivation is not included in the model since this is not relevant for the experimental periods studied in this thesis.

8.2.1 Standard approach

In a first step the model describing the dynamic behaviour of the different reagents in the LDH subsystem, i.e. Eq. (2.9), was calibrated using experimental design no. 1 (Table B.1 in the Appendix) and its corresponding data set. Though only the results of experiment no. 1 are shown, the presented conclusions are very similar for the other experimental designs.

The values of the different kinetic parameters are limited to a physically possible range given in Table 8.1. Although the search for the true system parameters is restricted by these boundaries, the continuous range is still enormous. Hence, a constrained SIMPSA algorithm was used for calibrating this model since this optimisation algorithm is rather robust with respect to local optima.

Table 8.1: Lower (LB) and upper bounds (UB) for the different kinetic parameters of the LDH subsystem.

	V_1	V_2	K_{mA}	K_{iA}	K_{mB}	K_{iB}	K_{mP}	K_{iP}	K_{mQ}	K_{iQ}
	[$\mu\text{M/s}$]	[$\mu\text{M/s}$]	[μM]							
LB	0	0	0	0	0	0	0	0	0	0
UB	3×10^7	3×10^7	5×10^6							

As initial parameter guess a random set of parameters was chosen, limited by the lower and upper boundaries. The optimal result is given in Figure 8.1 and Table C.1 in the Appendix.

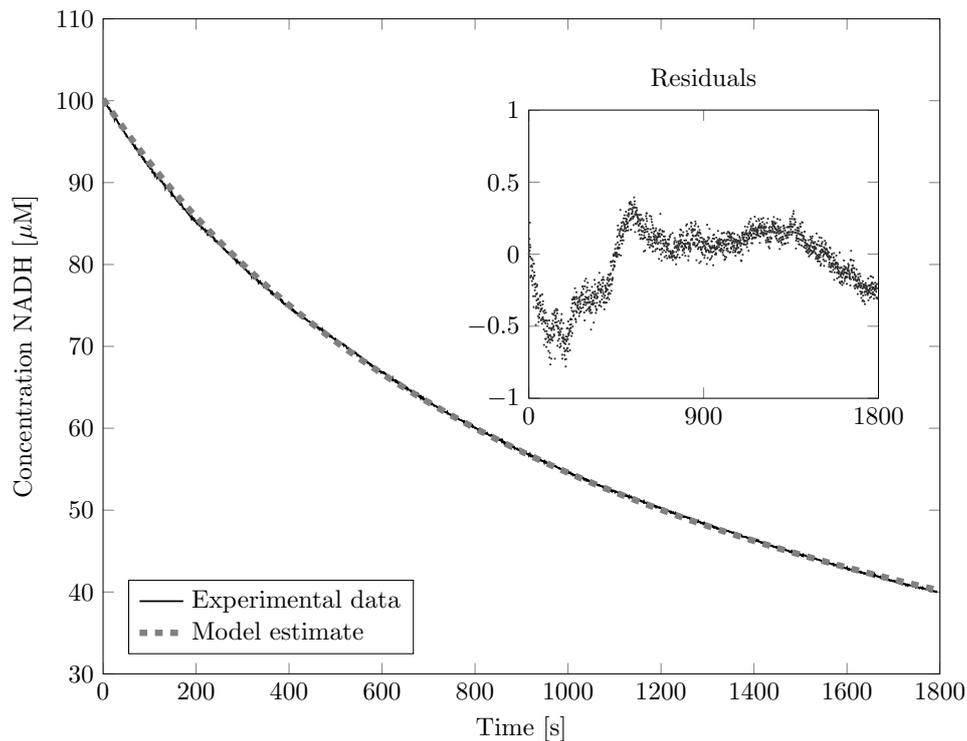


Figure 8.1: Progress curve of NADH concentration during an LDH batch reaction (experimental design no. 1). Experimental data vs. model prediction, note the systematic error of the residuals.

One could state that the model fits the single progress curve reasonably good since the two curves visually cover each other. However, when looking at the residual plot, valuable information on the fit is revealed. Clearly, a systematic error is observed. This means that the residuals of the same variable, i.e. the concentration of NADH, at different points in time are dependent on each other. This is also referred to as autocorrelation. The reason for this kind of correlation can have multiple origins, e.g. errors in the proposed model structure, the inputs or the used parameter set. However, it is most likely that in this case the estimated parameters are causing this systematic scatter since the model inputs and model structure are considered to be correct. Indeed, scientists are quite sure about the enzymatic mechanism, which is represented in the model structure (Borgmann et al., 1975; Cornish-Bowden, 2004).

When performing several similar parameter estimations starting from the same or different initial parameter guesses (IG), another problem arises: no unique set of parameters can be found (Figure 8.3, Table C.2 and C.3). This implies that the objective function has several local minima in which the optimisation algorithm gets stuck.

When looking at the total relative sensitivity functions (Figure 8.2), locally around the first estimated set of parameters, we may conclude that the model output is, in this case, most sensitive for perturbations in V_1 , K_{mB} and K_{iA} . Note that the sensitivity curve of K_{iA} overlaps

with the curve of K_{mB} . When taking a closer look at the SFs of the other parameters, we clearly see an increase in sensitivity of V_2 , K_{mP} and K_{iQ} . Although the SFs of experiments no. 1 and no. 10 look very different, they do have a lot in common.

In the case of experiment no. 1, the equilibrium which favours the product side is not yet reached at the end of the experiment. Hence, the sensitivity of the parameters which are in the forward initial rate model (V_1 , K_{mB} and K_{iA}) are sensitive for a much longer time because there is still a significant amount of substrate which can be converted into product. Unlike the former case, experiment no. 10 reaches the thermodynamic equilibrium much faster. Which translates to SFs of V_1 , K_{mB} and K_{iA} that are only significant in the beginning of the experiment during a short period of time. Looking at the SFs of the parameters represented in the reverse initial rate model (V_2 , K_{mP} and K_{iQ}) the opposite effect is observed, i.e. a slow increase for experiment no. 1, yet much faster for experiment no. 10. Indeed when the experiment progresses, more product is formed and the parameters of the reverse initial rate model become more important.

Variations in the model output are almost exclusively governed by the parameters for which the model is most sensitive. Hence, it is of utmost importance that these parameters are precisely known. However, from the sensitivity functions we can also elucidate that most of the parameters are highly correlated. This is because the extrema of their sensitivity curves occur at the same point in time. This basically means that, if the modeller wants to change the course of the NADH progress curve, this can be realised by changing either of these parameters. Considering that optimisation algorithms depend heavily on the individual effect of the different parameters in the model, this has serious consequences when trying to estimate the kinetic parameters of this model.

In what follows a brief overview is given of all the approaches that were investigated in order to solve the problem of non-uniqueness of the estimated parameters.

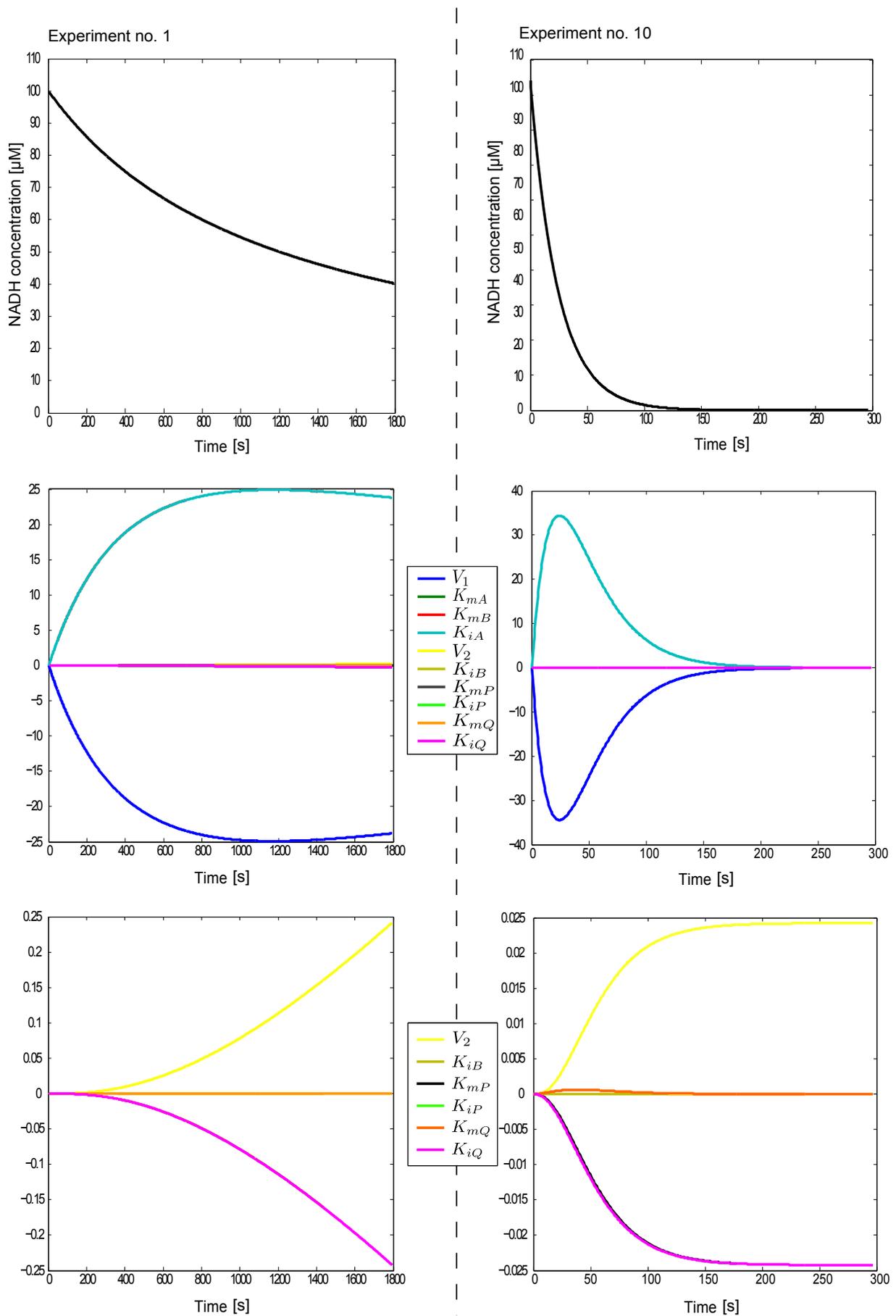


Figure 8.2: Using the parameters from Tabel C.1 the following progress curves and total relative sensitivity functions are obtained for both experiment no. 1 and no. 10.

8.2.2 Improvements of parameter estimation

Manual tuning of the SIMPSA algorithm

Cooling rate and minimal cooling factor: In order to enhance the search for the global optimum, fine-tuning some of the parameters of the SIMPSA algorithm seems an obvious choice. In this thesis most of the parameters of the algorithm were set to their recommended values or were calculated beforehand by the algorithm itself (Cardoso et al., 1996). However according to Vanhaute et al. (2012) and Donckels (2009) the cooling rate δ and the freezing temperature T_0 need to be tuned manually. To avoid the algorithm getting stuck in a local minima, the value for the cooling rate is lowered from 10 to 1 with a constant decrease of 1 (Figure 8.4). It should be remembered that as the temperature decreases, the probability of performing a step in the wrong direction decreases and the SIMPSA algorithm converges to the “local” Nelder and Mead method. Hence adjusting the cooling rate should result in slower, yet steadier convergence (Section 5.2.3). Note that the SIMPSA algorithm used throughout this thesis also contains a parameter called the minimal cooling factor (f_{min}). This latter parameter determines the minimal difference between subsequent temperatures, e.g. if the initial temperature is 1000 the subsequent temperature should at least be as low as 900 for an $f_{min} = 0.9$. Hence, in order to have a slower convergence the minimal cooling factor was raised from 0.9 to 0.99. The result of this modification is again unsatisfactory since there is no improvement in the value of the objective function and no unique solution for the different kinetic parameters was found. This is also the case when lowering the freezing temperature T_0 , below the default value of 1 (Figure 8.3 – Freezing temp. – and Table C.5).

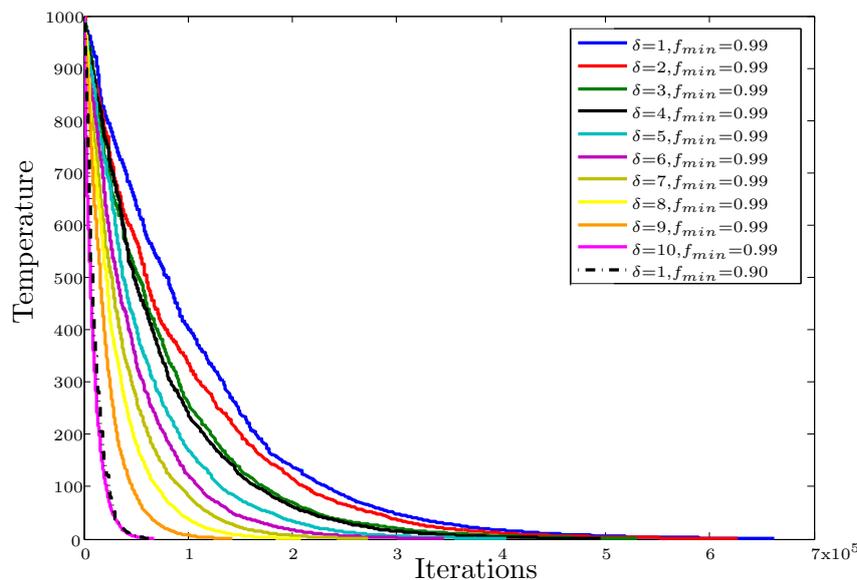


Figure 8.4: The temperature course for different settings of the cooling rate (δ) and minimal cooling factor (f_{min}).

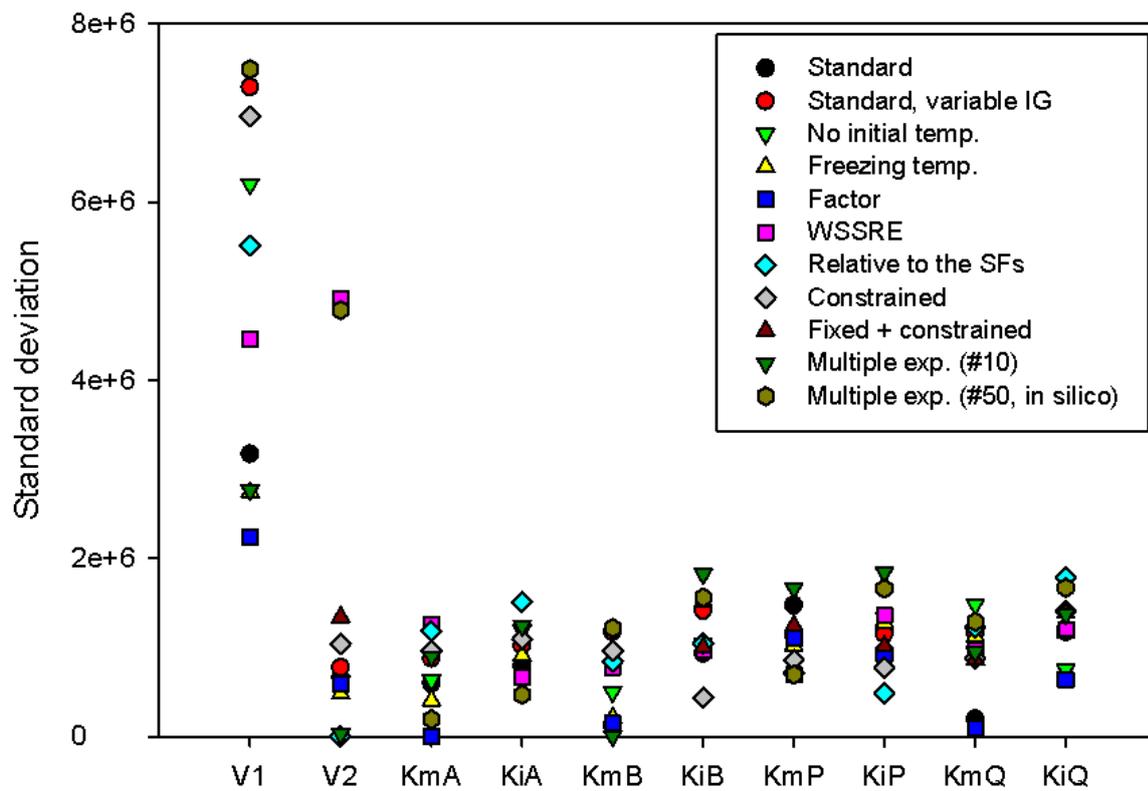


Figure 8.3: In order to estimate the parameters of the LDH model, different approaches were examined. In this figure an overview of the standard deviation between the estimates of different iterations is given. This is done for each of the proposed improvements.

Initial temperature: It was already mentioned in Table 5.1 that the initial temperature of the SIMPSA algorithm does not need to be declared beforehand, since the algorithm itself is capable of estimating this parameter. Therefore, the outcome of an optimisation with and without an arbitrarily defined initial temperature of 1000 was compared. Again, the result after a testing of 20 iterations, starting with a computed optimal temperature, was unsatisfactory (Figure 8.3 – No initial temp. – and Table C.4).

TOLX and TOLFUN: Using two different convergence tests, the SIMPSA optimisation algorithm determines whether an optimum is reached or not. If one of these test criteria is met, the algorithm stops and returns the values of the parameters in this optimum. A rather straightforward way to make both criteria more stringent at the same time, is by multiplying the value of the objective function with a large factor. To test whether this approach leads to a unique set of parameters, a testing of 20 iterations was performed using a multiplication factor of $1e6$. As shown in Figure 8.3, this approach lowers the deviation between the outcome of different iterations, yet it does not generate an exclusive set of parameters (Figure 8.3 – Factor – and Table C.6). Note that besides these two convergence tests, other termination criteria such as K_{max} , $FEVAL_{max}$ and t_{max} exist. However, increasing these termination criteria will not improve the accuracy of the algorithm.

Rescaling the objective function

Looking at the final part of the objective function, it is observed that there is little variation between the different iterations (knowing that the error of the initial guess was about a thousandfold larger). This clearly indicates that the shape of the multidimensional objective function does not show a well distinguishable global minimum. Hitherto, the objective function we used was the WSSE with no real weighting, i.e. \mathbf{Q} is the unity matrix \mathbf{I}_{n_m} constant in time. By changing how the objective function is defined, it is possible to alter its hyperdimensional shape. Note that, as a result of noise, there is a chance that the objective function may not have a global minimum at all.

Units and truncation: A possible way to reshape the objective function is through rescaling of the experimental data. The data obtained from the Danish lab at DTU were originally expressed in milimolar (mM), however, this were adjusted for all of the different approaches to micromolar (μM). The reason why this is done is because one would expect that higher values of the residuals lead to more precise parameters when the tolerance of the optimisation algorithm is kept constant (TOLFUN and TOLX for the SIMPSA algorithm). Sensitivity functions can also be used to rescale the experimental data. For example, by only taking into account the part of the progress curve that is highly sensitive to perturbations in the parameters. Clearly, less sensitive parts are redundant since these have less distinctive power. In this respect, the model output ought to have sections at which it is less sensitive to all of the parameters comprising the model. This is, however, not the case for experiment no. 1 (Figure 8.2).

WSSRE: As mentioned in Section 5.2.2 the weighted sum of squared errors only takes into account absolute errors. In other words, the same absolute errors at higher values of the output variables are of equal importance when compared to the same errors at lower values of the output variables. By using the weighted sum of squared relative errors (WSSRE) this can easily be circumvented. However, this does not improve the search for a unique set of parameters (Figure 8.3 – WSSRE – and Table C.10).

SFs: Parameter sensitivities can also be used for redefining the objective function. Most optimisation algorithms try to optimise a problem by iteratively changing a candidate solution and subsequently evaluating the newly proposed candidate. For continuous optimisation problems, iteratively changing a candidate solution basically means that one or more model parameters are changed each iteration. A change in the parameters can either significantly alter the model output or it could have no influence whatsoever. In the field of mathematical modelling the impact of a small perturbation in one of the parameters can be quantified using sensitivity analysis (Section 5.4). Local sensitivity functions reveal how the output of the model is affected by a slight change in one of the parameters. Redefining the objective function, with respect to parameter sensitivities, can be done by weighting the residuals according to the maximum value of all the total relative sensitivity functions. In this way, the residuals become more important if the model output is more sensitive to a parameter, resulting in a time dependent weighting. This means that the differences between model output and experimental data needs to be lower in these regions of the progress curve. Hence, it is expected that the algorithm will be able to make a better distinction between different sets of parameters. Note that only for *in silico* data and thus with known real parameter value, the local sensitivity functions can be calculated beforehand. Therefore, when using this modification for real experimental data, either some prior knowledge of the parameters should be available e.g. parameter values from a prior optimisation or an updated local sensitivity function at every iteration should be included. In this thesis the former approach, i.e. taking the maximum of the local relative SFs into account when evaluating the objective function, was used to refine the search for the global optimum. However, no satisfactory results were obtained (Figure 8.3 – Relative to the SFs – and Table C.9).

Constraints: Additional restrictions on the values of the kinetic parameters can also help to redefine the shape of the objective function. In Section 2.4 it was shown that the kinetic parameters of both dehydrogenase reactions must fulfill two constraints, i.e. the Haldane and the non-Haldane relationship. Taking into account such constraints is practically done by implementing a penalty function $C(\boldsymbol{\theta})$ (Section 5.2.2). In this thesis the following penalty function was arbitrarily chosen:

$$C(\boldsymbol{\theta}) = 0.5 \cdot \Delta H \cdot J(\boldsymbol{\theta}) + 0.5 \cdot \Delta nH \cdot J(\boldsymbol{\theta}) \quad (8.1)$$

in which,

$$\Delta H = \left(\frac{V_1}{V_2}\right) \frac{K_{mP}K_{iQ}}{K_{iA}K_{mB}} - \left(\frac{V_1}{V_2}\right)^2 \frac{K_{iP}K_{mQ}}{K_{mA}K_{iB}} \quad (8.2)$$

and

$$\Delta nH = \frac{K_{mA}}{K_{iA}} - \left[1 + \frac{V_1}{V_2} \left(1 - \frac{K_{mQ}}{K_{iQ}}\right) - \frac{K_{mP}K_{iQ}}{K_{iP}K_{mQ}}\right] \quad (8.3)$$

The result is given in Figure 8.3 – Constrained – and Table C.7. Again, no remarkable improvement in the result is observed.

Finally, besides the MATLAB implementation of the LDH model and the SIMPSA optimisation algorithm, a software package called Encora 1.2 (specifically designed for progress curve analysis of biocatalytic models) was also used. Research has shown that among computer programs, which use dynamic parameter estimation, large differences can occur in the result (Zavrel et al., 2010). Although Encora 1.2 is not the most recommended software for analysis of progress curves, it was nonetheless chosen since it is freely available (Straathof, 2001). In a first step, noise free *in silico* data for a single LDH conversion was generated using an arbitrarily chosen set of predefined parameters and initial conditions. Subsequently, the Encora 1.2 software was used to re-identify the ten kinetic parameters. Note that this software always takes into account the additional Haldane and non-Haldane constraints. However, also in this case the result was unsuccessful since the program was not able to find the true system parameters nor did it find a unique set of parameters starting from different initial guesses. Therefore a clear message to the community: it is not because you have a visually acceptable fit and a set of optimised parameters using specific software, that these parameters represent the true system parameters. It is consequently dangerous to use such set of parameters for extrapolation. Hence, the MATLAB implementation of the model and optimisation algorithm was used for the further course of this thesis.

Fixing parameters of the LDH model

Constants: Although the underlying mechanism of dehydrogenase enzymes are well known, little knowledge on the values of the kinetic parameters of the LDH system has been reported to date. Borgmann et al. (1975) were one of the first to determine the eight rate constants (k_i) of different LDH isoenzymes at varying temperatures using graphical initial rate analysis. In order to improve our knowledge of LDH kinetics, the values of the parameter in the initial rate model (Table C.12) were considered as constants. Which means that the number of parameters to be estimated is almost halved, i.e. from ten to six parameters. Note that these literature parameters are not valid for the experimental data available to us. Hence, *in silico* data had to be used. A prerequisite for such virtual data is a predefined set of parameters. Logically, the values of V_1 , K_{mA} , K_{mB} and K_{iA} were taken from the literature and the remaining six parameters had to be arbitrarily chosen. For this purpose, the values of the remaining six parameters were chosen

so that the additional Haldane and non-Haldane constraints were satisfied (Table C.13). The result is given in Figure 8.3 – Fixed + Constrained – and Table C.8.

8.3 Impact of the experimental design

Estimating the parameters of complex nonlinear biocatalytic models from one single reaction batch experiment was proven to be too ambitious. This is mainly because such experiments do not provide a wide variety of information in the measured data. Hence, strategies that increase the information content of an experiment were investigated in the next steps.

8.3.1 Optimal experimental design for parameter estimation

OED: Increasing the information content of experiments is vastly associated with optimal experimental design (OED). It was already mentioned in Section 4.1 that OED/PE is generally based on the inverse of the Fisher Information Matrix (FIM). Since, under the condition that the measurement errors and the residuals are white and uncorrelated and the model is linear, the FIM is equal to the lower bound of the parameter estimation covariance matrix. However, OED/PE can not be used for our case because there are some restrictions to the LDH and GDH systems. Firstly, calculation of the FIM requires information about the local sensitivity of the output variable around the optimal parameter values. Since we do not find a unique value for each of the parameters in the model using the SIMPSA algorithm, calculation of a unique FIM is not possible. Indeed, for different iterations the search algorithm ends in completely different regions of the parameter space. Secondly, the condition of white and uncorrelated residuals and a linear model is clearly not met for the system we are evaluating (Section 8.2) (De Pauw, 2005). We must thus conclude that it is not possible to use OED/PE directly as a tool for increasing the information contents for the LDH and the GDH subsystems.

8.3.2 Multiple experimental designs

Multiple experiments: Although OED can not directly be used for the different subsystems in this thesis, alternative methods are available to increase the information content of experimental data. A rather straightforward way of doing this is by just increasing the amount of available data. In this respect, two possibilities are available. Either more measurements need to be performed during a single experiment or more different experiments should be performed. In the former approach one could measure additional states or measure the same states but with a higher frequency.

In some cases, multiple states of a system can be measured. Hence, it is possible that some of the parameters in the model have different effects on different output variables. Indeed, this favours parameter estimation. According to the previous argument, we should somehow try to

measure more than one state of the system. However, care must be taken since this is only valid for certain cases. The system focused on is unfortunately not such a case, because although the LDH subsystem has four potentially measurable states (i.e. the concentration of NADH, pyruvate, lactate and NAD), these states do not contain different information. The fact is that the states of a system consisting of just a single batch reaction are merely transformations of one another. Hence, we use the same equations (Eq. (2.9)) to model the different states of the LDH system. Therefore, other strategies should be explored.

Since no additional information is obtained when measuring additional states and because of the high frequency (i.e. 1/60 to 1/30 Hz) of the measured data, measuring more of a single experiment will not significantly improve the information content of the data. Hence, using data from multiple experiments is the most attractive alternative.

In the same way as for the single experiment, a parameter estimation was done using experimental data obtained from multiple lab experiments, i.e. experiments no. 1 to 10 (Table B.1). Again, unsatisfactory results were obtained (Figure 8.3 – Multiple exp. – and Table C.11). Under the assumption that ten experiments still might be too few in order to determine the parameters of the considered LDH subsystem, even more *in silico* data were generated to confirm the previous conclusion. In this respect, different degrees of freedom were varied. Looking at the available experimental designs for both progress curves and initial rate experiments (Table B.1 and B.2), one can deduce six degrees of freedom in each experiment: the duration of an experiment, the measurement frequency and the initial concentrations of each of the four components. Considering the same possible ranges, it can easily be shown that there are 12,800 possible experimental set-ups. Although we can virtually perform this much experiments, in reality this is practically not possible without high-throughput techniques. Since in our case all the experiments are performed manually, we should limit the amount of virtual experiments. Therefore, a testing was done using only 50 arbitrarily chosen experimental designs and data sets (using the parameters in Table C.12 and C.13). However, each of the 20 repetitions took into account multiple adjustments, i.e. $\delta = 1$, $f_{min} = 0.99$, no user-defined initial temperature, a multiplication factor of $1e6$, the WSSRE objective function and the additional Haldane and non-Haldane constraints. The result for parameter V_1 is given in Figure 8.5. Even with the adjustments and more data sets, the SIMPSA optimisation algorithm is still not able to find the true parameter values nor a unique set of parameters.

8.4 Structural identifiability analysis

Since none of the aforementioned approaches leads to a unique set of parameters, we can conclude that progress curve analysis is not an option for accurate and precise estimation of all the kinetic parameters in the LDH model. It was already mentioned in Section 3.2 that nonlinear biocatalytic models which contain a large number of parameters, are often found practically

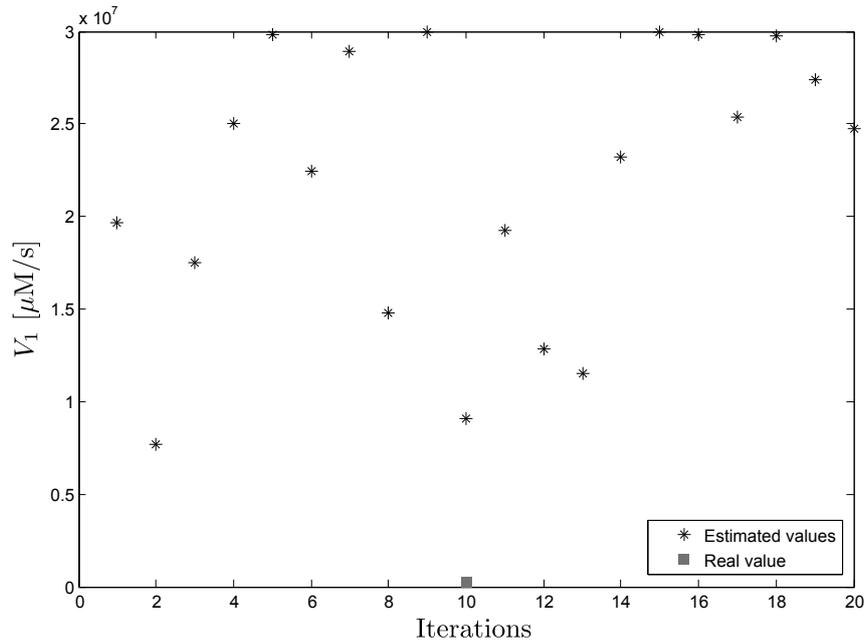


Figure 8.5: The estimated value of V_1 for different iterations taking into account multiple *in silico* generated datasets. Note that the result of different iterations is scattered over the whole parameter space.

and/or structurally unidentifiable, i.e. no unique value can be assessed to the parameters comprising the model (Chis et al., 2011). The approaches examined in the previous section clearly focused on practical identifiability of the LDH model. In this respect, we must conclude that the model describing the kinetics of lactate dehydrogenase, i.e. Eq. (2.9) is practically unidentifiable using the different approaches suggested by us. According to Chis et al. (2011) lack of practical identifiability is in general terms solvable if the experimental constraints allow for designing sufficiently rich experiments. Obviously, the crux of the problem is that the experiments in our case do not allow for such designs.

Considering the previous conclusion, a more appropriate question would thus be “Is the LDH model structurally identifiable?”. Different methods for testing the structural identifiability of a mathematical model exist. In this thesis, a software package called DAISY was used to examine structural identifiability by means of a differential algebra based method. For this purpose, we considered a batch LDH experiment in which only one state was measured. However, when testing the structural identifiability using DAISY software, computational problems were encountered due to the complexity of the rate equation. This was expected from the difficulties in determining the structural identifiability of large and complex models, mentioned in literature (Chis et al., 2011; Hattersley et al., 2011; Meshkat et al., 2011). However, when neglecting the reverse reactions and when only including the inhibiting effect of the products in the equations, some analyses could be performed. Unfortunately, the successful analyses revealed unidentifiable models, i.e. an over-specified model. This was not only the result when having only one

measurement and no inputs (i.e. a batch experiment), but also when having two measurements in a batch experiment. Note that it was not possible to finish an analysis that included system inputs. So, using DAISY, no conclusions could be drawn for a system that has the freedom of giving pulses.

A simple solution for unidentifiable models is reducing the complexity of the model. For example, by taking into account parameter sensitivities, the mathematical model can be simplified by removing the parameters for which it is considered insensitive (Santacoloma, 2012). From the modeller's perspective, this could definitely improve parameter calibration. However, care must be taken, since such an approach leads to empirical input-output models with no physical meaning whatsoever. Furthermore, scientists in the field of biocatalysis are most familiar with the kinetic models that use the notation proposed by Cleland (1963). One of the advantages of these models is that they allow for an easy comparison of enzyme activities. For example, the ratio k_{kat}/K_m , referred to as the specificity constant, is a simple measure for the intrinsic efficiency of an enzyme. However, if the model structure changes, these measures will no longer be valid since the parameters are lumped together.

CHAPTER 9

Initial rate analysis

9.1 A systematic approach for estimating the kinetic parameters of biocatalytic reactions

Thus far, it was not possible to estimate the kinetic parameters in Eq. (2.9) simultaneously. Therefore, one could ask whether it is possible to decompose the present model in such a way that the parameters of the complete model can be determined in a systematic manner. Using graphical analysis for deriving the kinetic parameters in a stepwise procedure is well documented in literature (Cornish-Bowden, 2004). However, when using numerical methods, such incremental approaches were not available until recently (Chen et al., 2008). Al-Haque et al. (2012) tackled this shortcoming intelligently by exploiting the best features of several of the current approaches. They propose a robust methodology allowing systematic calibration of complex enzymatic models. Four hierarchical steps, in which the output of each step is the input of the subsequent step, have to be performed. Clearly, this methodology can be used in this thesis.

In the first step of the aforementioned methodology, the region in which the initial reaction rate changes linearly with a change in enzyme concentration is defined, because beyond this region mass transfer limitations influence the overall reaction rate. Note that in this thesis we are always working in this linear region. Next, the full rate equation, i.e. Eq. (2.9) for the LDH subsystem, is decomposed into the initial rate equations for both the forward and the reverse reaction, i.e. Eqs. (9.1) and (9.2). This is done by eliminating all the terms in the rate expression which contain variables representing product concentrations. Indeed, this is a valid approach since experiments that contain only substrates and enzyme initially, have a negligible concentration of products during the initial part of the conversion. Using initial rate data derived from different experiments, the four parameters in each of the two equations can now be estimated utilizing algebraic parameter estimation (V_1 , K_{mA} , K_{mB} and K_{iA} for the forward reaction; V_2 , K_{mQ} , K_{mP} and K_{iQ} for the reverse reaction). Note that Eqs. (9.1) and Eq. (9.2) are nearly identical since they have the same model structure but different variables. This results in following similarities: $V_1 \leftrightarrow V_2$, $K_{mA} \leftrightarrow K_{mQ}$, $K_{mB} \leftrightarrow K_{mP}$, $K_{iA} \leftrightarrow K_{iQ}$, $a_0 \leftrightarrow q_0$

and $b_0 \leftrightarrow p_0$. In fact, this is the reason why the sensitivity functions at the beginning and near the end of the conversion show such a comparable course (Section 8.2.1).

$$r_{init,f} = \frac{V_1 a_0 b_0}{K_{iA}K_{mB} + K_{mB} a_0 + K_{mA} b_0 + a_0 b_0} \quad (9.1)$$

$$r_{init,r} = \frac{V_2 q_0 p_0}{K_{iQ}K_{mP} + K_{mP} q_0 + K_{mQ} p_0 + q_0 b_0} \quad (9.2)$$

In step three of the systematic approach, the additional core inhibition parameters (K_{iB} and K_{iP}) are regressed using entire progress curves. However, the values of the parameters estimated in the previous step are considered fixed. Following this step, the set of parameters obtained in the previous two steps is used as initial guess for a final optimisation. Note that the different methods to improve the precision of the estimated parameters as suggested in Section 8.2.2, can be used both steps three and four. Another remark is that the original case study of Al-Haque et al. (2012) does not indicate whether a global or a local optimisation algorithm was used in the last step of the calibration. However, it is believed that a local optimisation method was used since in global optimisation the influence of the initial parameter guess is much more limited.

9.2 Structural identifiability of the initial rate model

Unlike progress curve analysis, the approach suggested by Al-Haque et al. (2012) does not require a large number of parameters to be estimated at once. Intuitively, we would expect that this systematic approach allows for more precise parameter identification. Structural identifiability analysis of the initial rate model (Eqs. (9.1) and (9.2)) was conducted in the same way as for the complete rate equation, i.e. assuming the measurement of only one variable during a batch experiment. The DAISY software showed global identifiability for the initial rate model, thus a unique set of parameters should be identifiable using noise free data and an appropriate algorithm. Note that both initial rate equations have the same model structure, therefore they have similar structural identifiability properties. Henceforward, we will only focus on the forward initial rate expression.

Though the initial rate models are structurally identifiable, this does not imply that the SIMPSA algorithm at its original settings will be able to identify a unique set of parameters starting from noise free data. Therefore, a testing using noise free *in silico* data was carried out. Analysing the available initial rate data obtained from the PROCESS department at DTU (Table B.2) reveals that there are two degrees of freedom that vary between the initial rate experiments, i.e. the initial concentration of respectively NADH (a_0) and pyruvic acid (b_0). Four different initial concentrations of NADH were used in the experiments, whereas for the initial concentration of pyruvic acid (Pva), thirteen different concentrations were used. Hence, a total of 52 initial rate

experiments were performed in the lab. In the virtual experimental data set, an initial concentration of $400 \mu\text{M}$ was included for both components. Hence, 70 different virtual experimental designs are available. Visually this is represented in Figure 5.2. For each of the 70 experiments, noise free *in silico* data was virtually generated using the literature parameters (Table C.12). This is graphically represented in Figure 9.1. Note that the initial reaction rate increases as the initial concentration of both components increases. Since the concentration of enzyme is constant, saturation occurs at elevated substrate concentrations.

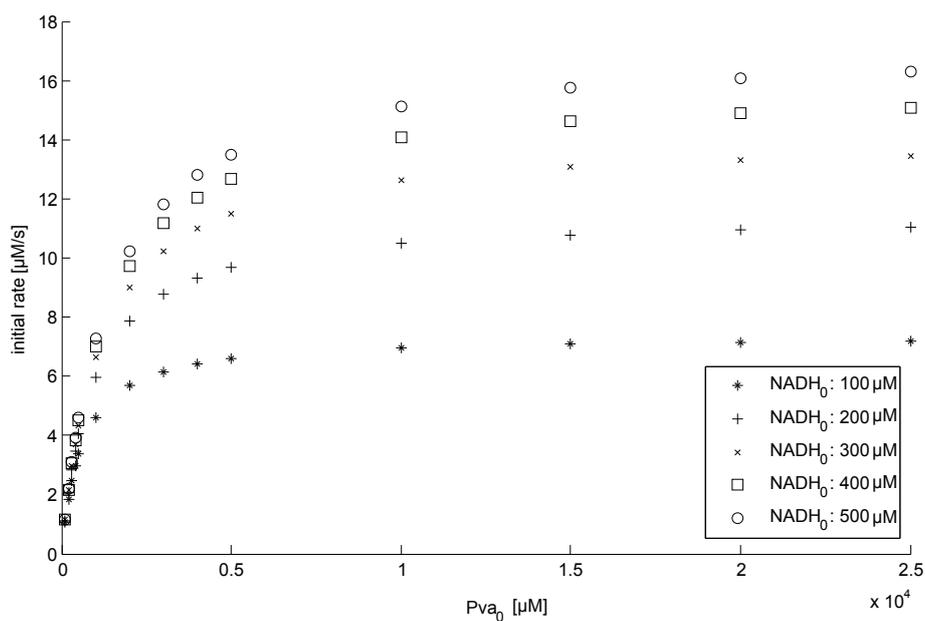


Figure 9.1: The initial reaction rate ($r_{init,f}$) as a function of the initial concentration of Pva and NADH.

Subsequently, this dataset is used to re-identify the literature parameters. In this respect, the more robust SIMPSA optimisation algorithm is compared with the Nelder and Mead simplex method. Maintaining the default settings and considering a WSSE objective function, the SIMPSA algorithm performed much better w.r.t. the optimal parameter values, than the local simplex algorithm when considering the upper and lower boundaries in Table 8.1. Nevertheless, the performance of the local optimisation algorithm can be increased by increasing the maximum number of function evaluations and iterations, as well as lowering the termination tolerance. Since the default settings of the SIMPSA algorithm did not allow for fast convergence, manual tuning is needed (Table 9.1).

Table 9.1: Settings used for the SIMPSA and simplex algorithm in order to accelerate estimation of the initial rate parameters.

SIMPSA		simplex	
Parameter	Value	Parameter	Value
T_0	-	$K_{T_0,max}$	15
T_{min}	10^{-6}	$K_{T_{min},max}$	50
δ	1	$K_{T,max}$	2000
ζ	0.95	K_{max}	10^6
f_{min}	10^{-6}	FEVAL _{max}	10^8
t_{max}	2500	TOLX	10^{-6}
Factor	10^6	TOLFUN	10^{-6}

For both the tuned simplex and tuned SIMPSA algorithm, the result after 50 different optimisations is given in Figure 9.2. The first 25 optimisations all had an initial guess in the entire parameter range — note the logarithmic scale of the vertical axis — whereas the range for the last 25 iterations was a thousandfold smaller. Several things can be deduced from this figure. Firstly, it seems that the SIMPSA algorithm performs better when compared to the simplex algorithm. Indeed, the values of the four parameters after SIMPSA optimisation are almost always around the optimal value, i.e. $\theta/\theta_{opt} = 1$, while this is not the case for the Nelder and Mead simplex algorithm. This clearly indicates the presence of different local minima. Hence, a global optimisation algorithm is needed for this model. However, if we ensure that the initial guess is closer to the optimal value (the last 25 optimisations), the simplex algorithm also finds the optimal parameter values, i.e. the literature values. Another remarkable observation is that according to this figure, both optimisation algorithms are better in predicting the values of parameter V_1 and K_{mA} . To validate this observation, a plot of the objective function near the optimal parameter values was made (Figure 9.3). Clearly, the curvature of the objective function is the steepest for V_1 , which means that optimisation algorithms will be able to estimate this parameter quite accurately and precisely. This is, however, not the case for parameter K_{iA} since the minimum is less distinguishable.

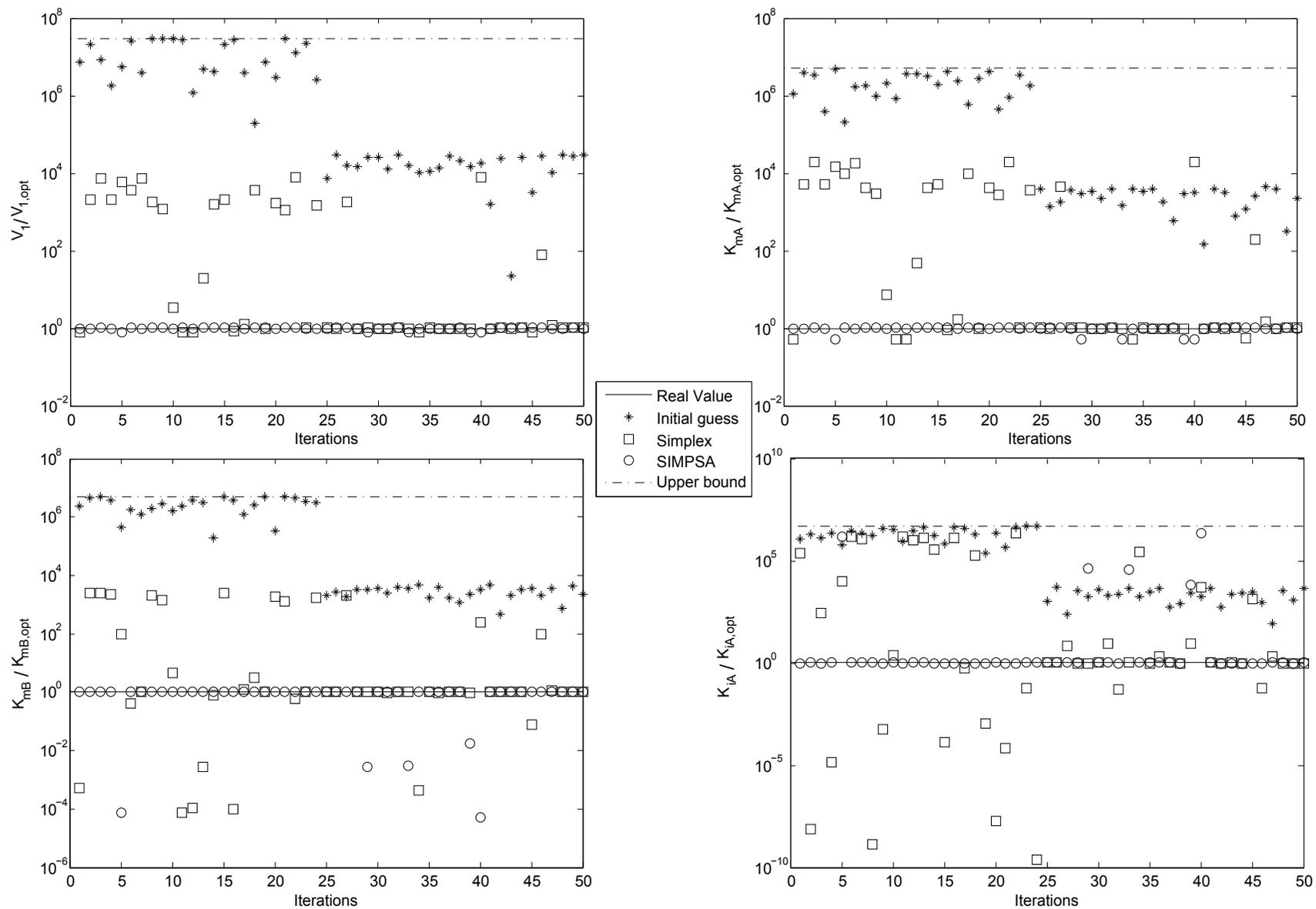


Figure 9.2: Comparison between the estimates of both the tuned simplex algorithm and the tuned SIMPSA algorithm. Note that the boundaries of the last 25 iterations are a thousandfold smaller than the first 25 iterations.

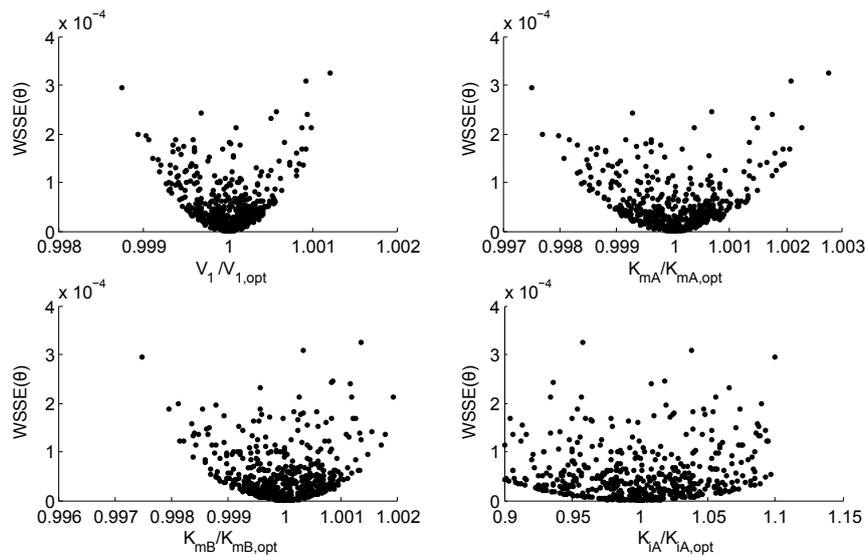


Figure 9.3: Visualisation of the WSSE objective function near the optimal parameter values. Each mark represents a point in the parameter space for which the value of the objective function was calculated.

For parameters K_{mA} and K_{mB} an intermediate curvature is observed. Looking at the local relative sensitivity functions around the literature parameter values (Figure 9.4), the behaviour of the objective function can easily be explained. It seems that the initial rate model is most sensitive for perturbations in V_1 . Hence, the value of the objective function increases rapidly when diverging from the optimal parameter values. On the other hand, the initial rate model is relatively insensitive for changes in parameter K_{iA} , which clearly explains the stagnant behaviour of the objective function near its optimal value.

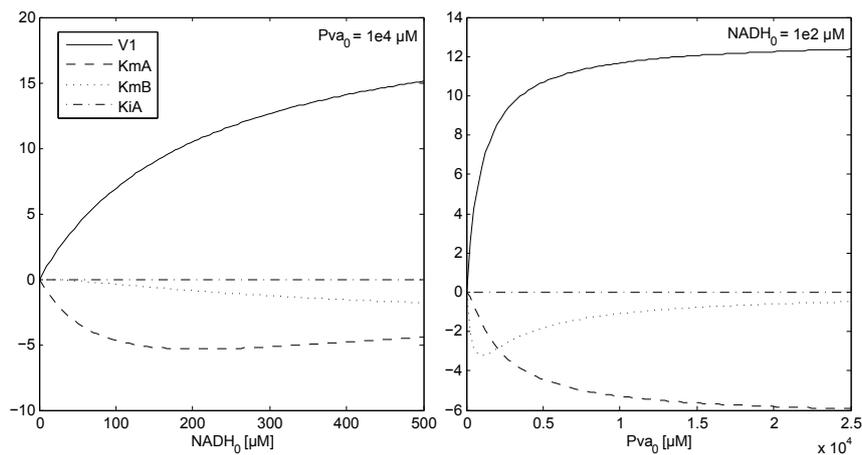


Figure 9.4: The local relative SFs of the initial rate model. Note that the model has two independent variables (NADH_0 and Pva_0), which explains the necessity for multiple plots.

9.3 Practical identifiability of the initial rate model

It should be noted that experimental data in the field of bioprocess engineering is almost never free of noise. Hence, noise-corrupted data is used to estimate the kinetic parameters. Second, initial rate analysis is based on linear regression through a (small) set of consecutive initial data points. To assess the influence of both features, an additional testing was done using *in silico* generated data. Obviously, real experimental data can also be used for this purpose. However, we chose not to, since the true system parameters are only known in the case of virtual data. In this respect, the initial reaction rate had to be disturbed with noise. The following steps were performed for each of the 70 experimental designs.

First, the initial rate model and literature parameters (Table C.12) were used to simulate the first seconds of the progress curve. Indeed, the use of the initial rate model is valid since the concentration of product ought to be negligible in the initial period of time. In order to mimic the measurement noise, Gaussian noise, $\mathcal{N}(0, \sigma)$, was added to these virtually generated data points. To do so, different values for the standard deviation σ were used:

- $\sigma = 0.0346 \mu\text{M}$: The standard deviation of 100 data points in the visually stagnant phase of experiment no. 10.
- $\sigma = 0.122 \mu\text{M}$: The standard deviation of the residuals when regressing a straight line through the first five seconds of the data obtained in experiment no. 10.
- $\sigma = 1.227 \mu\text{M}$: Equals to 5% of the average measured concentration in experiment no. 6. This approach was mentioned by Al-Haque et al. (2012).
- $\sigma = 1 / 5 / 10 \mu\text{M}$: Arbitrarily chosen values for the standard deviation.

Note that in this thesis σ was assumed to be identical at each time instant; this is also the case in other studies (Al-Haque et al., 2012; Zavrel et al., 2010). Finally, a straight line was regressed through the first 10 data points of the disturbed progress curve. This corresponds to 4.5 s of the conversion. Next, the slope of this line was set to be the noise-corrupted initial reaction rate. According to Zavrel et al. (2010), usually around 10% of the conversion is taken into account when deriving the initial reaction rate using graphical analysis. For our case, the fastest experiments reach this point around 5 to 6 s. Hence, by taking into account the first 10 data points our method is more or less in agreement with other studies.

The aforementioned method was repeated 50 times for each value of the standard deviation. Thus, 50 different data sets, each containing 70 disturbed initial rates, were available for analysing the effect of a certain degree of measurement noise. To do this, parameter estimations were conducted using only one data set at a time. The result can be seen in Figure 9.5(a). On the vertical axis the standard deviation (relative to θ_{opt}) between the different results of each of the

50 estimations is given. The outcome of this testing is quite straightforward, i.e. as the standard deviation on the measurements (σ) increases, the results after different optimisations deviate more from one another. This means that practical identifiability can no longer be guaranteed beyond a certain degree of noise. Remarkably, the deviation is not the same for each parameter. It seems that the value of parameter K_{mB} is highly influenced by measurement noise. Thus, this parameter can not be accurately estimated at high levels of noise using the SIMPSA algorithm. This is, however, not the case for V_1 . Hence, this parameter can be estimated with sufficient accuracy and precision. In practice, this plot can be used as a valuable tool to evaluate the precision of the estimated parameter values.

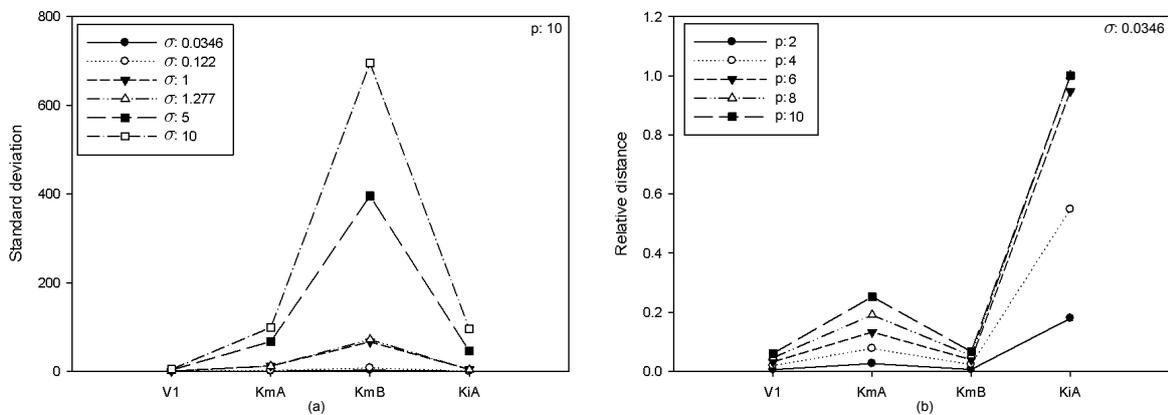


Figure 9.5: The influence of measurement noise (a) and the number of consecutive data points used for deriving the initial reaction rate (b). Note that the lines connecting the marks are only drawn to simplify the use of this figure. They should thus not be interpreted as continuous data.

Another important issue that was investigated is the influence of the number of consecutive data points used for linear regression on the estimates. In Figure 9.5(b) this is clearly visualised. Note that on the vertical axis, the relative Euclidean distance to the optimal parameter value is given (Eq. (5.17)). In contrast to Figure 9.5(a) which focused on the accuracy of the estimates, Figure 9.5(b) clearly focuses on the precision of the estimates. Obviously, the more data points (p) used in the regression, the larger the distance. This is because the initial reaction rate is underestimated to a larger extent. Again, not all parameters are equally influenced by this feature. It seems that the relative distance for the estimates of parameter K_{iA} is largest. However, we have already shown in Figure 9.4 that the model is rather insensitive to changes in this parameter. Hence, the effect on the model output will be fairly small. It is most likely that a change in parameter K_{mA} has a larger impact on the model output. Once again, these Figures clearly indicate that there are some serious issues which should be taken into account when fitting the initial rate model to data.

9.4 Subset selection for experimental design

It was clearly mentioned in Section 9.1 that, in order to estimate the true system parameters with sufficiently high precision, the use of a stepwise incremental method is compulsory. However, the problem with such robust methods is that they require lots of different data sets to calibrate and validate the complete model. Therefore, this can be quite laborious and costly, especially when fine chemicals such as pyridine nucleotide cofactors are needed (Serrano Briega, 2011). For example, the method proposed by Al-Haque et al. (2012) uses six different data sets to calibrate and validate biocatalytic models of the same complexity. Each of these sets is comprised of multiple experiments. The most reagent demanding steps in this method are these which need initial rate data (step no. 1 and 2). Indeed, when performing initial rate analysis each experiment delivers only a single data point. In this respect, obtaining 70 or even 30 (Al-Haque et al., 2012) initial rates for both the forward and the reverse reaction, is not considered to be very user friendly. Hence, if we are able to reduce the amount of experimental data needed, we could save time, money and labour.

Decreasing the number of experiments without compromising the precision of the parameter estimates can be considered a subset selection problem. Different approaches are available in literature to solve such problems, but in this thesis a straightforward case-specific algorithm was developed and implemented (Algorithm 1 in Section 5.5).

In a first step, the same noise free data as in Section 9.2 was used to test the performance of the newly proposed subset selection algorithm. Besides θ_{opt} and \mathbf{y}_{exp} , values for the termination criterion (d_{min}), the maximum allowed number of iterations ($ITER_{max}$) and the number of experiments in a subset (N_{sub}) need to be declared beforehand. The former two values were arbitrarily chosen: $d_{min} = 0.1$ and $ITER_{max} = 1000$. In contrast, N_{sub} was varied between 2 and 70. It is noteworthy that the subset selection algorithm includes parameter estimation. Hence, different optimisation algorithms can be used during the subset selection process. In this thesis, the performances of both the simplex and SIMPSA algorithm were compared when using the same settings as in Table 9.1.

Figure 9.6 gives a clear overview of the result. For each evaluated number of experiments, the subset selection algorithm was run fifty times and either termination information was saved. The bars in the figure show the number of occurrences for each of the three possible states after the selection process:

- Unsuccessful selection: the total number of tested subsets exceeded the maximum allowed number ($ITER_{max}$), which indicates a non-successful subset selection.
- Optimal selection: the subset selection algorithm was able to find a good subset after changing some neighbouring experiments.

- Direct optimisation: the algorithm did not select any new neighbouring experiments since the initial subset was already informative enough.

From this figure, it can be seen that when the number of experiments in a subset increases, the probability of directly finding an optimal subset, i.e. direct optimisation, increases as well. However, it seems that the simplex algorithm has a harder time finding an optimal subset. This can be seen from the fact that the number of iterations which need additional sampling, i.e. optimal sampling, is higher for the simplex algorithm. Unlike the simplex algorithm, the SIMPSA optimisation algorithm seems to be able to estimate the kinetic parameters easily when five or more random experiments are available. It is even possible to find the true system parameter by only taking into account three well-chosen experiments. This is definitely not the case for the deterministic simplex algorithm since the number and the location of the experiments must be well chosen. Clearly, these occurrences are due to the local character of the latter algorithm.

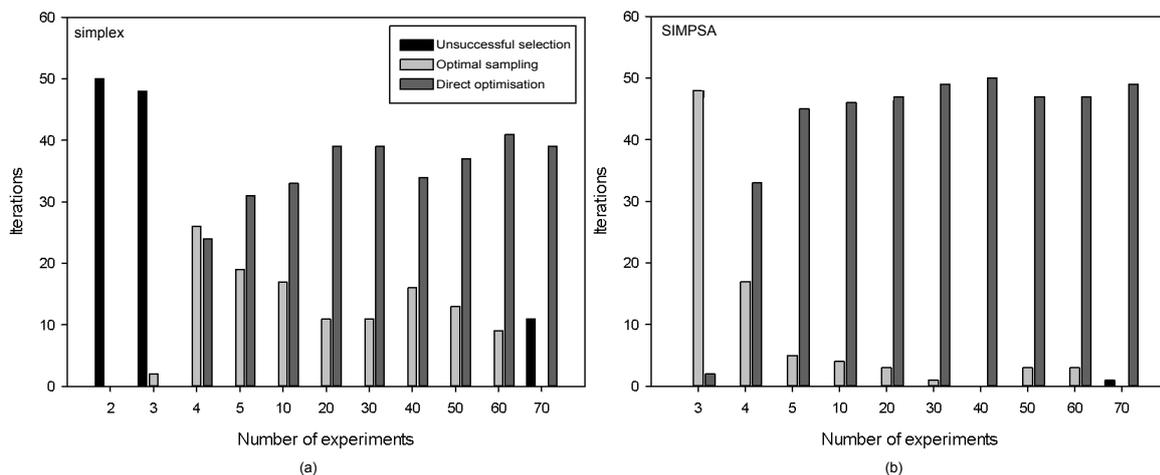


Figure 9.6: The result after subset selection using either the simplex (a) or the SIMPSA (b) optimisation algorithm for evaluating the goodness of a subset.

Remarkably, both optimisation algorithms have some problems when estimating the parameters using the total data set, i.e. 70 experiments. However, this is just the result of how the selection algorithm was designed. If the total data set is used, no additional experiments will be left for subsequent selection. Hence, the algorithm is forced to stop after its first attempt. Since the simplex algorithm performs only a local search, it could happen that the initial guess does not allow in finding the overall minimum (Figure 9.2). Note that the initial guess is kept constant during a run of the selection algorithm so that all proposed subsets are evaluated starting from the same point in the parameter space. The reason why the number of iterations is also exceeded once when using the SIMPSA algorithm, was not thoroughly investigated. Most likely, this is related to the settings of the algorithm which favour fast convergence.

The main aim of the subset selection algorithm is to distinguish subsets of experiments which are as informative as the total set of experiments. However, there will always be some degree of randomness in the path of this algorithm, for example, the randomly chosen initial guess and the stochastic behaviour of the SIMPSA algorithm. Hence, it is not capable of finding a single best subset of experiments. The goal of the algorithm is thus revealing either trends in the selection process or patterns in the obtained subsets. Indeed, a simple observation is that when experiments in the initial subset only vary in one of the two degrees of freedom, the selection algorithm proposes to vary the other degree of freedom as well. Clearly, more complex patterns could be observed. However, a thorough evaluation of the patterns requires an appropriate pattern recognition tool which was not yet fully exploited in this thesis.

Note that the aforementioned testing was done using noise-free *in silico* data. Intuitively, it is expected that the results for true experimental data will favour the ‘iterations exceeded’ state. However, looking back to Figure 9.5 learns us that the result might not be that different at all, especially not when the measurement noise and the number of consecutive points are kept relatively low. However, if the previous assumption is not valid, a potential pitfall for this method is the termination criterion d_{min} . In our case d_{min} only takes into account the relative distance to the true system parameters. However, these are only known in the case of *in silico* data. To deal with this shortcoming an alternative approach is suggested.

When using real experimental data, the only way of evaluating the goodness of an estimated parameter set is to compare the optimal parameter set for different repetitions of the optimisation i.e. starting from different initial parameter guesses. Hence, in the first step of this potential alternative approach multiple optimisations starting from the same subset of experiments, yet different initial parameter guesses, should be conducted. Next, the resulting parameter sets need to be compared based on the average euclidean distance between the parameter sets of the sets in parameter space. If this distance fulfills a predefined criterion, it is assumed that the global optimum is found, i.e. the true system parameters. If not, one of the experiments in the subset needs to be replaced so that the whole process can start over. This simple, yet powerful tool is just one of the possible ways to find the optimal subset of experiments in the case of noise-corrupted experimental data.

CHAPTER 10

The coupled tri-enzyme system

10.1 Proof of concept

The use of a coupled tri-enzyme system for the production of PEA has recently gained substantial interest. The reason being that, in practice, it has been confirmed that the production yield of PEA significantly increases if the coproduct pyruvate is removed *in situ* (Shin and Kim, 1999) (Figure 1.2). The aim of this section is to confirm that the same results are obtained using virtual simulations. Note that this part does not build on previous results, it is merely for proving the concept of multi-enzyme processes.

Due to the lack of some parameter values, preliminary parameter estimation was needed for the LDH and GDH subsystems. In this respect, the same approach as in Section 8.3.2 was used to estimate the unknown parameters. However, the upper bound for V_2 was set to be one tenth of the limiting rate in the forward direction (V_1). This is done because for both the LDH and GDH reactions, it is known that the equilibrium clearly favours the product side of the forward reaction. The resulting values are given in Table C.14 of the Appendix. As mentioned before, the complete model used for describing both subsystems is not practical identifiable. Therefore, the estimated values should obviously not be considered as the true system parameters since they only describe the system under the investigated experimental conditions.

To simulate the tri-enzyme system, i.e. the system with the cofactor regeneration cycle, Eq. (2.9) was used for both the LDH and GDH conversions. However, for the TA conversion substrate inhibition was included to the aforementioned model (Eq. (2.10)) in order to resemble reality as close as possible (Shin and Kim, 1998) (Figure 10.1).

$$r_{TA} = \frac{\frac{V_1 a b}{K_{iA} K_{mB}} - \frac{V_2 p q}{K_{iP} K_{mQ}}}{\frac{a}{K_{iA}} \left(1 + \frac{a}{K_{siA}}\right) + \frac{K_{mA} b}{K_{iA} K_{mB}} \left(1 + \frac{b}{K_{siB}}\right) + \frac{p}{K_{iP}} + \frac{K_{mP} q}{K_{iP} K_{mQ}} + \frac{a b}{K_{iA} K_{mB}} \dots} + \frac{\frac{a p}{K_{iA} K_{iP}} \left(1 + \frac{a}{K_{siA}}\right) + \frac{K_{mA} b q}{K_{iA} K_{mB} K_{iQ}} \left(1 + \frac{b}{K_{siB}}\right) + \frac{p q}{K_{iP} K_{mQ}}}{\dots} \quad (10.1)$$

In Figure 10.2 the result of both simulations with and without cofactor regeneration cycle

is depicted. The system without the regeneration cycle reaches equilibrium relatively fast, whereas the system with additional regeneration keeps on converting acetophenone (Ace) to (S)-1-phenylethylamine. The regeneration cycle continues until the the glucose concentration becomes limiting.

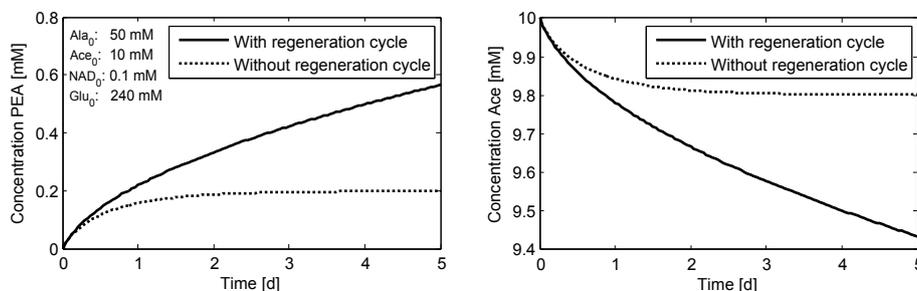


Figure 10.1: Simulation results for the tri-enzyme system. The initial concentrations are shown in the upper left corner.

Nowadays, protein engineering and *ab initio* design, i.e. computer aided design and subsequent *in vivo* expression, of biocatalysts enables scientists to increase enzyme activity. Increasing the activity of the enzymes involved in the regeneration step will have no significant effect since the thermodynamic equilibrium and kinetics of these reactions already favour the formation of the desired components. However, increasing the activity of the rate-limiting enzyme i.e. the TA leads to an increased product yield and a shorter equilibration time. The result is depicted in Figure 10.2. The data obtained from DTU showed a comparable time-scale for the conversion, indicating that the enzymes used in their lab experiments clearly have a higher activity than those first described by Shin and Kim (1998).

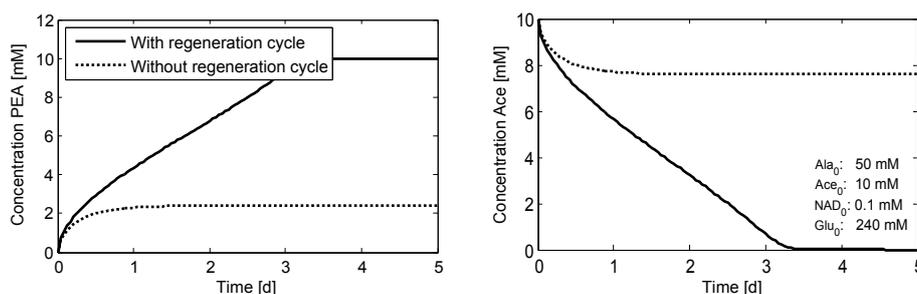


Figure 10.2: Simulation results for the tri-enzyme system if the limiting rate (V_1) of the TA reaction is multiplied with a factor 200 (Santacoloma, 2012).

PART V

CONCLUSIONS AND PERSPECTIVES

Conclusions

In this MSc. thesis, a rather theoretical study was conducted in order to gain more insight into next-generation biocatalytic processes, i.e. multi-enzyme processes. This study focused on calibration and experimental design for a bi bi compulsory-order ternary-complex mechanism. In this respect, special attention was paid to the recently published work of Al-Haque et al. (2012).

First, the practical identifiability of the model describing the complete mechanism of the LDH enzyme was investigated based on the available data i.e. batch reaction data at different initial concentrations of NADH and pyruvic acid. More specifically, a global optimisation algorithm (SIMPISA) was repeatedly used to estimate the parameters of the model based on one single batch experiment and using the sum of squared errors as objective function. This analysis revealed that no unique optimal parameter set could be found. In fact, this is the result of strong correlations between the parameters. In order to enhance the practical identifiability, several improvements like manual tuning of the optimisation algorithm, rescaling the objective function and taking multiple experimental data sets into account, were suggested. Even though these improvements did not lead to an exclusive set of parameters, some of them can definitely be used for other continuous optimisation problems.

Practical identifiability of the complete model based on entire progress curves of batch reactions was thus rejected. It should be noted that, a lack of practical identifiability does not imply that parameter estimation techniques fail in regressing the model. The problem is that the obtained estimates are not unique and in fact meaningless for these deterministic models because they do not allow for extrapolation or validation. Hence, software packages designed especially for estimating the kinetic parameters of enzyme kinetic models should be applied with caution. In addition, optimal experimental design (OED) could not be used since practical identifiability is a prerequisite for this method.

Besides the more practical approach, structural identifiability was also scrutinised using the differential algebra method, but to no avail. The used software bumped into computational problems due to the highly complex model structure. Hence, no conclusions could be drawn with regard to over-specification of the complete model. However, it was shown that even a simplified model revealed structural unidentifiability. Although over-specification of a model can be solved by simplifying the model structure, this is not recommended for kinetic models in the field of biocatalysis. Indeed, simplified models result in lumped parameters and a corresponding loss of information. A relevant example is the following: based on the ratio of different kinetic parameters, enzyme properties like the intrinsic enzyme efficiency can be deduced. This is, however, no longer the case when reducing model complexity.

Clearly, progress curve analysis can not be used directly for a precise estimation of all the parameters in the lactate dehydrogenase model. Therefore, alternative approaches such as the stepwise incremental method proposed by Al-Haque et al. (2012) are mandatory. In this method, the parameters related to the initial reaction rate are estimated beforehand.

Identifiability analysis revealed that both the forward and reverse initial rate models are structurally identifiable. This was also proved by the fact that the true parameters of the initial rate model could be uniquely estimated from noise-free *in silico* data at different initial concentrations. Furthermore, the initial rate model is quite insensitive to changes in the sole inhibition parameter of the model. In contrast, the limiting rate reveals to be the most sensitive parameter.

Moreover, it was shown that practical identifiability can be guaranteed when using noise-corrupted virtual data. In this respect, noise corrupted data were used to determine noise corrupted initial rates as the slope of the linear regression through the data. Both the variance of the noise and the number of data points used in the regression, were varied. Because even the parameters of identifiable models can only be estimated within a confidence interval, an easy-to-grasp figure was developed in order to perform a first assessment of the estimates' precision.

Regarding the number of experiments needed to calibrate the initial rate model, additional testing was done using a straightforward case-specific subset selection algorithm. Originally, it was assumed that a large data set was needed in order to calibrate the initial rate model. However, this research clearly showed that when using a robust optimisation algorithm and high-quality data, the number of experimental data sets might reduce to three well-selected experiments. In order to select a good experimental design, the suggested subset selection algorithm can be used as a valuable tool, even if the true system parameters are not known. This should be further verified on a real case.

Finally, the usefulness of multi-enzyme systems, in particular the tri-enzyme system for the production of (S)-1-phenylethylamine, was demonstrated by performing virtual simulations. Although the true system parameters are not thoroughly known, it was clearly shown that the use of *in situ* coproduct removal via an enzymatic cofactor regeneration cycle shifts the equilibrium towards the desired product side.

Future perspectives

It is often cited that the use of progress curves from enzymatic conversions hold particular promise regarding precise kinetic models. The main argument for this belief is that progress curves contain more information on the conversion in contrast to initial rate data. However, when using the entire time course of a conversion, the model becomes highly complex because additional phenomenons such as inhibition become important. Hence, identifiability can no longer be guaranteed.

That is why future research should not focus solely on progress curve analysis but rather on systematic approaches which combine progress curve analysis and initial rate analysis. The method proposed by Chen et al. (2008) and Al-Haque et al. (2012) are good starting points, but as the author claims himself, this is still a robust methodology which leaves room for improvements. A possible improvement is the inclusion of methodologies for optimal experimental design and thus reducing the number of experimental data needed.

Unlike the complete model, the initial rate models allowed for structural and practical identifiability. Hence, optimal experimental design should be performed for these models so that experimental work only focuses on what is relevant for the process and its model. In this respect, it is advised that the result of the suggested subset selection algorithm is evaluated using advanced pattern recognition tools. This could ultimately lead to a set of easy-to-use rules, that allow for defining informative sets of experimental designs. In addition, the subset selection algorithm should be validated on real experimental data.

Although it was not mentioned before, spiking of the system with substrates and/or products could also be examined w.r.t. identifiability. It was already described that for relatively simple enzyme systems spiking significantly improves the precision of the estimated parameters. In this respect, some preparatory work was done using the LDH model. However to date, no conclusions could be drawn.

Finally, if the models of the individual enzymatic steps in a multi-enzyme system are well understood, the transition towards modelling of a multi-enzyme process can be made. Intuitively, it is expected that additional phenomena such as inhibition and other interactions will become important in the case of a multi-enzyme system. Hence, model complexity will increase even more, which requires additional sophisticated evaluation techniques.

PART VI

APPENDICES

APPENDIX A

Relation kinetic parameters and rate constants

To facilitate the use of the rate expressions, the characters A and B are assigned to the substrates of a bi bi enzyme conversion, whereas the products are denoted using P and Q (Table A.1).

Table A.1: The following characters representing either substrate or product, will be used throughout this thesis.

	A	B	P	Q
TA	L-alanine	acetophenone	pyruvate	(S)-1-phenylethylamine
LDH	NADH	pyruvate	(S)-lactate	NAD
GDH	NAD	glucose	glucono- δ -lactone	NADH

The relations between the kinetic parameters and the rate constants are given in Table A.2.

While kinetic parameters (V , K_m , K_i) can always be expressed in terms of rate constants (k_i), an inverse relation is not always possible. Thus, a distinction between rate equations with identifiable and unidentifiable rate constants should be made. A simple example of a rate expression with unidentifiable rate constants is the Michaelis-Menten equation. Here, no unique set of rate constants can be found if the kinetic parameters are known. Hence, we call them unidentifiable rate constants. This is also the case for more complex mechanisms such as the rate equation for the substituted-enzyme mechanism (Straathof and Heijnen, 1996). An example of an expression with identifiable rate constants is the compulsory-order ternary-complex mechanism. The relations for this mechanism are given in Table A.3 (Cornish-Bowden, 2004).

Table A.2: Definitions of the kinetic parameters for both mechanisms. Although K_{iB} is not in Eq. (2.10), it can be derived using the Haldane relation given in Section 2.4

	Ternary-complex mechanism	Substituted-enzyme mechanism
V_1	$\frac{k_3 k_4 e_0}{k_3 + k_4}$	$\frac{k_2 k_4 e_0}{k_2 + k_4}$
V_2	$\frac{k_{-1} k_{-2} e_0}{k_{-1} + k_{-2}}$	$\frac{k_{-1} k_{-3} e_0}{k_{-1} + k_{-3}}$
K_{mA}	$\frac{k_3 k_4}{k_1 (k_3 + k_4)}$	$\frac{k_4 (k_{-1} + k_2)}{k_1 (k_2 + k_4)}$
K_{mB}	$\frac{k_4 (k_{-2} + k_3)}{k_2 (k_3 + k_4)}$	$\frac{k_2 (k_{-3} + k_4)}{k_3 (k_2 + k_4)}$
K_{mP}	$\frac{k_{-1} (k_{-2} + k_3)}{k_{-3} (k_{-1} + k_{-2})}$	$\frac{k_{-3} (k_{-1} + k_2)}{k_{-2} (k_{-1} + k_{-3})}$
K_{mQ}	$\frac{k_{-1} k_{-2}}{k_{-4} (k_{-1} + k_{-2})}$	$\frac{k_{-1} (k_{-3} + k_4)}{k_{-4} (k_{-1} + k_{-3})}$
K_{iA}	$\frac{k_{-1}}{k_1}$	$\frac{k_{-1}}{k_1}$
K_{iB}	$\frac{k_{-1} + k_{-2}}{k_2}$	$\frac{k_{-3}}{k_3}$
K_{iP}	$\frac{k_{-3} + k_4}{k_{-3}}$	$\frac{k_2}{k_{-2}}$
K_{iQ}	$\frac{k_4}{k_{-4}}$	$\frac{k_4}{k_{-4}}$

Table A.3: The inverse relations between the rate constants and the kinetic parameters for the compulsory-order ternary-complex mechanism

Rate constant	
k_1	$\frac{V_1}{K_{mA} e_0}$
k_2	$\frac{V_1(k_{-2}+k_3)}{k_3 K_{mB} e_0}$
k_3	$\frac{V_1 V_2 K_{iQ}}{e_0(V_2 K_{iQ} - V_1 K_{mQ})}$
k_4	$\frac{V_2 K_{iQ}}{K_{mQ} e_0}$
k_{-1}	$\frac{V_1 K_{iA}}{K_{mA} e_0}$
k_{-2}	$\frac{V_1 V_2 K_{iA}}{e_0(V_1 K_{iA} - V_2 K_{mA})}$
k_{-3}	$\frac{V_2(k_{-2}+k_3)}{k_{-2} K_{mP} e_0}$
k_{-4}	$\frac{V_2}{K_{mQ} e_0}$

APPENDIX B

Experimental designs

In the following tables an overview of different experimental designs is given. These experiments were carried out by The Center for Process Engineering and Technology (PROCESS) at the Technical University of Denmark (DTU).

Table B.1: Experimental design of the different LDH experiments.

	Initial concentration		Measurements	
	NADH ($\mu\text{M}/\text{l}$)	Pva ($\mu\text{M}/\text{l}$)	Duration (min)	Frequency (1/s)
Exp. 1	100	100	30	1
Exp. 2	100	200	15	1
Exp. 3	100	300	15	1
Exp. 4	100	400	15	1
Exp. 5	100	500	15	1
Exp. 6	100	1000	10	2
Exp. 7	100	2000	10	2
Exp. 8	100	3000	5	2
Exp. 9	100	4000	5	2
Exp. 10	100	5000	5	2

Table B.2: Experimental design of the different LDH initial rate experiments. All experiments were measured for 1 minute with a measurement frequency of 20 Hz.

	Initial concentrations	
NADH (μM)	100	300
	200	500
Pva (μM)	100	3000
	200	4000
	300	5000
	500	10000
	1000	15000
	2000	20000
		25000

Table B.3: Experimental design of the different GDH experiments.

	Initial concentrations		Measurements	
	NAD ⁺ ($\mu\text{M/l}$)	Glu ($\mu\text{M/l}$)	Duration (min)	Frequency (1/s)
Exp. 1	150.7	100	10.0	1
Exp. 2	150.7	250	10.0	1
Exp. 3	150.7	1000	10.0	1
Exp. 4	150.7	10000	5.6	1
Exp. 5	150.7	50000	10.0	1
Exp. 6	150.7	1000	5.3	1
Exp. 7	226.1	100000	10.0	1
Exp. 8	226.1	250	10.0	1
Exp. 9	226.1	1000	10.0	1
Exp. 10	226.1	10000	10.0	1
Exp. 11	226.1	50000	10.0	1
Exp. 12	226.1	100000	10.0	1
Exp. 13	301.5	100	10.0	1
Exp. 14	301.5	250	10.0	1
Exp. 15	301.5	10000	10.0	1
Exp. 16	301.5	50000	10.0	1
Exp. 17	301.5	5000	10.0	1
Exp. 18	301.5	100000	10.0	1

APPENDIX C

Results and discussion

Table C.1: Parameter values after optimisation.

Parameter	Value	
	Initial	Final
V_1	1.93655447×10^7	1.93572496×10^7
V_2	5.298662×10^5	1.2758126×10^6
K_{mA}	1.3574017×10^6	4.506756×10^5
K_{iA}	1.6979848×10^6	1.6979517×10^6
K_{mB}	1.3162406×10^6	1.3570537×10^6
K_{iB}	3.1432380×10^6	1.7679455×10^6
K_{mP}	3.6749689×10^6	3.1652505×10^6
K_{iP}	4.9496166×10^6	1.0538413×10^6
K_{mQ}	3.0927219×10^5	4.9266544×10^6
K_{iQ}	1.7755018×10^5	3.5763446×10^6

Table C.2: Summarizing result after 20 iterations of progress curve analysis. Standard settings; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	1.87E+07	1.08E-02	6.30E+02	2.50E+06	5.63E+05	4.86E+04	6.57E+03	2.84E+05	4.25E+06	5.00E+05	79.48
Max	2.95E+07	2.22E+06	2.13E+06	4.98E+06	9.59E+05	2.82E+06	4.58E+06	4.45E+06	5.00E+06	4.57E+06	100.27
Stdv	3.17E+06	6.58E+05	5.95E+05	7.92E+05	1.14E+05	9.29E+05	1.47E+06	1.21E+06	1.95E+05	1.20E+06	4.84
Rank	1	7	8	6	10	5	2	3	9	4	-
Stdv/range	0.106	0.022	0.119	0.158	0.023	0.186	0.295	0.243	0.039	0.239	-
Rank	7	10	6	5	9	4	1	2	8	3	-

Table C.3: Summarizing result after 20 iterations of progress curve analysis. Standard settings; starting from the different initial guesses.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	SSE
Min	6.43E+06	1.29E-01	3.28E+02	2.69E+05	2.77E+05	3.24E+05	8.97E+03	7.19E+05	1.35E+06	9.41E+03	9.51E+01
Max	2.93E+07	2.58E+06	2.71E+06	3.95E+06	4.92E+06	4.57E+06	4.77E+06	4.89E+06	4.99E+06	4.96E+06	1.00E+02
Stdv	7.29E+06	7.70E+05	8.77E+05	1.02E+06	1.17E+06	1.41E+06	1.14E+06	1.15E+06	1.16E+06	1.17E+06	1.33E+00
Rank	1	10	9	8	4	2	7	6	5	3	-
Stdv/range	0.243	0.026	0.175	0.204	0.235	0.282	0.229	0.230	0.232	0.235	-
Rank	2	10	9	8	4	1	7	6	5	3	-

Table C.4: Summarizing result after 20 iterations of progress curve analysis. No initial temperature; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	9.70E+06	2.25E-01	2.61E+02	1.06E+06	2.77E+05	6.89E+05	8.28E+03	1.54E+05	1.19E+06	2.42E+06	9.59E+01
Max	2.98E+07	2.58E+06	2.83E+06	4.66E+06	1.96E+06	4.29E+06	4.98E+06	4.93E+06	4.96E+06	4.78E+06	1.00E+02
Stdv	6.20E+06	6.18E+05	6.32E+05	1.19E+06	4.98E+05	1.04E+06	1.10E+06	1.33E+06	1.48E+06	7.50E+05	9.66E-01
Rank	1	9	8	4	10	6	5	3	2	7	-
Stdv/range	0.207	0.021	0.126	0.238	0.100	0.208	0.220	0.267	0.295	0.150	-
Rank	6	10	8	3	9	5	4	2	1	7	-

Table C.5: Summarizing result after 20 iterations of progress curve analysis. Freezing temperature of 0.1; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	1.73E+07	1.88E-01	1.36E+03	1.60E+06	5.07E+05	1.11E+04	4.62E+03	3.69E+04	8.97E+05	8.80E+05	9.38E+01
Max	2.83E+07	2.12E+06	1.80E+06	4.73E+06	1.29E+06	4.98E+06	4.74E+06	4.43E+06	5.00E+06	4.88E+06	1.00E+02
Stdv	2.74E+06	4.77E+05	4.00E+05	9.04E+05	2.08E+05	1.51E+06	1.02E+06	1.27E+06	1.12E+06	1.17E+06	1.44E+00
Rank	1	8	9	7	10	2	6	3	5	4	-
Stdv/range	0.091	0.016	0.080	0.181	0.042	0.303	0.204	0.254	0.224	0.233	-
Rank	7	10	8	6	9	1	5	2	4	3	-

Table C.6: Summarizing result after 20 iterations of progress curve analysis. Multiplying the value of the objective function with a factor; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	1.76E+07	2.47E-02	1.34E+01	1.67E+06	5.34E+05	3.91E+05	5.92E+03	2.59E+05	4.59E+06	2.62E+06	9.44E+07
Max	2.65E+07	2.59E+06	4.81E+03	4.82E+06	1.26E+06	4.49E+06	4.84E+06	3.81E+06	5.00E+06	4.76E+06	1.00E+08
Stdv	2.24E+06	5.86E+05	1.40E+03	6.67E+05	1.46E+05	9.84E+05	1.10E+06	9.26E+05	9.12E+04	6.41E+05	1.30E+06
Stdv/range	0.075	0.020	0.000	0.133	0.029	0.197	0.220	0.185	0.018	0.128	-
Rank	6	8	10	4	7	2	1	3	9	5	-

Table C.7: Summarizing result after 20 iterations of progress curve analysis. Taking into account the additional constraints; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$	ΔH	ΔnH
Min	1.14E+06	1.05E+06	8.71E+04	5.99E+05	2.55E+05	3.40E+06	1.75E+06	2.16E+06	2.06E+00	4.07E+00	1.00E+02	9.99E-07	9.70E-08
Max	2.62E+07	5.00E+06	3.26E+06	3.94E+06	3.37E+06	4.92E+06	4.80E+06	5.00E+06	3.49E+06	5.00E+06	2.92E+02	2.43E-05	3.38E+00
Stdv	6.96E+06	1.04E+06	9.56E+05	1.09E+06	9.61E+05	4.34E+05	8.57E+05	7.67E+05	8.72E+05	1.40E+06	4.26E+01	6.75E-06	7.56E-01
Rank	1	4	6	3	5	10	8	9	7	2	-	-	-
Stdv/range	2.32E-01	3.45E-02	1.91E-01	2.17E-01	1.92E-01	8.69E-02	1.71E-01	1.53E-01	1.74E-01	2.80E-01	-	-	-
Rank	2	10	5	3	4	9	7	8	6	1	-	-	-

Table C.8: Summarizing result after 20 iterations of progress curve analysis. Considering the literature parameters as constants; starting from the same initial guess.

	V_2 [$\mu\text{M/s}$]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$	ΔH	ΔnH
Min	1.18E+04	5.00E+03	3.77E+03	2.43E+05	1.17E+06	1.19E+03	8.92E+04	8.41E+02	1.11E+02
Max	3.80E+06	2.74E+06	5.00E+06	4.98E+06	5.00E+06	4.99E+06	4.63E+06	8.96E+02	1.13E+02
Stdv	1.34E+06	1.00E+06	1.24E+06	1.01E+06	8.60E+05	1.39E+06	2.06E+06	2.14E+01	5.97E-01
Rank	2	5	3	4	6	1	-	-	-
Stdv/range	0.045	0.201	0.247	0.202	0.172	0.277	-	-	-
Rank	6	4	2	3	5	1	-	-	-

Table C.9: Summarizing result after 20 iterations of progress curve analysis. Weighted by the maximum of the total relative sensitivity functions; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	1.37E+07	2.10E+04	1.53E+05	7.88E+05	7.74E+05	9.17E+04	2.83E+06	2.09E+04	7.75E+05	1.44E+06	1.91E-02
Max	3.00E+07	2.01E+07	4.46E+06	3.47E+06	3.39E+06	4.02E+06	4.99E+06	4.15E+06	4.46E+06	4.97E+06	1.45E-01
Stdv	4.46E+06	4.91E+06	1.25E+06	6.64E+05	7.72E+05	9.60E+05	6.88E+05	1.36E+06	9.69E+05	1.20E+06	3.79E-02
Rank	2	1	4	10	8	7	9	3	6	5	-
Stdv/range	0.149	0.164	0.251	0.133	0.154	0.192	0.138	0.271	0.194	0.240	-
Rank	8	6	2	10	7	5	9	1	4	3	-

Table C.10: Summarizing result after 20 iterations of progress curve analysis. Weighted by the maximum of the total relative sensitivity functions; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	1.26E+07	2.84E-01	3.72E+05	4.91E+05	5.61E+05	1.19E+06	2.39E+06	2.90E+06	1.88E+04	7.29E+02	4.07E+04
Max	2.92E+07	8.72E+02	4.57E+06	4.78E+06	3.49E+06	5.00E+06	4.76E+06	5.00E+06	4.59E+06	5.00E+06	4.08E+04
Stdv	5.51E+06	2.25E+02	1.18E+06	1.51E+06	8.42E+05	1.04E+06	7.10E+05	4.83E+05	1.23E+06	1.78E+06	1.09E+01
Rank	1	10	5	3	7	6	8	9	4	2	-
Stdv/range	0.184	0.000	0.237	0.303	0.168	0.208	0.142	0.097	0.246	0.357	-
Rank	6	10	4	2	7	5	8	9	3	1	-

Table C.11: Summarizing result after 20 iterations of progress curve analysis. Considering data from multiple experiments; starting from the same initial guess.

	V_1 [$\mu\text{M/s}$]	V_2 [$\mu\text{M/s}$]	K_{mA} [μM]	K_{iA} [μM]	K_{mB} [μM]	K_{iB} [μM]	K_{mP} [μM]	K_{iP} [μM]	K_{mQ} [μM]	K_{iQ} [μM]	$J(\hat{\theta})$
Min	2.00E+07	4.92E+06	5.36E+05	8.81E+05	8.97E-05	5.59E+03	4.72E+03	2.27E+02	2.05E+06	4.34E-03	5.59E+03
Max	2.88E+07	5.00E+06	2.93E+06	4.58E+06	9.59E+00	4.76E+06	4.42E+06	4.94E+06	4.72E+06	3.69E+06	4.76E+06
Stdv	2.76E+06	2.66E+04	8.95E+05	1.24E+06	3.12E+00	1.82E+06	1.66E+06	1.84E+06	9.44E+05	1.37E+06	1.82E+06
Rank	1	9	8	6	10	3	4	2	7	5	-
Stdv/range	0.091859979	0.000885387	1.79E-01	2.48E-01	6.25E-07	3.63E-01	3.32E-01	3.68E-01	1.89E-01	2.73E-01	-
Rank	8	9	7	5	10	2	3	1	6	4	-

Table C.12: Parameter values for the LDH-1 isoenzyme taken from the literature (Borgmann et al., 1975). Note that these are not valid for the experimental data which was available to us.

Parameter	Value	
V_1	25.74	$\mu\text{M}/\text{s}$
K_{mA}	2499.10	μM
K_{iA}	2030.30	μM
K_{mB}	2.23	μM
Temp.	37	$^{\circ}\text{C}$
pH	10	

Table C.13: Remaining parameter values needed to generate *in silico* data.

Parameter	Value	
V_2	0.22	$\mu\text{M}/\text{s}$
K_{iB}	436.75	μM
K_{mP}	41.00	μM
K_{iP}	585.05	μM
K_{mQ}	12.19	μM
K_{iQ}	848.58	μM
K_{iQ}	904.5	

Table C.14: Estimated parameter values using progress curve analysis. The values in bold were obtained from the literature.

	TA ¹	LDH ²	GDH ³
V_1 [$\mu\text{M}/\text{s}$]	8.63×10^{-3}	25.74	21.00
V_2 [$\mu\text{M}/\text{s}$]	7.00	2.40	2.09
K_{mA} [μM]	1070.00	249.91	3.17
K_{iA} [μM]	2850.00	2.23	6.21
K_{mB} [μM]	540.00	2030.30	4560.00
K_{iB} [μM]	130.00	1962.16	320693.37
K_{mP} [μM]	9850.00	4326.62	36244.22
K_{iP} [μM]	31.40	3377.87	1501799.30
K_{mQ} [μM]	35030.00	3868.18	2749662.30
K_{iQ} [μM]	10.20	2712.75	78601.38
K_{siA} [μM]	25.82	-	-
K_{siB} [μM]	1.24	-	-

¹ Shin and Kim (1998), ² Borgmann et al. (1975), ³ Carper et al. (1983)

PART VII

BIBLIOGRAPHY

Bibliography

- Aarts, E. H. L. and van Laarhoven, P. J. M. (1985). Statistical cooling: a general approach to combinatorial optimization problems. *Philips Journal of Research*, 40(4):193.
- Al-Haque, N., Santacoloma, P. A., Neto, W., Tufvesson, P., Gani, R., and Woodley, J. M. (2012). A robust methodology for kinetic model parameter estimation for biocatalytic reactions. *Biotechnology Progress*, (1):1–11.
- Alberton, A. L., Schwaab, M., Lobão, M. W. N., and Pinto, J. C. (2011). Experimental design for the joint model discrimination and precise parameter estimation through information measures. *Chemical Engineering Science*, 66(9):1940–1952.
- Ataíde, F. and Hitzmann, B. (2009). When is optimal experimental design advantageous for the analysis of MichaelisMenten kinetics? *Chemometrics and Intelligent Laboratory Systems*, 99(1):9–18.
- Bates, D. J. and Frieden, C. (1973). Treatment of enzyme kinetic data. 3. Use of full time course of a reaction, as examined by computer-simulation, in defining enzyme mechanisms. *The Journal of Biological Chemistry*, 248(22):7878–7884.
- Borgmann, U., Laidler, K. J., and Moon, T. (1975). Kinetics and thermodynamics of lactate dehydrogenases from beef heart, beef muscle, and flounder muscle. *Canadian Journal of Biochemistry*, 53(11):1196–1206.
- Briggs, G. E. and Haldane, J. B. S. (1925). A note on the kinetics of enzyme action. *Biochemical Journal*, 19(2):338–339.
- Burden, R. L. and Faires, J. D. (2005). *Numerical Analysis*. Brooks/Cole, Belmont, USA, 8th edition.
- Cardoso, M. F., Salcedo, R., and Feyo de Azevedo, S. (1996). The simplex-simulated annealing approach to continuous non-linear optimization. *Computers & Chemical Engineering*, 20(9):1065–1080.
- Carper, W. R., Campbell, D. P., Morriscal, S. W., and Thompson, R. E. (1983). Characterization studies of glucose dehydrogenase. *Experientia*, 39(11):1295–1297.

- Chen, B. H., Hibbert, E. G., Dalby, P. A., and Woodley, J. M. (2008). A new approach to bioconversion reaction kinetic parameter identification. *AIChE Journal*, 54(8):3–5.
- Chis, O.-T., Banga, J. R., and Balsa-Canto, E. (2011). Structural identifiability of systems biology models: a critical comparison of methods. *PloS ONE*, 6(11):1–16.
- Cleland, W. W. (1963). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I Nomenclature and rate equations. *Biochimica et Biophysica Acta*, 67(2):104–137.
- Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics*. Portland Press, London, third edition.
- De Pauw, D. J. (2005). *Optimal Experimental Design for Calibration of Bioprocess Models : a Validated Software Toolbox*. PhD Thesis, Gent University.
- De Pauw, D. J. and Vanrolleghem, P. a. (2006). Practical aspects of sensitivity function approximation for dynamic models. *Mathematical and Computer Modelling of Dynamical Systems*, 12(5):395–414.
- Dochain, D. (2008). *Bioprocess control*. John Wiley & Sons Ltd, New York.
- Dochain, D. and Vanrolleghem, P. A. (2001). *Dynamical Modelling & Estimation in Wastewater Treatment Processes*. IWA Publishing.
- Donckels, B. M. (2009). *Optimal Experimental Design to Discriminate Among Rival Dynamic Mathematical Models*. PhD Thesis, Gent University, <http://biomath.ugent.be/%7ebrecht/index.html>.
- Donckels, B. M., De Pauw, D. J., Vanrolleghem, P. A., and De Baets, B. (2010). An ideal point method for the design of compromise experiments to simultaneously estimate the parameters of rival mathematical models. *Chemical Engineering Science*, 65(5):1705–1719.
- Drauz, K., Gröger, H., and May, O. (1990). *Enzyme Catalysis in Organic Synthesis*, volume 64. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, third edition.
- Duggleby, R. G. and Clarke, R. B. (1991). Experimental designs for estimating the parameters of the Michaelis-Menten equation from progress curves of enzyme-catalyzed reactions. *Biochimica et Biophysica Acta*, 1080(3):231–236.
- Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63:4846–4872.
- Garcia-Junceda, E. (2008). *Multi-Step Enzyme Catalysis*, volume 2008. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

- GIA (2012). Global Industrial Enzymes Market to Reach US\$3.74 Billion by 2015, According to a New Report by Global Industry Analysts, Inc.
- Goudar, C. T., Harris, S. K., McInerney, M. J., and Suffita, J. M. (2004). Progress curve analysis for enzyme and microbial kinetic reactions using explicit solutions based on the Lambert W function. *Journal of Microbiological Methods*, 59(3):317–26.
- Goudar, C. T., Sonnad, J. R., and Duggleby, R. G. (1999). Parameter estimation using a direct solution of the integrated Michaelis-Menten equation. *Biochimica et Biophysica Acta*, 1429(2):377–83.
- Hattersley, J. G., Pérez-Velázquez, J., Chappell, M. J., Bearup, D., Roper, D., Dowson, C., Bugg, T., and Evans, N. D. (2011). Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis. *Computer Methods and Programs in Biomedicine*, 104(2):70–80.
- Hunter, W. G. and Reiner, A. M. (1965). Designs for discriminating between two rival models. *Technometrics*, 7(3):307–323.
- Illanes, A. (2008). *Enzyme Biocatalysis: Principales and Applications*. Springer Netherlands, Dordrecht.
- King, E. L. and Altman, C. (1956). A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *The Journal of Physical Chemistry*, 60(10):1375–1378.
- Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):98–104.
- Koszelewski, D., Tauber, K., Faber, K., and Kroutil, W. (2010). ω -Transaminases for the synthesis of non-racemic alpha-chiral primary amines. *Trends in Biotechnology*, 28(6):324–32.
- Laidler, K. J. (1997). A brief history of enzyme kinetics. In *New Beer in an Old Bottle: Eduard Buchner and the Growth of Biochemical Knowledge*, pages 127–133. Universitat de València, València.
- Lindner, P. O. F. and Hitzmann, B. (2006). Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the Fisher information matrix. *Journal of Theoretical Biology*, 238(1):111–23.
- Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, second edition.
- Marquardt, W. (2005). Model-based experimental analysis of kinetic phenomena in multi-phase reactive systems. *Chemical Engineering Research and Design*, 83(6):561–573.

- Meshkat, N., Anderson, C., and Distefano, J. J. (2011). Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input-output equations. *Mathematical biosciences*, 233(1):19–31.
- Michalik, C., Schmidt, T., Zavrel, M., Ansorge-Schumacher, M., Spiess, A., and Marquardt, W. (2007). Application of the incremental identification method to the formate oxidation using formate dehydrogenase. *Chemical Engineering Science*, 62(18-20):5592–5597.
- Michalik, C., Stuckert, M., and Marquardt, W. (2010). Optimal experimental design for discriminating numerous model candidates: The AWDC criterion. *Industrial & Engineering Chemistry Research*, 49(2):913–919.
- Murphy, E. F., Gilmour, S. G., and Crabbe, M. J. C. (2002). Effective experimental design: enzyme kinetics in the bioinformatics era. *Drug Discovery Today*, 7(20 Suppl):S187–91.
- Nelder, J. A. and Mead, R. (1964). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.
- Nestl, B. M., Nebel, B. a., and Hauer, B. (2011). Recent progress in industrial biocatalysis. *Current Opinion in Chemical Biology*, 15(2):187–93.
- Nugent, T. and El-Shazly, M. (2010). Chiral Amine Synthesis - Recent Developments and Trends for Enamide Reduction, Reductive Amination, and Imine Reduction. *Advanced Synthesis & Catalysis*, 352(5):753–819.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, volume 29. Cambridge University Press, Cambridge, second edition.
- Saccomani, M. P. and Audoly, S. (2010). Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Computers in Biology and Medicine*, 40(4):402–407, <http://www.dei.unipd.it/%7epia/>.
- Santacoloma, P. A. (2012). *Multi-enzyme Process Modeling*. PhD Thesis, Danmarks Tekniske Universitet.
- Scheerlinck, K. (2012). *Metaheuristic versus tailor-made approaches to optimization problems in the biosciences*. PhD Thesis, Gent University.
- Schmidt, H. and Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics (Oxford, England)*, 22(4):514–515, <http://www.sbtoolbox2.org>.
- Schulz, A. R. (1994). *Enzyme Kinetics - From Diastase to Multi-enzyme Systems*. Cambridge University Press, Cambridge.

- Segel, L. A. and Slemrod, M. (1989). The quasi-steady-state assumption: a case study in perturbation. *Society for Industrial and Applied Mathematics*, 31(3):446–477.
- Serrano Briega, G. (2011). *Evaluation of Coupled Dehydrogenase Systems*. Number June. Master Thesis, Danmarks Tekniske Universitet.
- Shin, J.-s. and Kim, B.-g. (1998). Kinetic modeling of ω -transamination for enzymatic kinetic resolution of α -MBA. *Biotechnology and Bioengineering*, 60(5):534–540.
- Shin, J.-s. and Kim, B.-g. (1999). Asymmetric synthesis of chiral amines with ω -transaminase. *Biotechnology and Bioengineering*, 65(2):206–211.
- Straathof, A. (2001). Development of a computer program for analysis of enzyme kinetics by progress curve fitting. *Journal of Molecular Catalysis B: Enzymatic*, 11(4-6):991–998, <http://www.tnw.tudelft.nl/?id=35959&L=1>.
- Straathof, A. and Heijnen, J. J. (1996). New constraints between kinetic parameters explain the (un)identifiability of enzymatic rate constants. *Biotechnology and Bioengineering*, 52(3):433–7.
- Swann, W. H. (1969). A survey of non-linear optimization techniques. *FEBS Letters*, 2:S39–S55.
- Thompson, S. (1986). Biotechnology-Shape of things to come or false promise? *Futures*, 18(4):514–525.
- Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Wiley-VCH Verlag GmbH, New York.
- Tufvesson, P., Jensen, J. S., Kroutil, W., and Woodley, J. M. (2012). Experimental determination of thermodynamic equilibrium in biocatalytic transamination. *Biotechnology and Bioengineering*, 109(8):2159–62.
- Tufvesson, P., Lima-Ramos, J., Jensen, J. S., Al-Haque, N., Neto, W., and Woodley, J. M. (2011). Process considerations for the asymmetric synthesis of chiral amines using transaminases. *Biotechnology and Bioengineering*, 108(7):1479–93.
- Vanhaute, W. J., Vandenberghe, S., Scheerlinck, K., De Baets, B., and Verhoest, N. E. C. (2012). Calibration of the modified Bartlett-Lewis model using global optimization techniques and alternative objective functions. *Hydrology and Earth System Sciences*, 16(3):873–891.
- Versyck, K. J., Bernaerts, K., Geeraerd, a. H., and Van Impe, J. F. (1999). Introducing optimal experimental design in predictive modeling: a motivating example. *International Journal of Food Microbiology*, 51(1):39–51.
- Willner, I. and Mandler, D. (1989). Enzyme-catalysed biotransformations through photochemical regeneration of nicotinamide cofactors. *Enzyme and Microbial Technology*, 11(8):467–483.

- Zavrel, M. (2009). *Model-based Experimental Analysis of Enzyme Kinetics in Aqueous- Organic Biphasic Systems*. PhD Thesis, Rheinisch-Westfälischen Technische Hochschule Aachen.
- Zavrel, M., Kochanowski, K., and Spiess, A. C. (2010). Comparison of different approaches and computer programs for progress curve analysis of enzyme kinetics. *Engineering in Life Sciences*, 10(3):191–200.