Multi-enzyme Process Modeling

Andrade Santacoloma, Paloma de Gracia

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Multi-enzyme Process Modeling

Ph.D. thesis
Paloma de Gracia Andrade Santacoloma

Center for Process Engineering and Technology
Department of Chemical and Biochemical Engineering
Technical University of Denmark
March, 2012
Summary

The subject of this thesis is to develop a methodological framework that can systematically guide mathematical model building for better understanding of multi-enzyme processes. In this way, opportunities for process improvements can be identified by analyzing simulations of either existing or potential process configurations operated under different conditions. In these cases, process engineering, enzyme immobilization and protein engineering are presented as fields that can offer feasible solutions for better process configurations or biocatalyst modification to enhance actual process implementation, especially at an industrial level.

Multi-enzyme processes are characterized by a high degree of complexity due to the mixture of enzymes that catalyze several reactions. Therefore, it is necessary to understand how enzymes act in a coordinated and interactive way, and also how enzymes are affected (in a positive or negative way) by the presence of other compounds in the media.

In this thesis the concept of multi-enzyme in-pot term is adopted for processes that are carried out by the combination of enzymes in a single reactor and implemented at pilot or industrial scale. In order to understand the difference between multi-enzyme processes, a number of concepts are discussed in the second chapter of this thesis and has also been published as a review. Furthermore, a classification of multi-enzyme processes is suggested to clarify the ambiguous definitions found in the scientific literature.

Reliable mathematical models of such multi-catalytic schemes can exploit the potential benefit of these processes. In this way, the best outcome of the process
can be obtained understanding the types of modification that are required for process optimization. An effective modeling formulation of these processes can be achieved by applying a methodological framework. In this way, a systematic procedure, guidance and documentation are provided to support the modeler.

The methodological framework developed here brings many benefits to multi-enzyme process modeling. This framework identifies generic features of the process and provides the information required to structure the process model by using a step-by-step procedure with the required tools and methods. In this way, this framework increases efficiency of the model development process with respect to time and resources needed (fast and effective model development). Furthermore, this methodology incorporates state-of-the-art methods and provides background and insight into their applications for model development purposes.

The methodological framework, which comprises five steps, is the main result of this thesis. The novel feature of this methodology is the emphasis on the multi-enzyme process concepts that is introduced in all steps. In this way, the most relevant and necessary modeling issues can be precisely identified in order to achieve reliable mathematical structures of the processes. In the same way, specific mathematical techniques, for model quality evaluation such as uncertainty and sensitivity analyses, are included in this methodology. Multi-enzyme process modeling is tremendously benefited with the introduction of these analyses which mark a big difference in the formulation of reliable models for the multi-enzyme processes. In this way the model parameters that drives the main dynamic behavior can be identified and thus a better understanding of this type of processes.

In order to develop, test and verify the methodology, three case studies were selected, specifically the bi-enzyme process for the production of lactobionic acid, the bi-enzyme process for the production of N-acetyl-D-neuraminic acid, and the tri-enzyme process for the production of 1-phenylethylamine. Furthermore, different capabilities of the methodology are developed due to the valuable contributions of each case study. In this way, the methodology was also proven to be useful for a fast model formulation of multi-enzyme processes. Additionally, programming codes were developed using MATLAB (The Mathworks, Natick, MA) which were also used as computational tools to support the implementation, solution and analysis of all the mathematical problems faced in the case studies.
Denne afhandling har formålet at udvikle en metodisk rammme, der kan systematisk bygge matematiske modeller til en bedre forståelse af multi-enzymatiske-processer. Mulighederne for at forbedre processerne kan identificeres ved at analysere simuleringerne af enten eksisterende eller potentielle proceskonfigurationer. I simuleringerne er procesteknikken, enzymimmobiliseringen og proteinteknikken præsenteret som muligheder, der kan tilbyde løsninger for en bedre konfiguration af processerne. Eller som modifikation af biokatalysatore for at forbedre implementeringen af aktuelle processer, især på et industriel niveau.

Multi-enzym processer karakteriseres for at have en høj grad af kompleksitet. Dette er på grund af at blandingen af enzymer, kan katalyserer flere reaktioner i en enkelt reaktor. Det er derfor nødvendigt at forstå, hvordan enzymerne arbejder på en koordineret og interaktive måde og hvordan enzymerne påvirkes (enten en positiv eller negativ måde) ved tilstedeværelsen af de andre enzymer og komponenter i det samme miljø.

I denne afhandling er begrebet multi-enzym in-pot anvendt for processer, der udføres ved en kombination af enzymer i en enkelt reaktor og gennemføres på pilot eller industriel niveau. For at forstå forskellen mellem multi-enzym-processer, diskuteres en række begreber i det andet kapitel af denne afhandling. Dette er også blevet udgivet som er review. Endvidere er en klassificering af multi-enzym processer foreslået for at tydeliggøre de tvetydige definitioner, der findes i litteraturen.

Pålidelige matematiske modeller af disse multi-katalytiske ordninger kan udnytte de potentielle fordele af disse processer. På denne måde kan det bedste resultat af processen opnås ved at forstå de typer af ændringer, som kræves for
procesoptimering. En effektiv evaluering af disse processer opnås ved at anvende en metodisk række som giver en systematisk vej til modellering, struktur, vejledning, dokumentation og støtte til modelarbejderen.

De metodologiske rammer udviklet her giver mange fordele til multi-enzym proces modellering, fordi dette identificerer generiske funktioner af processen, der giver de nødvendige oplysninger for at strukturere procesmodellen. Dette opnåes ved hjælp af en trin-for-trin procedure med de nødvendige værktøjer og metoder. Denne metode forøger effektiviteten af model udviklingsprocessen med hensyn til tid og ressourcer (hurtig og effektiv model udvikling). Desuden indeholder denne metodologi state-of-the-art metoder og giver derudover også baggrund og indsigt af disse anvendelse til modeludviklingens formål.


For at udvikle, teste og verificere metoden er tre case-studier valgt, mere specifikt: bi-enzym proces til produktion af lactobionic acid, bi-enzym proces til produktion af N-acetyl-D-neuraminic acid, og tri-enzym proces til produktion af 1-phenylethylamine. Desuden er forskellige kapabiliteter i denne metode udviklet på basis af de værdifulde bidrag der kommet fra hvert enkelt case studie. I denne sammenhæng har metoden også vist sig at være nyttig for en hurtig model formulering af multi-enzym processer. Derudover er programmeringen udviklet ved hjælp af MATLAB (The MathWorks, Natick, MA), dette program anvendes også som beregningsværktøj til implementering, løsning og analyse af de matematiske problemer i case-studierne.
This thesis is submitted in partial fulfillment of the requirements for acquiring a PhD degree at the Technical University of Denmark (DTU). The research has been carried out from November 2008 to February 2012 and was financed by DTU. This thesis was prepared at the Center for Process Engineering and Technology (PROCESS) at the Department of Chemical and Biochemical Engineering under the main supervision of Professor John M. Woodley and co-supervision of Associate Professors Krist V. Gernaey (PROCESS) and Gürkan Sin (Computer Aided Process Engineering Center - CAPEC - DTU).

During the course of the work presented in the thesis, a number of people have provided their help and support, for which I am very grateful. First of all, I would like to thank my supervisors for their motivation, guidance, and inputs during the fruitful discussions I have had with them. Special thanks to John not only for excellent academic support (to the point of making me believe in the project) but also for the wonderful words of wisdom and unconditional support that have enabled me to grow as a person.

I don't have enough words to express how much pleasure it was for me to work at PROCESS not only for its excellent academic research which provides the perfect environment for fruitful discussions but also for the wonderful people with whom a lot of social activities were successfully developed (happy endings always). For sure without your participation none of these activities could have been possible and I really hope the same dynamic is kept for ever.

I would like to thank Carlos for the unconditional support and love that he gave me during all phases of this work. Also, thanks for encouraging me to develop good habits of reading books and playing music. These two fields have been the
greatest combination to put my head out of the academic work for a bit.

Special thanks to my Colombian friends, Naty and Oscar, because we have created our own family here in Denmark for some years now. I am really grateful for their personal and academic support throughout all phases of my life.

Finally the most sincere thanks to my family, specially to my sister because having her close is simply fantastic, and to my mother who is still showing me from a distance how wonderful life is.

Paloma Andrade Santacoloma
Paloma Andrade Santacoloma
Lyngby, February 2012
List of publications

  Type: Full journal article

  Type: Oral communication

  Type: Full conference paper publ. in proceeding/book and Oral communication

  Type: Full journal article


Type: Full conference paper publ. in proceeding/book and Poster communication


Type: Full conference paper publ. in proceeding/book and Poster communication


Type: Invited conference contribution


Type: Oral communication


Type: Poster communication


Type: Poster communication
Acknowledgements

I thank Dr. Wouter Van Hecke and Prof. Herman Van Langenhove (Ghent University - Belgium), for providing the experimental data used in the development of the first case study of this thesis which was related to the bi-enzyme production of lactobionic acid.

I also thank PhD candidate Katrijne Cierkens and Prof. Ingmar Nopens (Ghent University - Belgium), for the collaboration we had at the end of my project with the conceptual understanding of optimal experimental design.
Contents

Summary i
Resumé iii
Preface v
List of publications vii
Acknowledgements ix

1 Introduction 1
  1.1 Overview of multi-enzyme process modeling 2
  1.2 Brief description of the overall framework 3
  1.3 Motivation 6
  1.4 Objectives 6
  1.5 Thesis Outline 7

I Multi-enzyme process concepts and modeling methodology 9

2 Multi-enzyme processes 11
  2.1 Multi-enzyme ex-vivo processes 14
  2.2 Multi-enzyme in-pot processes 20
  2.3 List of symbols 23

3 Methodology for the model-building of multi-enzyme ex-vivo processes 25
  3.1 Introduction 25
## CONTENTS

3.2 Methodological framework ........................................ 26  
3.3 Step 1. Modeling objective ..................................... 27  
3.4 Step 2. Information ............................................. 31  
3.5 Step 3. (Re) Formulation ....................................... 36  
3.6 Step 4. Calculations ........................................... 41  
3.7 Step 5. Experimentation ....................................... 46  
3.8 List of symbols and nomenclature ............................... 49  

4 Mathematical methods for model analysis ......................... 51  
4.1 Model implementation and dynamic simulation .................. 52  
4.2 Identifiability of parameters ................................... 53  
4.3 Parameter estimation, confidence intervals and correlation matrix 55  
4.4 Uncertainty analysis: Monte Carlo method ...................... 58  
4.5 Sensitivity analysis: Morris Screening method ................. 62  
4.6 Optimal experimental design for parameter estimation (OED-PE) 66  
4.7 List of symbols .................................................. 67  

5 Multi-enzyme process improvements ............................... 71  
5.1 Process Engineering ............................................. 72  
5.2 Enzyme immobilization technology .............................. 80  
5.3 Protein Engineering ............................................ 81  
5.4 List of symbols and nomenclature ................................ 83  

II Case studies .......................................................... 85  

6 Introduction to case studies ......................................... 87  
6.1 List of symbols ................................................... 94  

7 Case study 1: Bi-enzyme production of lactobionic acid ........... 95  
7.1 General description ............................................. 95  
7.2 Step 1: Modeling objective ..................................... 96  
7.3 Step 2: Information ............................................. 96  
7.4 Step 3: Formulation ............................................ 98  
7.5 Step 4: Calculations .......................................... 102  
7.6 Conclusions .................................................... 113  
7.7 List of symbols and nomenclature .............................. 114  

8 Case Study 2: Bi-enzyme production of Neu5Ac .................... 117  
8.1 General description ............................................. 117  
8.2 Step 1. Modeling objective .................................... 118  
8.3 Step 2. Information ............................................ 118  
8.4 Step 3. Formulation .......................................... 121  
8.5 Step 4. Calculations .......................................... 123
## CONTENTS

8.6 Step 3. (Re) Formulation ........................................ 128  
8.7 Step 4. Calculations ................................................ 129  
8.8 Step 5. Experimentation ............................................ 130  
8.9 Conclusions .......................................................... 130  
8.10 List of symbols and nomenclature ............................... 131  

9 Case Study 3. Tri-enzyme production of 1-phenylethylamine 133  
9.1 General description .................................................. 133  
9.2 Step 1: Model objective ............................................ 134  
9.3 Step 2: Information .................................................. 134  
9.4 Step 3: Formulation .................................................. 137  
9.5 Step 4: Calculations .................................................. 141  
9.6 Conclusions .......................................................... 146  
9.7 List of symbols and nomenclature ............................... 147  

10 Discussion and conclusions ........................................ 151  

11 Suggestions for future work ........................................ 155  

A Reactions rate expression of one and two substrates 159  
A.1 Nomenclature .......................................................... 163  

B General programming schemes implemented in MATLAB 165  

Bibliography ............................................................ 173
Chapter 1

Introduction

Successful application of biocatalysis has been demonstrated in several fields, mainly in the pharmaceutical sector but also in other sectors where new opportunities are now arising in the synthesis of lower value chemicals and bio-fuels [1]. The exquisite selectivity of biocatalysts under mild process conditions is the main reason why interest today in biocatalysis is experiencing significant growth at both an academic and industrial level. Furthermore, the necessity for ‘greener’ processes is pushing the chemical industry to find more innovative alternatives in order to fulfill market demands, maintaining or even improving product quality [2-4]. The idea of these emerging technologies is the generation of more sustainable processes (being based on either the principles of green chemistry and/or the use of renewable resources) and they provide many interesting opportunities for biocatalysis [5-7]. However, synthetic sequential reactions are often required in a process to generate a desired product. One consequence is that when implementing a single biocatalytic step into such a sequence it is often found that the conditions for that single step, while very favorable for that step itself, are frequently different from the others. This raises an obvious question about whether it would be possible to catalyze several, if not all of the steps, using enzymatic methods. In nature that question has already been answered since ‘cell factories’ do this with great efficiency [8]. However at an industrial level, the concentration of components is too low. Essentially this results in major process investment in the product recovery.
Interestingly, the processes designed by a combination of enzymes in a single reactor are also having great impact in the synthesis of different compounds using e.g. cascade reactions, integrated deracemization and cofactor recycling [8 11]. In these cases, the catalytic activity of all enzymes working together is exploited. In this way a intermediate product which is produced from a substrate can be used by the next enzyme and so on, in cascade, parallel or network reactions. As a further advantage, separation and purification steps of intermediate products are no longer required because of the unit integration in the process. Actually, promising future applications are envisaged for multi-enzyme processes which can be considered as the next-generation of biocatalytic applications [4, 8, 12, 13]. Hence multi-enzyme processes are today at the point to be fully exploited taking advantage of developing fields such as process engineering, enzyme technology and protein engineering. These fields specifically aim to improve biocatalyst characteristics and to overcome process/biocatalyst limitations that can be inherent to a mixture of enzymes [14 17].

1.1 Overview of multi-enzyme process modeling

Historically the term multi-enzyme has most commonly been used to describe the group of bioreactions that are involved in the metabolic mechanism of living microorganisms. In this case, the term multi-enzyme in-vivo processes is given to the group of enzymes that are inside of a cell. This means that the activity of each enzyme is regulated by the cell according to the metabolism requirements. Hence what happens if a process is carried out by a mixture of free-enzymes in a single reactor? It must also be called as multi-enzyme process, specifically multi-enzyme ex-vivo process. Consequently a classification of this type of processes is necessary in order to define the main characteristics. For that reason Chapter 2 presents a review on the different types of multi-enzyme processes that have been reported to date. Likewise a classification of multi-enzyme processes is suggested based on the particular characteristics of each process (see Figure 2.2). From the different types of multi-enzyme processes that are distinguished in the classification, this project is specifically focused on investigating the multi-enzyme ex-vivo processes (more specifically the multi-enzyme in-pot process) (see Chapter 2 for further details). For simplicity, in this thesis, the term multi-enzyme process is used to refer to a multi-enzyme ex-vivo process. Hence the concepts and ideas developed in this thesis involves neither mechanistic nor modeling-based understanding of processes within a microorganism.

Through most of the biocatalyst developments, enzymes have been studied as individual compounds which are isolated from the natural environment due to their specificity in a given reaction, and not so much as a set of complementary
1.2 Brief description of the overall framework

In general multi-enzyme processes contain a high degree of complexity. Hence realistic behavior interpreted by models is limited by assumptions made about the real process. In fact what it is interesting to understand about multi-enzyme process is how enzymes act as a coordinated and interactive system, and how enzyme activities can be affected (in a positive or negative way) by the presence of other compounds. Consequently, it is believed that the combination of enzymes produces further effects in the full multi-enzyme model (if compared with single enzyme reactions). In those cases, an addition or reduction of kinetic parameters in the model structure is to be expected. However, it has been experienced that in order to have a good picture of a multi-enzyme process model, the individual reactions (enzyme kinetics) need to be understood first. Hence a decomposition of the full multi-enzyme system is required as a starting point for the model formulation. Afterwards the reliability of all kinetic parameters in the full multi-enzyme process model is evaluated based on sensitivity analysis which can quantify and identify the model parameters that are significant or insignificant in the suggested model structure at the particular operating conditions for each experiment. However a mathematical model is built not only by combining equations, but also requiring a sequence of decision steps and analyses (statistical tests) that can quantify the reliability of the model structure. Therefore, a systematic strategy is required to make this job tractable and straightforward for different multi-enzyme process configurations. Furthermore, the methodological approach must be accompanied by computational tools that facilitate and promote the mathematical model implementation and rapid analysis.

In the following subsection a general description of the methodological framework, developed here, is introduced. A detailed description is found in Chapter 3.

1.2 Brief description of the overall framework

In order to define the structure of a mathematical model, a number of different considerations must be carefully discussed. The formulation of any kind of model is done in an iterative manner, because every decision made in one step can affect the decisions in the subsequent steps and therefore, model reformulations are frequently required. This is why a methodological framework is considered as
a powerful achievement for the formulation of multi-enzyme process models. The development of a step-by-step procedure with required tools and methods provides a systematic way of modeling, a structure, guidance, documentation and support to the modeler. Consequently, this aims to be a step towards standardization of model development efforts.

Figure 1.1 shows a summarized work-flow for multi-enzyme process modeling. In general, the model structure depends on the relationship between reaction and process characteristics (see Chapter 3). For a multi-enzyme process, this evaluation is critically important to achieve a better understanding of the process and to achieve useful process model applications.

The methodological framework developed here collects and adapts general modeling methodologies and model analyses. This means that in principle it can be used for any type of process. However in some stages of the methodology specific information is provided for the description multi-enzyme processes, especially in the model formulation. This methodology also incorporates state-of-the-art methods and provided background and insight into their application for model development purposes. In this manner this approach becomes particular and innovative to multi-enzyme processes. Furthermore the model quality techniques here suggested have not been applied before to multi-enzyme processes and they are found relevant in the evaluation of the model structure and reliability of the model.

The methodological framework consists of the following steps in order to build a reliable model: model objective, information, (re) formulation, calculations, and experimentation.

- **Step 1. Modeling objective**: an important issue to address is the formulation of the modeling objective because it is directly related to the model complexity. Hence, models with less phenomena details can be generated for general structural process understanding. In this case, limited output capabilities of the model are expected and therefore less accurate input information is required. In contrast, validated models with high detailed phenomena description can be used for process optimization, controller performance evaluation and prediction. However a more complex model structure can be required, as well as good quality input information in order to obtain reliable results.

- **Step 2. Information**: compilation of the information about the system such as characteristics and experimental data reported in literature. This builds the basis for the model formulation. This compiled information is extremely useful for the generation of candidate models according to kinetic expression formulations and process designs.
1.2 Brief description of the overall framework

Figure 1.1: An overall methodological framework to guide the model-building of multi-enzyme processes

- **Step 3. (Re) Formulation:** having a basic knowledge about the system, one can start selecting relevant information about the enzymatic reactions and process operation. In this part, mathematical considerations are formulated based on first principles developing equations to describe reaction rates and mass balances for the desired system.

- **Step 4. Calculations:** computational implementation of the mathematical model is required to study and understand the dynamic behavior of the process when it is exposed to different changes in the input variables. If experimental data are available, such information can be used for model calibration. Furthermore, uncertainty and sensitivity analyses can be performed in order to identify critical factors affecting the system. Based on the results, optimal experimental design can be formulated if some preliminary experimental data are available and there is opportunity to carry out more experiments. Otherwise model reformulation might be considered again. In this case, the modeling procedure must return to the formulation block i.e. Step 3 in the framework (see Figure 1.1).

- **Step 5. Experimentation:** experimental data are always required when models need to be validated for a defined modeling objective e.g. optimization, control, prediction. Results from Step 4 can be used to formulate an experimental design where specific highly relevant variables are investigated. Therefore complementary information can be obtained and used for model calibration as mentioned in Step 4.
1.3 Motivation

Bioprocesses are attracting increasing interest at industrial level. One consequence is that there is a growing necessity to apply more detailed design, optimization and control of these types of processes. In order to achieve such objectives, mathematical models are required in most cases since they offer a quick way to understand, implement and improve the process.

For multi-enzyme processes, one of the main modeling benefits is to understand the interaction between enzyme, substrates, and products in a single reactor. This process understanding can be achieved by using computational tools where a large number of simulations of the process can be performed. In this way, different scenarios and conditions can theoretically be studied in order to identify potential process implementations. Furthermore time and effort in the laboratory can be minimized since only promising process configurations and conditions need to be carried out experimentally monitoring the information that is necessary to improve the model and process.

In the scientific literature there are few mathematical models developed for multi-enzyme processes and no evidence of a systematic way of getting reliable models. Consequently the lack of mathematical models makes it difficult to move forward with the development of these types of processes, especially to scale them up. Even though most of the multi-enzyme processes are still implemented at laboratory scale, they are seen as building-blocks for this project since they provide useful and important information to develop a methodological framework. In this case, relevant case studies provide knowledge, experience and insight that can be reused each time to improve and tune the methodology in order to achieve a more generic approach.

1.4 Objectives

As mentioned in the motivation, there is a need for reliable mathematical models that can describe the behavior of multi-enzyme processes with enough accuracy and can be suited for optimization, control or prediction of a desired process. Hence the key objective of this project is to establish a methodological framework that can systematically drive the model-building for multi-enzyme processes. The final methodology must be able to guide not only the model formulation (i.e. kinetics and mass balances based on first engineering principles) but also the mathematical implementation of the model and further techniques for the assessment of model quality (e.g. uncertainty and sensitivity analyses).
1.5 Thesis Outline

A further objective of this project is to choose relevant case studies in order to validate and exemplify the step-wise methodology.

1.5 Thesis Outline

So far this chapter gave a brief overview about the work developed in this PhD project. Here the general problem statement was presented highlighting not only the benefits of multi-enzyme processes but also the limitations and challenges that are necessary to consider in order to successfully understand and implement multi-enzyme processes. In the same manner, the motivation and therefore the objectives for this project were also stated in this chapter. Apart from the introduction, final conclusions, future work, references and appendices, this thesis has been organized in two main parts covering the contributions in multi-enzyme processes. The first part covers from Chapter 2 to Chapter 5, and the second part covers from Chapter 6 to Chapter 9.

The aim of the first part of this thesis is to provide understanding of the multi-enzyme processes, including not only the key concepts, but also the required steps for model-building. This first part is also considered as the pillar for the second part of this thesis which compiles the case studies selected to build, test and verify the methodology developed in this thesis. As mentioned previously, the 'multi-enzyme' term has been ambiguous in literature and therefore Chapter 2 suggests a classification of multi-enzyme processes based on what is reported in the scientific literature. Since the primary focus of the work is the methodological framework for modeling of multi-enzyme processes, Chapter 3 is devoted to a step-by-step establishment of the methodology, as briefly illustrated in Figure 1.1. In a similar manner, Chapter 4 is dedicated to explain and suggest mathematical methods that are specifically required to develop the calculation steps in the methodology. This block mainly contains the methods for implementation and simulation of the model, identifiability analysis, model calibration, uncertainty analysis, global sensitivity analysis and optimal experimental design. With the information collected during the whole modeling procedure and the model analysis itself, limitations, which make industrial application of a given process infeasible, can be identified (e.g. low reaction rates, enzyme inhibition/deactivation and solubility of components among many others). In this manner, Chapter 5 presents how fields such as process engineering, enzyme immobilization and protein engineering can offer feasible process configurations or biocatalyst modification that could solve the multi-enzyme process limitations to enhance real process implementation especially at an industrial level.

The aim of the second part is to provide not only the results of different case
studies but also to explain how they contribute in the development of the methodology. Hence, Chapter 6 gives an overview of all the case studies including a summary which shows the type of information that was found in the scientific literature and the calculations that were performed following the steps described by the methodology developed in this thesis. In this way, Chapter 7 develops a model for the production of lactobionic acid using two enzymes in one pot. In this case, the first enzyme is used as the main biocatalyst to synthesize the desired product and the second enzyme is added to the process to help in the recycle of the redox mediator that is required to promote this reaction (see Figure 7.1). Chapter 8 shows the analysis of the model for the production of N-acetyl-D-neuraminic acid (Neu5Ac) also using two enzymes. In this case both enzymes are necessary to synthesize the main product in a cascade. This means that the first enzyme is responsible for generating an intermediate product that is used by the second enzyme to achieve the final product (see Figure 8.1). Chapter 9 develops a mathematical model for the analysis of the production of 1-phenylethylamine. This process is carried out by three enzymes. The first enzyme drives the main reaction to produce the desired product, and then a second enzyme is added to remove the byproduct produced from the first reaction. However, the second enzyme is cofactor dependent and then a third enzyme is added to the process in order to recycle the cofactor (see Figure 9.1).

Finally this thesis concludes with Chapters 10 and 11, which present the most significant conclusions and possible future directions for research. Complementary material is presented in appendices A and B.
Part I

Multi-enzyme process concepts and modeling methodology
Multi-enzyme processes are characterized by using two or more enzymes which catalyze and drive a group of reactions in a defined pathway via a cascade, parallel or network configuration \[18\]. Some basic reaction structures are illustrated in Figure 2.1. In the first case (Figure 2.1a), a cascade is carried out when the product of a first reaction is used by another enzyme to obtain a product. Parallel reactions (Figure 2.1b) can be found when two or more enzymes compete for the same substrate to obtain different products. Another case of parallel reactions is found in the degradation of complex substrates. In that case several enzymes attack the substrate at a certain chemical group in the structure. In the third case (Figure 2.1c), a network reaction is found if there are components that directly connect two or more reactions in the process in order to drive all reactions to a desired side. A common example is seen in cofactor regeneration.

Historically, the term ‘multi-enzyme’ has most commonly been used to describe the metabolic activity of living microorganisms using established, modified, or de-novo pathways, in the context of biosynthesis \[19, 20\]. However for organic synthesis, multi-enzyme processes outside the cell hold particular promise, since individual enzyme expression and regulation may be decoupled from the metabolic network. In this way at least two isolated enzymes can be combined in an optimal way, driving the synthesis towards a primary product. It must also be clarified that using this concept, there is no attempt to reproduce the metabolic networks that are involved inside a microbial cell. On the contrary,
Multi-enzyme processes

Figure 2.1: Basic reaction schemes for multi-enzyme processes, (a) cascade or series reactions, (b) parallel reactions, (c) network of reactions

the cell metabolism is taken here as inspiration for multi-enzyme ex-vivo processes in order to understand the enzymes behavior in the presence of other enzymes and compounds.

Figure 2.2: Classification of multi-enzyme processes

A hierarchical classification of multi-enzyme processes is shown in Figure 2.2. In the first division, multi-enzyme processes are distinguished as multi-enzyme in-vivo and multi-enzyme ex-vivo processes. The first case refers to all enzymatic reactions carried out by the cell in order to keep the physiological role, taking into account the self regulatory mechanisms, as well as the diffusion effects that exist through the cellular membrane. The second case refers to the reactions that can be carried out by the combination and coupling of free enzymes in order to drive the process to the desired product. The advantage of this approach is
that the catalytic activity of each individual enzyme can be exploited in the process. Furthermore, a second division is shown for the multi-enzyme *ex-vivo* processes. These may be classified as artificial cells, multi-enzyme *in-vitro* processes or multi-enzyme *in-pot* processes. This classification is not only based on the evolution that this topic has had in its historical development, but also based on the process scale which lead to different phenomena description, and so the mathematical model formulation. In Sections 2.1.1, 2.1.2 and 2.2, each subcategory is described in a more detailed manner.

The versatility of isolated enzymes opens the possibility to explore different process configurations which could bring advantages with respect to the operability of the process depending on enzyme and compound compatibility. As shown in Figure 2.3, multi-enzyme processes can be operated with all the enzymes in a single reactor if they can match the reaction conditions. Alternatively, the process can be better operated in multiple reactors when some enzyme and compound incompatibilities are present. Clearly, the metabolic network of a cell cannot be decoupled into individual reactions and then it is not feasible to carry out the process in more than one reactor. Further discussion about each case is addressed in Section 2.1.

There are clear advantages when working with multi-enzyme *ex-vivo* processes in comparison with *in-vivo* processes. For instance Table 2.1 compares some characteristics related to the catalyst constraints, process modeling, monitoring and controllability among others between multi-enzyme processes. For example, modeling of *ex-vivo* processes is less complex than modeling the whole metabolic network of the cell, making it more likely to be reliable for the monitoring and
control of different variables during the process.

2.1 Multi-enzyme \textit{ex-vivo} processes

A process that uses an isolated enzyme as biocatalyst is considered as an \textit{ex-vivo} process. Consequently, the process that involves a combination of several isolated enzymes or free enzymes in a single system can then be labeled as a multi-enzyme \textit{ex-vivo} process. For organic synthesis, this approach holds particular promise since individual enzyme expression contributes in the system to drive a given transformation to the subsequent one until the desired product is obtained.

The application of multi-enzyme \textit{ex-vivo} processes has been explored since the 1970s [8, 12, 25], although it was not until 2003 that Bruggink and coworkers published their seminal review [8]. Interestingly, the application of promising cascade conversions was the central theme of this paper and this was supported by a compilation of different cases in which cascade catalysis was achieved by using multiple enzymes, multiple chemocatalysts, and/or a combinations of both, enzymes and chemocatalysts. In a more recent overview, Findrik and Vasic-Racki (2009) [12] again emphasize the importance of developing multi-enzyme processes. Furthermore, the necessity of using mathematical models and simulation tools was introduced by these authors as a means to achieve process understanding and optimization. To date, several processes have been successfully proven at a laboratory scale for \textit{in-situ} cofactor regeneration [26, 27], deracemization [9, 10], and in general cascade catalysis [8, 12, 28, 29]. Furthermore, a limited number of cases have also been reported at pilot and industrial scale [8]. Consequently, although relevant conversions using multiple enzymes have been proven as a valuable concept, only few scaled examples exist. Therefore there are still potential multi-enzyme process configurations to be exploited at an industrial level.

As previously mentioned, three different types of multi-enzyme \textit{ex-vivo} processes can be classified. Figure 2.4 shows a scheme for each multi-enzyme \textit{ex-vivo} concept. They are defined as a multi-enzyme artificial cell, a multi-enzyme \textit{in-vitro} system and multi-enzyme in-pot system. Historically, each category has contributed to the conceptual evolution of this topic.
Table 2.1: Comparison of multi-enzyme in-vivo and ex-vivo processes in a single reactor

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>In-vivo process [21, 22]</th>
<th>Ex-vivo process [18, 23, 24]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell/Biocatalyst constraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substrate inhibition</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Product inhibition</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Catalytic stability</td>
<td>Low</td>
<td>Higher (if immobilized)</td>
</tr>
<tr>
<td>Production cost</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Reaction constraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction reproducibility</td>
<td>Variable</td>
<td>Reproducible</td>
</tr>
<tr>
<td>By-products</td>
<td>Possible</td>
<td>Unlikely</td>
</tr>
<tr>
<td>Operating conditions</td>
<td>High dependence</td>
<td>High dependence</td>
</tr>
<tr>
<td>Process modeling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process understanding</td>
<td>Mechanism not fully understood</td>
<td>Possible</td>
</tr>
<tr>
<td>Reaction structure</td>
<td>Complex metabolic networks</td>
<td>Simpler reactions</td>
</tr>
<tr>
<td>Mathematical model interpretation</td>
<td>Difficult</td>
<td>Difficult</td>
</tr>
<tr>
<td>Potential process controllability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory control (T, pH, DOT)</td>
<td>Possible (if on line monitoring )</td>
<td>Possible</td>
</tr>
<tr>
<td>Supervisory control (concentrations)</td>
<td>Difficult (if intermediates)</td>
<td>Possible</td>
</tr>
<tr>
<td>Process monitoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On line measurements</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Intermediate products</td>
<td>Unlikely</td>
<td>Possible</td>
</tr>
<tr>
<td>Downstream processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product recovery</td>
<td>Possible (if extracellular product)</td>
<td>Possible</td>
</tr>
<tr>
<td>Recycling (cell/biocatalysts)</td>
<td>Possible</td>
<td>Possible (if immobilized)</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green/renewable process</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Current research activity</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>
2.1.1 Artificial cells

Early work on multi-enzyme processes was developed in so called ‘artificial cells’, which can be seen today as a particular case of enzyme immobilization by encapsulation [25, 30, 31]. In reported cases, a combination of multiple enzymes was microencapsulated within a polymeric membrane of 20 μm in diameter (i.e. cellular dimensions). A clear advantage of this technique is that enzymes are kept in solution inside the membrane and then the free movement of enzymes within the capsule facilitates the reactions because of the close contact with the compounds e.g. intermediates. Furthermore, diffusion of substrates and products through the permeable membrane is possible, while remaining impermeable to the larger enzyme molecules. A schematic representation of this type of system is shown in Figure 2.4a. Furthermore such a concept has been used for detoxifiers, immunosorbents, blood substitutes and drug carriers, among other medical applications [31, 32].

A specific example of such an application is illustrated by the production of 6-phosphogluconolactone, as shown in Figure 2.5. Here, two sequential reactions are carried out using two encapsulated enzymes, hexokinase (EC 2.7.1.1) and glucose-6-phosphate dehydrogenase (G6PD) (EC 1.1.1.49), respectively. Figure 2.4 gives a schematic idea of how the reaction takes place in the capsule. In this case, the NADPH formation was followed in time in order to analyze and compare the coupled reactions for the encapsulated system and for the soluble enzymes. Results demonstrated that the encapsulated system showed faster dynamic behavior, reaching steady-state more rapidly than the corresponding soluble system [33].
2.1 Multi-enzyme \textit{ex-vivo} processes

![Diagram](image)

\textbf{Figure 2.5:} Bi-enzymatic (hexokinase and Glucose-6-phosphate dehydrogenase) reaction for the production of \textit{\&}phosphogluconolactone using glucose as the substrate and glucose-6-phosphate as the intermediate product.

Today, the term ‘artificial cells’ is at the center of a wider scientific discussion, since the concept has changed over the years, together with philosophical considerations of what constitutes a living cell [34, 35]. However for synthetic chemistry, the multi-enzyme concept has evolved into the study of a feasible mixture of enzymes that can perform simpler reactions promoting in addition better understanding of enzyme behavior, and then higher possibilities for developing mathematical models which can contribute to process understanding and optimization [12, 27, 36],

\subsection{In-vitro processes}

To date, most of the knowledge on multi-enzyme processes has been obtained from multi-enzyme \textit{in-vitro} processes (see Figure 2.4b) [8, 12]. Although most of these reactions have been carried out at laboratory scale, the information obtained is actually of considerable value for reaction engineering since laboratory operation has promoted the analysis of individual enzyme and reaction characteristics. Multi-enzyme \textit{in-vitro} processes are mostly carried out in batch mode. Furthermore the main objectives to date have been to analyze the compatibility among the mixture of enzymes which are expected to work under similar environmental conditions, and the efficiency of the natural driving force in order to maximized the desired product.

Most applications of multi-enzyme \textit{in-vitro} processes have been developed for cascade catalysis and \textit{in-situ} cofactor regeneration (see Figures 2.1a and 2.1c). However, particular attention has been devoted to cofactor regeneration since many biochemical reactions require one or more cofactors to activate the enzymes and thus to carry out conversions effectively [11, 12]. For example in
cofactor-dependent biocatalysis, the proper addition of certain enzymes to the process promotes the efficient regeneration of the cofactor which can then be recycled multiple times in the same process. Consequently, a significant cost reduction can be achieved, since the use of many cofactors is still expensive for large-scale application. In a similar way, metal or chemical intermediate redox mediators, which are also widely used, can be better exploited since smaller amounts are then required in the system. As a result, cofactors are no longer compounds that limit the desired reaction and therefore higher conversions of the main substrates can be achieved [15].

One example is the intermediate redox regeneration reported for the production of lactobionic acid [27] (see also Chapter 7). In this reaction, the first enzyme promotes the generation of the main product and the second enzyme is added to regenerate the required cofactor, as illustrated in Figure 2.6. In this case, the first enzyme (flavocytochrome cellobiose dehydrogenase (CDH)(EC 1.1.99.18)) catalyzes the dehydrogenation of lactose to lactobiano-1-lactone, which is spontaneously hydrolyzed to lactobionic acid, and a second enzyme (laccase (EC 1.10.3.2)) allows the full reduction of the available oxygen to water. Furthermore, the double action of the redox mediator 2,2'-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) is exploited, regenerating the oxidation states of both enzymes.

![Figure 2.6: Reaction scheme for the production of lactobionic acid using lactose and oxygen as substrates, two enzymes (CDH and Laccase) and an intermediary redox mediator (ABTS) [27], (a) lactose; (b) lactobiano-lactone and (c) lactobionic acid (see also Chapter 7)]

In that work [27], a mathematical model was formulated to describe the kinetics of the process. Furthermore, the model allowed successful prediction of the system behavior carried out in a mini-reactor with integrated bubbleless oxygenation. Interestingly, a model-based graphical tool was used to determine the optimal process conditions when oxygen is used as a co-substrate. In this case, the formulated model contributed not only with a better process understanding, but also to suggest feasible process improvements.
2.1 Multi-enzyme ex-vivo processes

On the other hand, cascade catalysis is often required in many important synthetic routes [37-41]. One example is the transformation of D-methionine into L-methionine using four enzymes in a cascade [26]. In that case a mathematical model was formulated and validated. A further example is the synthesis of a non-natural carbohydrate (5-deoxy-5-ethyl-D-xylulose) from glycerol where four enzymatic steps were also carried out in in-vitro using a pH switch method which temporarily reduced the action of one enzyme in the system at a given time [42]. Several reports in the scientific literature offer overviews of specific types of reaction, and the relevant laboratory procedures. [4, 8, 12]

Figure 2.7: Reaction scheme for the production of optically-pure chiral amines employing two ω-transaminases with opposite stereopreference, lactate dehydrogenase (LDH) for pyruvate removal and glucose dehydrogenase (GOD) for cofactor regeneration. (a) S- and R-amine, (b) pyruvate, (c) D-alanine, (d) ketone, (f) L-alanine, (g) S-amine, (h) lactate, (i) glucose and (j) gluconolactone [10]

Another illustrative example is the deracemisation of ω-chiral primary amines to optically pure amines by the action of specific ω-transaminases (EC 2.6.1) [10]. Deracemisation is achieved in a one-pot, two-step procedure, as shown in Figure 2.7. In the first step, kinetic resolution of the chiral racemic amine is carried out by an ω-transaminase to yield an intermediate ketone and the residual optically pure amine. In the second step, the ketone intermediate is transformed into the amine by employing alanine as the amine donor and an ω-transaminase displaying the opposite stereo preference to that in the first step. In addition
in the second step, lactate dehydrogenase (LDH) (EC 1.1.1.27) is used to remove the pyruvate (by-product) in order to shift the reaction equilibrium to the product side. Here two procedures were applied. In the first procedure, the second enzyme was added to the system after the kinetic resolution of the first step was complete. At the end of the conversion, the optical purity of the final product was moderate. The reason for this was that the ω-transaminases also accept, to a certain extent, the opposite alanine enantiomer, meaning that the ω-transaminase of the first step also catalysed the amination reaction although at a reduced rate. In the second procedure, the conversion of the reaction was improved by introducing a heat treatment before starting the second step. In that way, the desired enantiomer was obtained with ee values >99%. From the point of view of multi-enzyme ex-vivo processes, the first case is clearly an interesting study to analyze since high interaction between the enzymes was found, while in the second case both enzymatic reactions were completely independent due to the heat treatment in the first step. Thus even though the conversions were carried out in a single reactor, multi-enzyme activity did not take place.

2.2 Multi-enzyme in-pot processes

The so-called multi-enzyme in-pot process describes a system where the synthetic reactions are carried out using two or more enzymes in a single reactor specifically at pilot/industrial scale. This implies that not only a batch reactor is considered, but also many other reactor options as well, including different operating modes and configurations for the process.

A multi-enzyme in-pot process is very well suited to run in an integrated fashion since the conditions in each reaction (i.e. media, concentrations of substrates and products, catalyst pH, catalyst temperature) are typically well matched. In the same manner as other multi-enzyme processes, the catalytic activity of all enzymes working together is exploited, and thus a substrate transformed to a first intermediate product can be used by another enzyme and so on, in a cascade, parallel or a network of reactions. As a further advantage, separation and purification steps of intermediate products are eliminated from the process [8]. Consequently, the multi-enzyme in-pot approach potentially leads to considerable process improvements such as a potential reduction in downstream processing and operating costs which can be significant at the industrial level.

To date, most enzymatic reactions have been operated in single reactors since the enzyme kinetics and operating conditions are well known for that single step. However, the combination of enzymes opens a window for the next generation of biocatalysis, not only as an innovative way of carrying out the process, but
2.2 Multi-enzyme in-pot processes

also such that different advantages can be exploited. This concept is illustrated in Figure 2.8. The first part (a) illustrates a process in single biocatalytic steps while in the second part (b) the sequence of biocatalytic reactions is carried out in a single reactor.

![Figure 2.8: Tri-enzyme process: enzyme 1 (○), enzyme 2 (□) and enzyme 3 (△). a) Process carried out in single biocatalytic steps. b) Multi-enzyme in-pot process to carry out the reactions in a single reactor.]

Given a feasible mixture of enzymes, the multi-enzyme in-pot processes can be analyzed to explore the optimal conditions of the process. This includes operating conditions, reactor design and process control [43].

In fact the operating conditions can be evaluated by analyzing the role that each enzyme has in the system. In multi-enzyme processes, enzymes can be divided into two groups: main enzymes and helper enzymes. The first group refers to those enzymes that are directly involved in the synthesis of the desired product. The second group refers to those enzymes that are added to the system in order to improve the characteristics of the main reactions. They must have higher activity and preferably use cheaper substrates than the main enzymes, and they must not be directly involved in the synthesis of the desired product [44]. As illustrated in Figure 2.7, the multi-enzymatic production of optically-pure chiral amine involves four enzymes, two ω-transaminases, lactate dehydrogenase and glucose dehydrogenase. Following the classification, the main enzymes are ω-transaminases, and the helper enzymes are LDH and GDH. The point with this classification is to find the perfect balance among enzyme activities where the productivity of the process is maximum compromising preferably the activity of the helper enzymes which are expected to have lower cost with high activity [44].

The reactor design can be evaluated according to the following characteristics: (1) the format that each enzyme has in the system (e.g. soluble, immobilized
or a mixture of both), (2) number of phases present in the system, and (3) reaction limitation (e.g., substrate and product inhibition). At the end, the selection is based on achieving appropriate process operating conditions that promote better enzyme compatibility, and therefore better performance of the process [45, 46].

In the same manner, process control must be taken into account since it is responsible for achieving and maintaining the optimal operating conditions. Furthermore, a robust control can effectively reject disturbances in the process. Process control is really important at all process scale levels. However, it becomes much more relevant at pilot-plant and industrial scale, especially if run in continuous mode. However, most of the reported cases of multi-enzymatic reactions are mainly at laboratory scale where little if any process control was applied. Nonetheless, a few exceptions are found at industrial scale such as the continuous production of L-amino acids (e.g., L-methionine, L-norleucine, and L-2-aminobutyric acid) [8].

As mentioned previously, a number of challenges need to be addressed in order to exploit all the benefits of this approach and one technique to assist in this development is by using computational tools to simulate the desired process under different conditions. However, as a basis for the simulation, a mathematical model needs to be built and solved, and then the question is whether such models are reliable enough to guide decision-making. Applying a systematic framework, the mathematical expressions are formulated while the final model structure will depend on the decisions taken during the analysis of the reaction and process characteristics (as shown in Figure 3.2 Chapter 3) that are related to the real process.

![Figure 2.9: Basic steps for process development](image)

In order to fully develop, implement and scale-up a multi-enzyme process some
fundamental steps are necessary [47]. These steps are illustrated in Figure 2.9. In this case, the old and new ways of developing processes are compared in order to show the advantages that are earned when using mathematical models. The main difference lies in the amount of experimental work that is required when implementing a process in the old fashioned manner. On the contrary, the number of experiments may be minimized when this development is combined with modeling. In the new approach, the model building also consumes time. However this is compensated with a reduction in the experimental time since this can be supported by the formulation of experimental design based on the model. In the same way, the model can be exploited to analyze potential scenarios that can be applied at larger scale. That is why a pilot plant implementation can be facilitated if compared with the old approach. At larger scale, recalibration of the model must be necessary. However, further advantages can be achieved when using the model for process control and optimization of an implemented process.

2.3 List of symbols

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6PD</td>
<td>Glucose-6-phosphate dehydrogenase</td>
</tr>
<tr>
<td>CDH</td>
<td>Flavocytochrome cellulbios dehydrogenase</td>
</tr>
<tr>
<td>ABTS</td>
<td>2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt</td>
</tr>
<tr>
<td>LDH</td>
<td>Lactate dehydrogenase</td>
</tr>
<tr>
<td>GDH</td>
<td>Glucose dehydrogenase</td>
</tr>
</tbody>
</table>
CHAPTER 3

Methodology for the model-building of multi-enzyme ex-vivo processes

3.1 Introduction

Mathematical process models are a particular form of representing what happens in reality, in a quantitative manner. The set of equations that are developed must have certain characteristics to describe the system according to the specific use of the model. In general, the main purpose of modeling is to gain insight into the behavior of systems in order to understand them, control them, or optimize them, among other applications [23, 48, 50].

For multi-enzyme ‘ex-vivo’ processes, a mathematical model can be the key to exploit the potential of this approach since the individual enzymes may still operate under unfavorable conditions, even though the overall process is optimal. This is well illustrated by the synthesis of non-natural carbohydrates from glycerol which can be produced by the use of e.g. four enzymes. In this
example, the activity of some enzymes was 'switched off' by increasing the pH from 4 to 7.5 and 'switched on' again by lowering the pH back again to 4 [42]. Interestingly, such a procedure enabled the successful production of the desired product keeping at the same time the stability of the enzymes even though some of them were not active at certain points in time during the overall process.

For the case of multi-enzyme in-pot processes, it is advantageous to formulate a mathematical model before proceeding with real implementation, because modeling gives an opportunity to evaluate the process feasibility, e.g., analyzing different scenarios such as the operating conditions and alternative reactor configurations. Thus, the effect on the performance can be analyzed in order to explore and clarify the advantages as well as the limitations of a given process [43, 51]. As a result, process feasibility can be proven conceptually ahead of experimentation and be used for obtaining more detailed information about the process and how to proceed with the implementation (e.g., formulating reliable experimental designs). In reality most of the mathematical models are formulated based on processes that are already implemented (at least at laboratory level). Consequently, there are some process characteristics and conditions that are known and defined for a given process. Some examples from literature are the production of lactobionic acid [27] and L-methionine [26], where mathematical models were formulated based on real process implementation and then successfully validated with available experimental data.

Models can be used for simulation as a relatively cheap option to investigate feasibility of improvements of existing processes, the scale-up of desired processes or the implementation of new in-pot syntheses [23, 36, 52]. Furthermore model simulation enables the integration of process understanding (e.g., dynamics, productivity, controllability, stability, etc.) and more fundamental and knowledge-based process design as opposed to empirical process design. One example to mention is the bi-enzyme modeling formulation and simulation in the production of optically pure lactone. In this case, 'windows of operation' were used as a tool in order to identify feasible economic scenarios to carry out the process [53]. Another example is the bi-enzyme modeling of aminotriol/aminodiol synthesis. In that case, the enzymes transketolase-transaminase (TK-TAm) were used and the effects of the TAm/TK activity ratio was analyzed by simulation to identify optimal process operation [54].

### 3.2 Methodological framework

Mathematical models are formulated and adapted according to the conditions where processes can be feasibly applied. In order to define the structure of a
3.3 Step 1. Modeling objective

mathematical model, a number of different considerations must be carefully discussed. The formulation of any kind of model is done in an iterative manner, because every decision made in one step can affect the decisions in the subsequent steps and therefore, model reformulations are constantly required before a validated model is obtained [43, 48].

In the introduction a general overview of the methodological framework has been presented (see Figure 1.1). Furthermore, this section shows a more detailed scheme and information of the systematic evolution of the methodology in which the same five suggested stages are kept i.e. 1) model objectives, 2) information, 3) (re) formulation, 4) calculations, and 5) experimentation (see Figure 3.1).

In the following sections a detailed description of each stage is given. The three first stages are mainly based on human expertise and literature, while step 4 is based on a combination of numerical methods and computational tools for solving the model formulation (computer work). However most of the computational implementation must be developed by the modeler and then some programming skills are required. Subsequently, the calculations follow basically the same routine for all cases. Step 5 is based on the procedures and analytical techniques that are required for real implementation of the desired process (laboratory work).

3.3 Step 1. Modeling objective

An often asked question about modeling is related to the model complexity. Indeed, it is an important issue to address before the model is formulated and implemented. The primary reason is that model complexity depends upon the model purpose. Hence, simple models can be generated for general structural process understanding with limited output capabilities and therefore requiring less accurate input information. In contrast, dynamic models to be used in process optimization, controller performance evaluation and prediction, typically require a more complex model structure. Additionally, good quality and quantity input information is required to obtain reliable results [43, 48].

Modeling objectives can be arranged in increasing order of complexity, thus (see also Figure 3.1):

- **To develop and understand the model structure**: this objective is possibly to analyze since the mathematical model formulation, suggested here, is based on first principles i.e. variables and parameters have physical meaning. In this way, the modeling procedure can start including
Figure 3.1: Schematic framework for multi-enzyme process modeling.

The methodology as described and the model is ready to be used.
the main phenomena that drive the process, and later on more complex phenomena can be added to achieve a more realistic process model. In this case, enzymes follow a defined reaction mechanism according to the type and number of the substrates that are involved in the reaction. Consequently they are the main structural units to be analyzed in the model. For multi-enzyme processes, the model structure analysis opens the opportunity to evaluate the terms in the model structure that can appear due to extra inhibition or disappear due to a fast consumption of the intermediate products. For this model objective, neither kinetic parameter values, experimental data, nor computational implementation are required. In this case, the model may be able to follow the multi-enzyme process behavior in a qualitative manner.

- **To simulate the process model:** building on the previous model objective, the computational implementation of a model structure gives the opportunity to analyze the dynamic process behavior through model simulation. In this case, the influence of the input variables and parameter values over the process can be explored, e.g. to analyze what happens when the input flow rate or kinetic rates are increased or decreased in the process. It gives a fast interpretation and a better idea about the expected behavior that the process can have under certain circumstances. For this model objective, parameter values and initial conditions of the process are required (obtained from literature if experimental data are not available). However, neither validation of the model structure nor experimental data are really required. As described, the model is able to follow the multi-enzyme process behavior in a quantitative manner, but not validated.

- **To simplify the process model:** Unless an enzyme follows a Michaelis-Menten kinetic, enzyme mechanisms are commonly described by complex terms with many kinetic parameters which describe the interaction between the enzymes and compounds in the reactions e.g. substrates and products (see Appendix A). Most kinetic parameters of enzymatic reactions are estimated using initial reaction rate data or progress curves obtained in batch mode for single enzymes. However the degree of parameter significance can change in the mathematical model with the integration of reactions, such as a multi-enzyme process. Consequently, model simplifications can be evaluated in order to reduce model complexity under given operating conditions keeping the reliability of the model.

- **To formulate an optimal experimental design (OED):** this objective for the model is highly recommended when a model needs to be validated. Random experimental data do not always offer the necessary information to validate a model and then experimental work must be focused on those variables that offer relevant information about the process dynamics. In this case, a defined objective for the OED is required (e.g. to
select the best model structure, to estimate parameters or to understand the process\cite{55}, as well as the initial conditions of the process. However, no validated (true) parameter values are required for the formulation of optimal experimental design. As a result, the obtained model is able to follow the multi-enzyme process behavior in a quantitative and validated manner.

- **To optimize the process**: sometimes a validated model is not enough to enable an optimal performance of a multi-enzyme process. Actually, the process conditions must be optimized in order to guarantee high productivity of the process, minimizing the cost. In this case, process constraints need to be identified as well as the type of parameters that can be adjusted to affect optimal performance e.g. equipment, kinetic parameters and operating procedures. A general optimization problem may be described as:

\[
\begin{align*}
\min/\max & \quad F_{\text{obj}} = f(x) \\
\text{Subject to} & \\
g(x) & \leq 0 \\
h(x) & = 0
\end{align*}
\]

where \(f(x)\) is the function to be minimized or maximized over the variable \(x\), and \(g(x)\) and \(h(x)\) are the set of constrains, equalities or inequalities, that also reduces the search space in order to find feasible solutions for the process. Notice that the constrains depends upon the region within which the model is formulated. A reliable model can be used as a powerful quantitative tool to optimize multi-enzyme processes and to lead the industrial decision making.

- **To control the process**: after finding optimal process conditions, those must track a trajectory during the process in order to maintain the same productivity and quality of the product. Furthermore, it is desired that controllers can also handle disturbances. Consequently, validated models under optimal operating conditions can be used for controller design, especially if they are based on the model structure such as inventory control, linear quadratic regulator (LQR) and model predictive control (MPC) \cite{43, 56, 57}. However, for the implementation of modern control such as LQR or MPC, a linearization of the model is required. In this cases, Taylor series expansion can be used as linearization method.

- **To predict process outputs**: reliable predictive capabilities of a model are desired in order to evaluate and optimize different scenarios in which the process can be operated. In this case, the model must contain as many
3.4 Step 2. Information

In order to facilitate model building, it is necessary to understand the function of the multi-enzyme process since this gives the basis to formulate an appropriate model structure in order to achieve the model objective. It definitively requires a deep literature search and discussion with experts in the field. As understanding of the full multi-enzyme system is not achieved easily, it is necessary to decompose the system in the individual enzymatic reactions in order to facilitate the information search. In that manner, reactions are analyzed individually and then the contribution of each single enzyme can be understood in terms of the full system.

This step in the methodology compiles all information that allows us to know more about the process. This step can be considered as one of the most important for modeling, even though mathematical calculations are not involved. In the same manner, for a person that is not familiar with a given case study, this step can consume much more time than the quantitative model development itself.

Having a clear idea of the process and phenomena that are involved, the mathematical translation becomes easier. Furthermore, a good scanning of the information can reduce some of the subsequent steps since parameter values or even experimental data can be found and then used in the analyses and calculation.
of the further steps. As mentioned, different sources must be considered to look for information. Scientific literature is the most common and accessible source. However, the help and discussion with experts in the field can offer valuable information and reduce the searching time.

Useful information can come from changing operating conditions of the process, preliminary kinetic parameter estimates, mass and energy transfer phenomena, enzyme mechanisms, types of reactors, uncertainty values of measurements and parameters, among others [24, 58]. As shown in Figure 3.1, a classification of the information is required to promote the process understanding and then the translation to the corresponding mathematical expressions. It can be divided in two tasks: the first task classifies the information related to the process characteristics. This then involves assumptions, control volumes, conservation equations, constitutive equations, and conditions (detailed information can be found in Section 3.5). As a support to the first task, the second task sorts the information according to the reaction and process characteristics which are directly related to reaction rates and mass balances of the process (detailed information of these two categories is presented in Section 3.4.1).

The amount of information that can be found in the scientific literature is extraordinary. However, the information is often presented in a sub-optimal way (e.g. inconsistent units and values for kinetic parameters, unavailable progress curves, unclear and incomplete reaction/process conditions). In this way, the scientific information is not helpful neither for model building nor for further applications. Consequently, the selection of reliable information becomes more difficult for particular case studies. Today, more effort is focused on developing a guide that can lead to the reporting of theoretical and experimental information of biocatalytic reactions [59]. In this way, the reliability of the information can be ensured since it can be presented in a more standardized way, and with a certain level of quality.

3.4.1 Reaction and Process considerations

In this section, relevant information to be considered for modeling of multi-enzyme processes is more highlighted, specifically, information related to the reaction and process characteristics. Several important issues are described below and illustrated in Figure 3.2. For a multi-enzyme process, this evaluation is critically important to achieve a better theory understanding of the process and to achieve useful modeling and process design.

The reaction considerations describe the key characteristics to understand how the interaction between enzymes and components can be interpreted for mod-
Figure 3.2: Reaction and process considerations for multi-enzyme process modeling (see details in Section 3.4.1)
Methodology for the model-building of multi-enzyme ex-vivo processes

eling. Furthermore, such information forms the basis for the formulation of reaction rates for the different enzymes that are involved in the multi-enzyme process. In this manner a preliminary idea of enzyme mechanisms and kinetic parameters, that can be expected when developing a model, may be obtained. Key information may be described as:

- **Knowledge about the compounds involved**: physical and chemical properties of the compounds involved in the reaction must be known and for such a task the use of existing databases facilitates the compilation of the available information [60]. Some of the properties can be density, water solubility, viscosity, boiling point, among others that could be relevant for multi-enzyme processes.

- **Structure of the reaction**: a graphical identification of all reactions in the multi-enzyme process is the basis for describing the final model structure. It includes the main reactions, secondary reactions, competitive reactions, reversible reactions and undesired reactions. For a single enzyme, reaction mechanisms are well developed and then they are included into the full model to describe the multi-enzyme process by combining the effect of the individual enzymes. In this way, the different possible reaction schemes are generated such as a cascade, parallel or network structures, as illustrated in Figures 3.2 and 2.1. In the model structure of a multi-enzyme process, additional terms can appear due to inhibitory effects or can disappear due to the fast consumption of intermediates. Furthermore, the importance of all the reactions must be analyzed because they increase the complexity of the model.

- **Interaction matrix**: this matrix is suggested to identify the different interactions that can exist between compounds and enzymes in the process. In this case, the reaction structure defined in the previous step is useful to visualize and classify those relationships that can happen with a higher degree of probability. Similar ideas about the interaction between compounds can be found in the scientific literature or from experimental experience in the laboratory. In order to build the matrix, the compounds involved in the process (i.e. substrates, intermediates, by-products, products, etc.) are arranged in rows (i.e. A, B, C, . . .), and the enzymes (E_i) are arranged in columns (for i = 1, 2, 3, . . .), as illustrated in Figure 3.2. In this way, the matrix is filled defining the relationship between each compound and enzyme in turn, i.e. (S) for substrate, (P) for product, (I) for inhibitor, or (X) when no interaction between one compound and one enzyme. This compiled information is extremely useful to make decisions about the relevant terms or kinetic parameters that must be added or removed from the reaction rate expressions and process model. For example, if a compound (D) of a secondary reaction inhibits one enzyme
in the process, there is a parameter $K_D^i$ that should be added to the reaction rate formulation. The position of the new term/parameter in the final expression is defined by the enzyme kinetic mechanism which shows how the compound inhibits the enzyme e.g. competitive, uncompetitive, non-competitive or mixed inhibition.

Similarly, the process considerations describe the key characteristics that can affect or modify the mass balances when formulating the model. In this manner a preliminary idea, of which process configurations and further phenomena can be expected in the model, is obtained. They may be described as:

- **Operating mode**: the operating mode is related to the liquid flow exchange characteristics in the process. The operating mode must be identified in order to have a clear idea of the terms that must be considered in the mass balances of the process (see Figure 3.2). For multi-enzyme processes, operating modes such as batch, fed-batch or continuous are used according to the process characteristics (e.g. inhibitory effects and transport limitations). However other operating modes can also be considered such as the pH switching [42]. In this case, the pH in the process was modified (increased/reduced) in order to activate or deactivate certain enzymes in the mixture and thus prioritize some reactions in the media at a different point in time.

- **Type of reactor**: the identification of the type of reactor is relevant to analyze the physical characteristics and constraints that can be present in the process, such as: transport phenomena, number of phases involved and the degree of mixing (see Figure 3.2). In this way, this characteristics can be included into the mass balances using the corresponding mathematical expression that described the phenomena. To date, mainly stirred tank reactors are used in the application of multi-enzyme process. However, membrane reactors and packed bed reactors have been also reported when process characteristics, such as substrate or product inhibition and immobilized enzymes, are presented in the process [27]. Good reactor design is a key issue to achieve multi-enzyme process improvements, and then it is also important to consider and analyze different type of reactors and configurations that can best handle the process limitations. Hence further details and suggestions about reactor selection for multi-enzyme processes are discussed in Section 5.1.1.

- **Component characteristics**: in addition to the physical and chemical properties of the components, it is also necessary to provide specifications such as purities of substrates, concentrations and amounts of cofactors, enzyme format (e.g. the whole cell, isolated, immobilized). This information
is part of the initial conditions of the process and then used for the simulation of the model. Furthermore, they contribute to the decision-making about the previous issues i.e. operating mode and type of reactor.

- **Process control**: in multi-enzyme processes, variables such as pH and temperature are often controlled during the process in order to reduce the influence that they produce on the dynamic of other variables and enzymes. For modeling, the controlled variables need to be identified in order to limit the capabilities of the model. In this case, they are included as assumptions of the model. Process control can be divided into two basic control layers [61] (more details in Section 5.1.2). The first is known as regulatory layer which controls variables such as pH, temperature, dissolved oxygen (DO). In this case simple controller design can be implemented. The second is known as supervisory layer which manages variables with more impact on the process, such as concentrations of the compounds. In this case more detailed controller design is required. For multi-enzyme processes, this issue is highly relevant especially to achieve process improvements therefore further details and discussion are later presented in Section 5.1.2.

### 3.5 Step 3. (Re) Formulation

There is certain information that needs to be defined for better understanding of the building model limitations. Hence, the model is a set of mathematical equations which aims to describe how a process performs in simulation. The quality of the model depends on how well the model represents experiments within the desired operating window. No matter the type of equations developed to describe the system, the procedure follows almost the same steps for the model formulation, and the difference is faced in the application of the numerical methods to solve the equations (see Section 4.1). In the case of multi-enzyme processes, one must take into account the following steps [48]:

#### 3.5.1 Model assumptions

These describe the particular characteristics in which the model is developed as well as the conditions in which the experiments are carried out in terms of the desired processes, i.e. time, space (e.g. perfect mixing), flow conditions (e.g. equal inflow, outflow), controlling mechanisms (e.g. chemical reaction, diffusion of mass), safe process ranges, physical limitations (e.g. water solubility of compounds, media compatibility, enzyme deactivation), physico-chemical
properties (e.g. constant density), dependences such as temperature, pH, compound inhibitions (e.g. product inhibition), types of inhibition (i.e. competitive, uncompetitive, noncompetitive) etc.

3.5.2 Boundaries and balance volumes

It refers to the selected volume in which the mass balances are formulated. This information helps to give clear idea on how the conservation equations are combined in order to build the process model. Modeling can be done in an incremental way, starting from small balance volumes and interconnecting them in order to describe a larger phenomenon, when required, e.g. the model of a single particle of an immobilized biocatalyst can be developed, and then it can be repeatedly used in order to describe a combination of several beads. Afterwards it can be used for example in the model of a continuous packed bed reactor (CPBR). This is also illustrated in Figure 3.3

![Figure 3.3: Different balance volumes for defining the limits for model construction (dashed line). (a) Balance volume for a unit of an immobilized enzyme and (b) Balance volume for two transversal sections in a PBR](image)

3.5.3 Conservation equations

The construction of mathematical models is based on conservation laws i.e. first engineering principles. In general, the modeling information that is collected, in Step 2 of the methodology, will lead to one type of model either dynamic or steady-state. For each balance volume selected for modeling, a mass balance is formulated. Depending on the problem complexity, the mathematical expression is described by a specific set of equations, changing from algebraic equations (AE’s) when time and spatial dependences are not considered, to ordinary differential equations (ODE’s) when only time dependence is considered,
Methodology for the model-building of multi-enzyme ex-vivo processes

and to partial differential equations (PDE’s) when both time and spatial dependencies are taken into account. Accordingly, the following classification for model structure is given in the scientific literature [48].

![Figure 3.4: Classification of the feasible mathematical structures for modeling multi-enzyme processes [48]](image)

Having defined one type of model according to the process characteristics, a clear idea of the mathematical problem can be visualized, and so the required numerical methods and for example MATLAB toolbox/functions for the solution.

- Dynamic models - Lumped parameter models

As illustrated in Figure 3.4 these types of model do not include spatial variation of the states in the balance volume. This means that mainly intensive variables are described in time (e.g. substrate/product concentrations).

The total mass balance is given as:

$$\frac{dM}{dt} = \sum_{j=1}^{p} F_j - \sum_{k=1}^{q} F_k$$  \hspace{1cm} (3.1)

where $p$ and $q$ represent the number of input and output streams, respectively.
In the same manner, the component mass balances can be generalized as

\[
\frac{dm_i}{dt} = \sum_{j=1}^{p} f_{i,j} - \sum_{k=1}^{q} f_{i,k} + g_i \quad i = 1, 2, ..., n \quad (3.2)
\]

where \( m_i \) is the mass hold-up of component \( i \), and \( g_i \) is the rate of generation or consumption of species \( i \).

Generally the generation term can be expressed as

\[
g_i = r_i V = 1, 2, ..., n \quad (3.3)
\]

Where \( r_i \) is the net reaction rate of the compound \( i \) and \( V \) is the reactor volume.

In multi-enzyme processes, both continuous and batch stirred tank reactors (CSTR - BSTR) follow this type of model structure. This means that an homogeneous mixture of all compounds is assumed.

**Dynamic models - Distributed parameter models**

As illustrated in Figure 3.4 these types of model do include spatial variation of the states in the balance volume. This means that intensive variables are described as a function of both time and position. This leads to a mathematical representation as PDEs in one, two or three spatial dimensions. The general differential conservation expression described in rectangular co-ordinates, thus becomes

\[
\frac{\partial \phi}{\partial t} = D \left( \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \right) - \left( \frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y} + \frac{\partial \phi}{\partial z} \right) + g \quad (3.4)
\]

Where \( \phi \) represent an extensive quantity in the balance volume which includes the intensive potential counterpart \( \varphi \) coming from the diffusion/convection term, and \( g \) is the generation and/or consumption term.

Plug flow reactors (PFR) and packed bed reactors (PBR) follow this type of model structure. This means that concentration gradients are assumed through the reactor. This type of model structure is also applicable for some implementations of multi-enzyme processes.

For multi-enzyme processes, the formulation of a net reaction rate for the reaction network involves extensive calculations which can be sensitive to manipulation errors. Therefore the generation term is expressed by the sum of all reaction rates coming from the individual enzymes involved in the process. In this way, the reaction rates are properly included in the conservation equations and then the full model is able to take into account the different behavior and trends in the full process.
Model Implementation in Chapter 4 suggests mathematical methods and tools to be used for the solution of the conservation equations.

3.5.4 Constitutive equations

These are algebraic equations which describe thermodynamics, kinetics, controls, etc., and they are incorporated into the conservation balances in order to include these relationships in the process. In the multi-enzyme modeling the reaction rates for the enzymes involved in the process can be considered as the most important constitutive equations. Nonetheless, there are many multi-enzymatic reactions that require molecular oxygen, specifically when oxidases are involved in the process, e.g. glucose oxidase and laccase [24]. In those cases, diffusion of the gas (air, enriched air or pure oxygen) in the liquid media must also be considered in the model, and then it is represented by the gas to liquid transfer rate equations [48]. In the scientific literature different types of reaction rate expressions are well developed to describe enzyme mechanisms (i.e. how the enzyme binds the substrate(s) and how the product(s) is released). These types generate different mathematical equations depending on the mechanism described for each enzyme in the system [18]. For reaction rate expressions, the Michaelis-Menten expression is generally used for one substrate while there are other expressions for more than one substrate such as, the compulsory-order ternary-complex mechanism, the substituted-enzyme mechanism and the random-order ternary-complex mechanism. Furthermore, substrate(s) and product(s) influence, e.g. with enzyme inhibition, is also necessary to include this phenomenon in the reaction rate expression when required. In cases where reaction rate expressions are not available in the scientific literature for a given enzyme, they can be derived from steady-state kinetics e.g. using the King-Altman or Wong-Hanes methods [18, 62]. Even though these methods are not explained in this thesis, they provide a systematic way of describing enzyme mechanisms when they are not available. Appendix A presents both the most representative single enzyme mechanisms for reactions of one and two substrates and their corresponding reaction rate expressions.

Defining a preliminary reaction rate for each enzyme, the next question is how many kinetic parameter values are assumed known i.e. from the scientific literature. In case of missing parameters in the suggested reaction rate model, they must be estimated from reliable experiments that allow the possibility to find a proper description of these enzymes. In those cases, there are different methods that guide the parameter estimation e.g. based on linear plotting and nonlinear regression. [63, 64]
3.6 Step 4. Calculations

3.5.5 Conditions

Many reactions of importance in biochemistry are characterized by being reversible, and then the thermodynamic equilibrium is an important condition to meet in the model. Commonly for enzymatic reactions, the equilibrium constant of the reaction is expressed by the relationship between the kinetic parameters. The resulting algebraic equation is well known as the Haldane relationship [18, 62]. The structure of the equilibrium constant depends on the enzyme mechanism, for the reversible Michaelis-Menten expression (see Appendix A), the Haldane relationship may be described as:

\[ K_{eq} = \frac{K_{cat}^e K_m^p}{K_{cat}^m K_m^p} \]  

(3.5)

where \( K_m \) represents the Michaelis-Mente constant and \( K_{cat} \) represents the turnover number for the enzyme.

For the numerical solution of the mathematical models, initial and boundary conditions are necessary. In this way, equations described by a set of ODE's (lumped parametric models) require as many initial conditions of the states (e.g. initial concentrations of substrate(s) and product(s)) as the number of ODE equations. On the other hand, equations described by PDE's (distributed parameter models) not only require initial conditions but also boundary conditions of the system which say something about the states at certain positions in the control volume [48].

3.6 Step 4. Calculations

In the calculations, there are six subsequent and iterative activities (see Figure 3.1). They are divided into model implementation, identifiability analysis, model calibration, uncertainty analysis, global sensitivity analysis and optimal experimental design. In this section, the main purpose is to explain these substeps mathematically with focus on multi-enzyme processes, and to suggest suitable mathematical techniques for the model analysis. Furthermore, procedures for the mathematical implementation of the suggested techniques are covered in Chapter 4 where concepts and applicability are explained in more detail. However, for the application of the suggested techniques, preliminary experimental data is required. Basically, this information is used as the basis for the generation of more relevant and improved results according to the formulated model objective.
3.6.1 Model implementation

Computational implementation of the model is required to solve the set of equations that describe the system, specifically formulated in Step 3 of the methodology (Section 3.5). Process simulation gives the possibility of studying and understanding the dynamic behavior of the system when different changes are performed in the input variables or initial conditions of the process. Depending on the model structure and software tools, the implementation can be directly applied in the computational program (e.g. MATLAB) using available functions; otherwise further algorithms must be developed and implemented in order to give the required instructions to solve the specific problem.

3.6.2 Identifiability of the parameters based on experimental data

Enzymatic reactions are well known for involving non-linearities in behavior [18, 62]. Furthermore, they are characterized by a high degree of complexity which is directly translated to a large number of parameters in the model structure.

Identifiability problems can be directly related to the model structure (structural identifiability) or to the quality/quantity of experimental data (practical identifiability) [55]. However, the second reason seems to be more relevant for multi-enzyme process modeling since practical identifiability not only considers the structural identification of the parameters in the model, but also takes into account the impact of the available experimental data in the procedure. Usually two conditions must be fulfilled to obtain a subset of identifiable parameters. The model output must be sensitive to individual changes in each parameter and the parameters must be uncorrelated.

The main purpose of this analysis is to indicate whether it is possible to obtain unique values of the model parameters given certain experimental data. However, this analysis is a trade-off between the experimental data and the number of parameters in the mathematical model. An identifiability problem can be noticed when the number of identifiable parameters are less than the total number of parameters in the model. This can occur for two reasons: the first one is when the model structure is assumed correct but the quality of experimental data is poor, and the second one is when the quality of experimental data is assumed to be good but the model structure is wrong.

For improving parameter identifiability when experimental data is poor, optimal experimental design can be formulated based on the model. In this manner, the
3.6 Step 4. Calculations

The most suitable experimental procedure is obtained in order to get reliable information for the model calibration. If more experiments cannot be carried out and all parameters in the model are sensitive, only the number of parameters indicated by the test can be identified and the others can be assumed as correct and then fixed at the original values. This means that the most certain parameters can be assumed correct based on *a priori* knowledge and the most uncertain parameters can be estimated with the available experimental data.

The identifiability problem, when good experimental data and wrong model structure is assumed, can be an indication of model over-parametrization. Hence model simplification can be considered to improve the model calibration. However, an inadequate model structure simplification can loose the insight of the mechanism. This means that the model can exhibit inaccuracies when performing extrapolation or predictions of the process (limited predictive power) in the neighborhood of the conditions where the experiments were performed. Consequently, sensitivity analysis is required in order to reveal which parameters can be best removed from the model without affecting the overall process behavior. In principle, insignificant parameters can be removed.

3.6.3 Parameter estimation, confidence intervals and correlation matrix

This step deals with the calibration of the formulated mathematical model. This means the estimation of the unknown parameters of the model using some preliminary experimental data. This step is not only important in order to find physically meaningful parameter values, but also in order to assess the uncertainty of the resulting model.

For model calibration using computational tools, an estimator attempts to approximate the unknown parameters using the experimental measurements. Hence, a satisfactory model is achieved when the set of estimated parameters for a model can best explain the experimental observations within a specified uncertainty. However, good visual fitting does not always mean good performance of the parameter estimation. Hence, further techniques may be also applied in order to validate or reject the reliability of such information. Consequently, confidence intervals, the correlation matrix of parameters and output predictions must also be considered in the model building analysis (relevant mathematical expressions for the mentioned calculations are shown in Chapter 4 [65, 66]).

As mentioned, multi-enzyme models involve several parameters in the formulation therefore identifiability and sensitivity analyses must come together with the parameter estimation. In this way, identifiability analysis provides the set
of parameters that can be estimated from the available experimental data, and sensitivity analysis identifies the parameters that are most significant in the model structure, and that in fact drive the output model of the process.

Notice that during the procedure of parameter estimation, it is always required to divide the experimental data, since a set of experiments must be used for model validation. The selected set of data cannot be part of the parameter estimation and should preferably be performed under different operating conditions within the desired operating window.

### 3.6.4 Uncertainty analysis

Calibrated models do not necessarily represent reality, since reality is usually much more complex. In that manner, the prediction capability of the model is limited by the presence of an inherent error that comes with reality. Sources of uncertainty are also related to the used analytical methods which can offer more or less precision and accuracy of the measured variables. Such error cannot be ignored, but should rather accounted for by making decisions on the basis of model outcomes.

For multi-enzyme processes, this analysis is important, especially for scaling-up the process, since it can show the risks to which the process can be exposed if working outside the predictable range in a clear and simple manner, e.g. areas where concentrations are higher than required, so enzyme activity can be affected, or reaching limits where some compounds in the media are denatured.

The scaling-up procedure also involves control decisions if a profitable process is desired. Consequently, uncertainty analysis also provides information about the variables that can best be controlled in the process, since in process control reliable measured variables are required for good performance of controllers. Definitely, variables with a narrow range of uncertainty or propagation error need to be measured (see Figure 4.5 Chapter 4), because it is practically a measurement that can be trusted and thus has less bias in the control response.

A further application where uncertainty analysis plays an important role is in the implementation of on-line parameter estimation. In the same manner, trustworthy variables are required to be continuously measured in order to recalibrate the model when process operating conditions are changed. This means that the model can have adaptability properties [43].

In this thesis, the Monte Carlo technique is used in the evaluation of uncertainty since the method offers more global results due to the amount of model evalu-
3.6 Step 4. Calculations

In this thesis the Morris Screening method is used for the sensitivity analysis of parameters [68]. It measures the degree of importance of input factors in generating output variations. The advantage of using the Morris method is that the sensitivity analysis is global due to the sampling that is employed in the model evaluation [68, 69]. Chapter 4 presents the technique to implement this kind of analysis. Furthermore, sensitivity analysis is applied for different modeling objectives in the three case studies of this thesis.
3.6.6 Optimal experimental design

In practice, there are variables that are much easier or more difficult to measure in a process, so more or less experimental data points can be obtained. The main idea of making experiments is to capture the dynamics of a process under given conditions. However, many data points do not always mean more information. Therefore, it is important to understand the type of experiments and measurements that are required for model calibration, especially in multi-enzyme systems.

Optimal experimental design can be formulated for purposes such as:

1. To get a better understanding of some phenomena of the system under study
2. To estimate (some of) the parameters of a model or
3. To discriminate between several possible model structures

Even though all purposes are interesting to investigate, to date multi-enzyme models are lacking information about the type of experimental data that is required for the estimation of the kinetic parameters. Therefore, more effort must be put in the optimal experimental design for parameter estimation. Consequently, it is desired to design experiments in such a way that the data which will be collected from these experiments are as information-rich as possible. The main questions that the design can answer are related to whether, where and how the system under study can be manipulated and where, how and when measurements can be performed on this system. Furthermore, there are methodologies that allow design of experiments that can combine and reach several of these goals [55, 70].

The methods and procedure for optimal experimental design are explained Chapter 4. However, no application of this technique was implemented in the case studies described in this thesis, as shown in Figure 6.1.

3.7 Step 5. Experimentation

This step covers the information that can be obtained from experimentation of the process. The type of experiments and measurements of the process are key information to interpret what happens in reality. As mentioned in previous
steps, experimental data are necessary to formulate reliable models which involve complex objectives.

High quality experimental data are always desired for modeling. However, experimentation always carries inherent errors such as human, systematic and random errors which are responsible for deviations from true values. After detecting the source of the error in the first two cases, these can be reduced by taking more care in carrying out experiments or refining the measurement methods or techniques. However, random error occurs because no measurements can be made with infinite precision and then it may only be reduced by averaging several measurements. In this case, the measurement error is quantified by the standard deviation which can be included in the modeling procedure. In this way, measurements with small standard deviation have more relevance in the calculations.

Experimental data can be obtained from real process implementation, the scientific literature or databases. However, when experimental data are collected from the scientific literature or databases, a clear idea of how the experimental set up was made is required. In this case, the operating conditions of the experiments and process dynamics must match the modeling objectives in order to consider this information as relevant for the methodology. For each set of experiments, issues such as operating conditions, enzyme concentrations, initial component concentrations, time-dependent variables or in steady state measurements and sampling time must be clearly specified in order to make this information useful for further application. However, often experimental data in the scientific literature is presented in a sub-optimal way such as incomplete reporting of process conditions or even inconsistent units for variables and parameters. In this case, a quantitative database ensures a more standardized experimental information since there is certain requirements which are necessary to report the information. However, databases for enzymes have been focused on nomenclature and general properties rather than on reaction mechanisms, kinetic parameters and process conditions which may be more useful for mathematical model developments.

In Figure 3.1 at Step 5, experimental information such as initial parameter values, initial conditions and boundary conditions (when required), is necessary for the computational implementation of the model. Furthermore for parameter identification, calibration of the dynamic models and optimal experimental design, progress curves are necessary since most of the process behavior is captured by measuring variables as function of time, as illustrated in Figure 3.5. In multi-enzyme processes, progress curves are often taken in batch operation mode (see Figure 3.5a). However fed-batch and continuous operation can show much more interesting dynamics and useful information for the model since a controlled perturbation (e.g. change in the input flow rate or input concentra-
48 Methodology for the model-building of multi-enzyme ex-vivo processes

tion) can be performed in the process in order to analyze the different aspects of the effects on the outputs than in batch (see Figures 3.5b and 3.5c).

Figure 3.5: Expected behavior from experimental progress curves. (a) Batch operation mode, (b) fed-batch operation mode and (c) continuous operation mode

Depending on the variable and the applied measurement technique, a number of data points can be collected in a given period of time. There are variables that are easy to measure such as pH, temperature or dissolved oxygen, and others (such as component concentrations) that require more sophisticated analytical techniques to determine their values (e.g. spectroscopy and chromatography). However, not all measurements supply the same type of information about the process. That is why reliability of experimental data increases when it is based on optimal experimental design since precise effort is put in collecting the measurements that can provide most significant information about the process dynamics for the model development.

An important issue to include when working with experimental data is data analysis. This primarily includes a group of techniques that quantify the reliability of the experiments and measurements e.g. to analyze the influence on the data if there are data filtering, missing observations or extreme observations (outlying observations). This analysis is performed to identify and treat the information which perturbed the measurements from the true value or to avoid redundancy in experimental data [71].

Experiments, such as enzyme stability, degradation of components in the media, influence of substrate concentration(s), temperature and pH, can also be carried out in order to analyze different effects related to the process behavior. This information gives more understanding of the process such that a greater operability range can be achieved by the model when these effects are included.
3.8 List of symbols and nomenclature

Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK</td>
<td>Transketolase</td>
</tr>
<tr>
<td>TAM</td>
<td>Transaminase bulk</td>
</tr>
<tr>
<td>OED</td>
<td>Optimal experimental design</td>
</tr>
<tr>
<td>LQR</td>
<td>Linear quadratic regulator</td>
</tr>
<tr>
<td>MPC</td>
<td>Model predictive control</td>
</tr>
<tr>
<td>CSTR</td>
<td>Continuous stirred tank reactor</td>
</tr>
<tr>
<td>BSTR</td>
<td>Batch stirred tank reactor</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>CPBR</td>
<td>Continuous packed bed reactor</td>
</tr>
<tr>
<td>AE</td>
<td>Algebraic equation</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary differential equation</td>
</tr>
<tr>
<td>PDE</td>
<td>Partial differential equation</td>
</tr>
<tr>
<td>PFR</td>
<td>Plug flow reactor</td>
</tr>
<tr>
<td>PBR</td>
<td>Packed bed reactors</td>
</tr>
</tbody>
</table>

Symbol

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Substrate</td>
</tr>
<tr>
<td>P</td>
<td>Product</td>
</tr>
<tr>
<td>I</td>
<td>Inhibitor</td>
</tr>
<tr>
<td>X</td>
<td>No interaction between components</td>
</tr>
<tr>
<td>p</td>
<td>Number of inputs to the mass balance</td>
</tr>
<tr>
<td>q</td>
<td>Number of outputs from the mass balance</td>
</tr>
<tr>
<td>n</td>
<td>Number of components in the system</td>
</tr>
<tr>
<td>F_{obj}</td>
<td>Objective function for optimization</td>
</tr>
</tbody>
</table>

Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Concentration of the component A (M)</td>
</tr>
<tr>
<td>B</td>
<td>Concentration of the component B (M)</td>
</tr>
<tr>
<td>C</td>
<td>Concentration of the component C (M)</td>
</tr>
<tr>
<td>E_i</td>
<td>Concentration of the enzyme i (M)</td>
</tr>
<tr>
<td>M</td>
<td>Total mass (mol)</td>
</tr>
<tr>
<td>t</td>
<td>Time (s)</td>
</tr>
<tr>
<td>F_j</td>
<td>Input flow rate (mol s(^{-1}))</td>
</tr>
<tr>
<td>F_k</td>
<td>Output flow rate (mol s(^{-1}))</td>
</tr>
</tbody>
</table>
Methodology for the model-building of multi-enzyme ex-vivo processes

$m_i$  Mass hold-up of component $i$ (mol)
$f_{i,j}$  Input flow rate of the component $i$ (mol s$^{-1}$)
$f_{i,k}$  Output flow rate of the component $i$ (mol s$^{-1}$)
$g_i$  Rate of generation or consumption of species $i$ (mol s$^{-1}$)
$r_i$  Net reaction rate of the compound $i$ (mol L$^{-1}$s$^{-1}$)
$V$  Reactor volume (L)
$K_m$  Michaelis-Mente constant (M)
$K_{cat}$  Turnover number of the enzyme (s$^{-1}$)
$x$  Spatial variation in $x$ direction (m)
$y$  Spatial variation in $y$ direction (m)
$z$  Spatial variation in $z$ direction (m)
$g$  Rate of generation (mol s$^{-1}$)

Greek letters

$\varphi$  Intensive quantity (M)
$\phi$  Extensive quantity (mol)

Subscripts

$i$  Index for the number of components
$j$  Index for the number of inputs
$k$  Index for the number of outputs
Chapter 4

Mathematical methods for model analysis

Today, computational tools provide a number of opportunities in order to solve complex and simple mathematical problems in an effective manner. Computer aided software enables the possibility to perform multiple simulations of a process, giving the opportunity to explore, evaluate and analyze the process under different conditions and scenarios.

In this thesis, programming codes are developed using MATLAB (The Mathworks, Natick, MA). Most of the used functions are mainly taken from those that are already included in the program, and some others are created in order to solve specific mathematical problems in the analysis of each case study. Furthermore the codes are developed in such a way that they are flexible and it is possible to modify them without too much effort. In this way, other multi-enzyme process models can easily be implemented for their analysis. Furthermore, Appendix B shows, in flow sheets, the sequence of the different calculation tasks that were required to develop the MATLAB codes. The flow sheets also specify the required input information and its corresponding output for the main MATLAB functions and the independent calculations that were required.
4.1 Model implementation and dynamic simulation

From Step 3 (Section 3.5), the conservation balances are described by ordinary differential equations (ODEs) as a function of time (dynamic models - lumped parameter models), and partial differential equations (PDEs) as a function of time and space (dynamic models - distributed parameter models). Furthermore, algebraic equations (AEs) are obtained for steady state models [43, 48]. As mentioned, the software MATLAB is used to facilitate and optimize the implementation and solution of the models. In order to solve the set of equations (i.e. the mathematical model of the process), solvers which represent different numerical methods can be applied. For instance, the solver 'ode45', in MATLAB, is used to solve models described by a set of nonstiff ODEs. This solver is based on the Runge-Kutta method which is the most widely used for solving ordinary differential equations [48]. Eventually, the solver 'ode15s' is used when the numerical solution with 'ode45' is too slow because the model that is to be solved forms a stiff system of ODEs. The implementation mainly requires a function with the set of differential equations (i.e. the mathematical model), time span (i.e. initial and final integration time), the initial conditions of the states for each differential equation (e.g. initial concentration estimates of the variables in the process), and a set of model parameter values (see Figure B.2). Furthermore, options related to the numerical method used by the solver can be customized according to the needs of the problem and the required precision of the solution.

In this thesis, case studies with spatial dependence (i.e. distributed parameter models) were not included. However for those cases, discretization of the partial differential equation is required to solve the set of equations for the model. If the model includes the time dependence (see Equation 3.4), after the discretization the set of PDEs is reduced to a set of ODEs with one for each state in the model. If the accumulation term is neglected, the set of PDEs is reduced to a set of AEs for each state in the model. In any case, the final number of equations to be solved depends on the discretization number of nodes and the number of model states (i.e. one ODE or AE per model state variable in each node). In order to solve this type of equations, the boundary conditions are also required to be specified. After that the mathematical problem can be solved using existing MATLAB functions, as mentioned above.
4.2 Identifiability of parameters

There are many techniques to evaluate identifiability of a model and most of them are based on the local sensitivity analysis of the parameters which is the derivative of the model output with respect to each parameter in the model, \( \partial y / \partial \theta \). Some techniques for analyzing sensitivity are: model reduction, regression methods, visual interpretation of local sensitivity functions and parameter selection based on collinearity index, among others [55].

Since the local sensitivity analysis of parameters is fundamental for further calculations, the previous definition is mathematically expressed in Equation 4.1. For simple models, the analytical derivation is the most precise. However, for complex models, the calculation of this function can be approximated by considering a forward difference method, thus [43, 55]

\[
S_{i,j} = \frac{\partial y_i}{\partial \theta_j} = \frac{y_i(\theta_j + \Delta \theta_j) - y_i(\theta_j)}{\Delta \theta_j}
\]  

(4.1)

where \( \Delta \theta_j \) is the change in the parameter, \( i \), and \( j \) are the number of model outputs and parameters in the model, respectively. In practice, \( \Delta \theta_j \) is implemented as the nominal value of the parameter \( \theta_j \) multiplied by a perturbation factor \( \xi \) which should not be too big because the non-linearities of the model can start affecting the calculation, and should not be too small either because it can produce numerical inaccuracies [55].

The identifiability technique used in this thesis is based on the collinearity index which also takes into account, in the procedure, previous experimental data of the analyzed process (see the sequence of steps in Figure B.3). An interesting characteristic of the collinearity index is that it is independent on \( \Delta \theta_j \) because the sensitivity function is the normalized (non-dimensional) \( \tilde{S} \), thus not only can the result interpretation be more general, but also a parameter importance ranking can be obtained [55, 72, 73]. The normalized sensitivity function is calculated as:

\[
S_{i,j} = \frac{\Delta \theta_j}{s_{c_i}} \times \frac{\partial y_i}{\partial \theta_j} \quad \tilde{S}_{i,j} = \frac{S_{i,j}}{\sqrt{\sum_{k=1}^{N} S_{k,j}^2}}
\]  

(4.2)

where \( s_{c_i} \) represents the measurement accuracy of \( y_i \), \( \tilde{S} \) is the normalized sensitivity matrix, and \( N \) is the number of experimental data points.
The importance of a single parameter can be evaluated by the sensitivity measure which is defined as:

\[ \delta_j^{msqr} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} S_{i,j}^2} \]  

(4.3)

A high value of the sensitivity measure means that the parameter has an important influence on the simulation result. The significance of individual parameters is important to investigate regarding the structure of the model.

Finally the collinearity index can be calculated, thus

\[ \gamma(\theta) = \frac{1}{\sqrt{\lambda_{\min} [\tilde{S}^T \tilde{S}]} } \]  

(4.4)

Where \( \lambda_{\min} \) is the smallest eigenvalue of the matrix of squared normalized sensitivity functions \( \tilde{S} \). The collinearity index represents the linear dependences of the sensitivity function for different parameters in a subset. Hence an index equal to one means that the sensitivity functions are independent, and the index value is increased when the linear dependency is increased. It has been observed that for collinearity indexes higher than 15, the subset is poorly identifiable [55, 72, 73]. The collinearity index can be easily analyzed by comparing it to the threshold and each subset of parameters that can be identified. Figure 4.1 shows an example of how results can be analyzed.

In Figure 4.1, each step change in the red line separates the parameter subset by number of parameters within the subset i.e. 2, 3, \( \cdots \), \( p \) where \( p \) is the total number of parameters in the model. The red line represents the mean value of the collinearity index in the subset with the same number of parameters. The blue line represents the collinearity index calculated for all subsets of parameters. In the previous example, all subsets of parameters can be identifiable since the collinearity index is no greater than 15. However, different parameter combinations within a subset can generate lower values of the index which are desired.
4.3 Parameter estimation, confidence intervals and correlation matrix

For parameter estimation, an objective function must be formulated in such a way that it can quantify the deviation of the model from the experimental data. Commonly, the non-linear least squares method is utilized \cite{74}. Furthermore, experimental data contains a degree of uncertainty due to the noise or measurement error that can be incorporated during the experimental work. In this case, the objective function is formulated as the weighted sum, over all data points \( N \), of the squared errors between measurements \( y_{\text{meas}} \) and the corresponding model predictions \( y(\theta) \), as shown in Equation 4.5. The objective function value decreases as soon as the parameter values are close to the true model parameters assuming that the model structure is correct. Hence the goal is to obtain a set of parameter values that corresponds to the global minimum value of the

\[ \sum_{i=1}^{N} w_i (y_{\text{meas}}(i) - y(\theta(i)))^2 \]
objective function [75].

\[
J(\theta) = \sum_{k=1}^{N} \left( \frac{y_{k,\text{meas}}(t) - y_k(t, \theta)}{\sigma_k} \right)^2
\]  

(4.5)

In Equation 4.5, the weight on the deviations is represented by the standard deviation (\(\sigma\)) of each measured variable which is obtained from experimental data, and it is directly related to the precision of the analytical methods used in process monitoring. The existence of measurement error means that the model never fits the experimental data perfectly even if the model structure is correct. However, this information is also used to give more importance to those measurements that are more reliable (i.e. less noisy). This means that measurements with high error will contribute less in the overall objective function. On the other hand, there are some cases where the weight can be removed from the objective function, specifically if this information is not available or if the measurement error is equal for all measured variables. In this case, Equation 4.5 is converted to the standard nonlinear least square method.

For reversible enzymatic reactions, a modification in the objective function can be incorporated in order to better drive the parameter estimation since more information about the process is taken into account. In that case, the Haldane relationship to calculate the equilibrium constant can be added as a condition to be met for the parameter estimation. If the relationship between parameters does not meet the required value or range for the equilibrium constant, the objective function is penalized with a high value. Consequently, the optimizer is forced to change the set parameter values to another set which can meet all conditions in the model.

The ideal result for parameter estimation is to obtain a global minimum in the optimization. However, enzymatic mechanistic models have many parameters that rarely can be estimated with a unique solution. Actually it is an important issue to consider in this type of system since the correlation that can exist between parameters is expected to be high and then identifiability problems for the model will result.

As mentioned, an idea of the quality of the estimation can be obtained by calculating the confidence interval and correlation matrix of the parameters involved in the estimation.

In order to estimate the confidence interval of the parameters, a linear approximation of the covariance matrix must be evaluated. It can be calculated using
4.3 Parameter estimation, confidence intervals and correlation matrix

Equation 4.6 [65]

\[
\text{Cov}(\theta) = \frac{\min J(\theta)}{N - p} \left[ \left( \frac{\partial y}{\partial \theta} \right)^T Q^{-1} \left( \frac{\partial y}{\partial \theta} \right) \right]^{-1} \tag{4.6}
\]

where \( N - p \) represents the degrees of freedom, \( N \) is the number of measurements and \( p \) represents the number of parameters, \( \partial y/\partial \theta \) represents the Jacobian matrix which corresponds to the local sensitivity of model variables \( y \) to parameters \( \theta \), and \( Q \) is the covariance matrix of measurement errors.

The confidence interval of parameters is determined with a confidence level of \((1 - \alpha)\) corresponding to the 95th percentile of the t-distribution value calculated at the \( \alpha/2 \) percentile with \( N - p \) degrees of freedom, as described in Equation 4.7 [65].

\[
\theta_{1-\alpha} = \theta \pm \sqrt{\text{diag}(\text{Cov}(\theta))} \cdot t \left( N - p, \frac{\alpha}{2} \right) \tag{4.7}
\]

where the \( \text{diag}(\text{Cov}(\theta)) \) takes into account only the diagonal values of the covariance matrix of parameters (see Equation 4.6). As a result, the narrower the confidence interval is, the higher the quality of the parameter estimation.

Notice that the confidence interval tends to be small when more experimental data points are included in the parameter estimation due to the degrees of freedom. However, it is important to check redundancy of the data before using it, because too many data points do not always contribute with more information about the process, and then no better estimation can be expected.

Finally, the strength and direction of the linear relationship between parameters is evaluated with the correlation matrix based on the covariance matrix \( \text{Cov}(\theta) \), as formulated in Equation 4.8 [65].

\[
\text{Cor} (\theta_i, \theta_j) = \frac{\text{Cov}(\theta_i, \theta_j)}{\sqrt{\text{Cov}(\theta_i, \theta_i) \cdot \text{Cov}(\theta_j, \theta_j)}} \tag{4.8}
\]

Correlation coefficients between parameters around unity indicate that any change in one parameter could be compensated by a change in another, meaning that it is too difficult to find a unique solution for those parameters. Consequently, identifiability analysis can be performed in order to find the subset of
parameters that can be estimated with a unique solution from the information contained in the experimental data.

4.4 Uncertainty analysis: Monte Carlo method

The Monte Carlo technique is used as a method for analyzing uncertainty propagation in models and the input variables that give rise to this uncertainty (see also Section 3.6.4). The main purpose of this method is to determine how the process is affected when random variations and a lack of knowledge are considered as an inherent part of the reality. These conditions directly affect the sensitivity, performance, or reliability of the model. The advantage of the Monte Carlo technique is that it can be used for complex and non-linear models, and when the model involves more than just a couple of uncertain parameters [76]. The method consists in iterative model evaluations using sets of random numbers as inputs. Therefore it is categorized as a sampling method in which probabilistic distributions are evaluated and analyzed for generating applicable results. In reality, a state measurement must be understood not as a point in phase space but rather as a small region whose size reflects the finite precision of the measuring apparatus [76]. Hence, the results from the simulations can be represented as histograms, or converted to error bars, reliability predictions, tolerance zones, or confidence intervals [77, 78].

This method has had a higher impact during recent years since the computational cost is no longer a major limitation, and then a large number of virtual experimental trials can be simulated. The Monte Carlo analysis involves four steps, as illustrated in Figure 4.2[23]

4.4.1 Specifying input uncertainty

This covers the range and distribution of each parameter $\theta$ that needs to be studied. In case parameter estimation is performed, the range for each parameter value can be calculated by Equation 4.7 which corresponds to the confidence interval calculation (i.e. upper and lower bound of the kinetic parameters). Otherwise, ranges can be found consulting the relevant scientific literature sources and/or asking for the opinion of process experts [66]. The distribution of the parameters needs to be included for the calculations e.g. uniform, log uniform or triangular distributions. Again if parameter estimation is performed, the correlation matrix can be calculated by Equation 4.8 and then it corresponds to the real distribution of parameters. In this case, the correlation matrix is included
4.4 Uncertainty analysis: Monte Carlo method

in the method in order to limit the random sampling space of the evaluated parameters and then more precise uncertainty results can be obtained [77].

4.4.2 Sampling input uncertainty

For applying the method a random sampling must be generated within the range and distribution specified for each parameter in the previous step. The result is a sequence of parameter values expressed as:

\[ \boldsymbol{\theta}_i = [\theta_{i1}, \theta_{i2}, \ldots, \theta_{im}] \quad i = 1, 2, \ldots, m \quad (4.9) \]

where \( n \) is the number of inputs (i.e. parameters) and \( m \) is the sample size. Different methods can be used for the generation of the distribution of parameter values (e.g. random sampling, latin hypercube sampling and orthogonal sampling). Nonetheless, the latin hypercube sampling is the most widely and often applied technique in uncertainty analysis. Some of the properties of this sampling technique are full stratification across the range of each input parameter, relatively small sample size, and direct estimation of means, variances and distribution functions [78].
The generated matrix from the sampling can be represented in a scatter plot matrix as shown in Figure 4.3.

![Scatter plot matrix](image)

**Figure 4.3:** Matrix plot of scatter plot for the random sampling and correlation between the analyzed model parameters ($\theta_1, \theta_2, \theta_3, \cdots$)

Figure 4.3 shows the relation that exists between parameters and how the random sampling is distributed in the parameter space (defined by the confidence intervals). The sampling follows a specific trend when the correlation matrix of the parameters is included in the sampling method.

### 4.4.3 Monte Carlo simulation

In order to analyze the uncertainty of the model and thus the reliability of the prediction of the process, the model must be solved for each set of parameters obtained in the sampling. This generates a sequence of results in the following form:

$$y_i = f(\theta_i) \quad i = 1, 2, \ldots, m \quad (4.10)$$

where $y$ represents the output model which is function of the sampled parameters as expressed in Equation 4.9. In essence, this creates a mapping from the analyzed inputs to the analysis of the results.
4.4 Uncertainty analysis: Monte Carlo method

Each process simulation can be compiled in the same figure which then provides an overview of how much the model output is affected by a change in a parameter value. Figure 4.4 shows an example of the type of plot that can be obtained in this step.

![Figure 4.4: Distribution of the model simulation computed at each parameter sampling space](image)

4.4.4 Representation and interpretation of uncertainty

For a better interpretation of the model simulations, the result can be represented by the calculation of the mean value, thus:

$$ E(y) = \frac{\sum_{i=1}^{m} y_i}{m} $$ (4.11)

and the variance, thus:

$$ V(y) = \frac{\sum_{i=1}^{m} [y_i - E(y)]^2}{m} $$ (4.12)

Another way to represent the uncertainty is calculating the 10th and the 90th percentiles of the Monte Carlo output. These percentiles are ideal for graphical representation of uncertainty results. Figure 4.5 shows the propagation of
the uncertainty using the mean value and the corresponding percentiles of the distribution obtained from the model evaluations.

Figure 4.5: Uncertainty propagation represented by the mean value and the 10th and 90th percentile of the model simulations based on Monte Carlo method

4.5 Sensitivity analysis: Morris Screening method

This global method effectively identifies and measures the degree of importance of input factors which can generate variations in the model output [79]. For the Morris method, an input factor \( n \) is defined as a parameter or variable that is perturbed from its nominal value to see the effect in the model output. In fact, effects can be considered as: negligible, linear and additive, and nonlinear or involved in interactions with other inputs [69]. As briefly mentioned, the selected input factors are perturbed by a given percentage away from the base values (in turn) while the other input factors are kept constant. This parameter perturbation is called elementary effects \( EE_i \) (see Equation 4.15). However, this method is improved by Morris computing \( r \) number of those local measures following an efficient randomized sampling scheme extracted from the input space [68]. In this way, the advantages of a global method are also obtained. Furthermore, this method provides an unbiased estimate of \( F_i \) which is the distribution function of \( EE \) for the \( i^{th} \) factor. Consequently reliable inferences can be made by using the estimated mean and standard deviation of the distribution function \( F_i \). The Morris screening method involves the following steps (see Figure 4.6):
4.5 Sensitivity analysis: Morris Screening method

4.5.1 Input factor

In a similar manner as done for the application of the Monte Carlo method, an uncertainty for each factor value needs to be specified (i.e. maximum and minimum value for the analyzed parameters). If parameter estimation is performed, the confidence interval of parameters can then be calculated (see Equation 4.7). Otherwise, the opinion of process experts or relevant scientific literature must be sought for delimiting the ranges of parameters [66].

4.5.2 Morris sampling

In order to obtain a matrix with random sampling, the optimal number of samples needs to be determined in order to apply the method effectively. First, the number of levels $p$ and the number of samples $r$ need to be defined. Subsequently, the optimal perturbation or incremental factor $\Delta$ can be calculated.
as:

\[ \Delta = \frac{p}{2(p-1)} \]  

(4.13)

The sampling space can be visualized in a scatter matrix plot which is interpreted in the same way as done in Figure 4.3.

### 4.5.3 Model evaluation

Defining the input factors \((n)\) and the number of desired samples \((r)\), the computational cost of this method can be calculated as:

\[ me = r(n + 1) \]  

(4.14)

Thus the number of model evaluations corresponds to the same number of sampling points computed in the previous step. In this manner, the influence of each parameter variation can be seen and compared by overlapping the model outputs of all simulations, as similarly shown in Figure 4.4.

### 4.5.4 Interpretation of the results

The results of the global sensitivity analysis method are based on the calculation and interpretation of the elementary effects. The calculation of the elementary effects may be described by Equation 4.15.

\[ EE_i^r = \frac{y(\theta_1^r, \theta_2^r, \theta_3^r + \Delta, \ldots, \theta_m^r) - y(\theta^r)}{\Delta} \]  

(4.15)

where \(y(\theta_1^r)\) are the model outputs at each random sample, \(y(\theta^r)\) are the model outputs computed at the nominal parameter values, and \(\Delta\) is the optimal perturbation factor, as calculated in connection with Equation 4.13. The interpretation of the elementary effect is done by calculating the mean and standard deviation of the distribution function \(F_i\) for each model output obtained from the elementary effects.
4.5 Sensitivity analysis: Morris Screening method

4.5.5 Critical factors affecting the model output

The mean $\mu_i$ and the standard deviation $\sigma_i$ of the $r^{th}$ elementary effects are used as influence measurements. $\sigma_i$ is used to detect parameters involved in interaction with other parameter or whose effects are non-linear. The absolute mean value of the distribution function $F_i$, $||\mu_i||$, is used to rank the most significant factors in the model. The higher the absolute mean value, the higher the influence in the model.

The sensitivity analysis results can also be represented graphically comparing the standard deviation against the mean values of the distribution function, as shown in Figure 4.7.

![Figure 4.7: Graphical visualization of the estimated mean and standard deviation of the distribution of elementary effects for identification of the most significant parameters in the model](image)

Each point in the plot corresponds to one input factor ($\theta_i$). Furthermore, the plot shows two solid lines which help to interpret the degree of influence of each factor in the model. The two lines are obtained by calculating the $\pm 2$ times the standard error of means (SEM), as shown in Equation 4.16.

$$SEM_i = \frac{\sigma_i}{\sqrt{r}}$$

(4.16)

The plot can be interpreted as follows: any parameter that lies between the lines is considered as insignificant. In contrast, if the parameter lies outside of the delimited area, the parameter is considered significant.
4.6 Optimal experimental design for parameter estimation (OED-PE)

In order to formulate an optimal experimental design for parameter estimation, a modification of the objective function used for the parameter estimation is then required. In this case, the formulated objective function is the minimum value of the sum of the square errors, Equation 4.5 [43, 55, 70]. For optimal experimental design, the expected value of the objective function is written as

\[
E[J(\theta + \delta \theta)] = \sum_{i=1}^{N} (y_i - y_i(\theta + \delta \theta))^T \times Q_i^{-1} \times (y_i - y_i(\theta + \delta \theta)) \tag{4.17}
\]

Hence a linearization of the model with respect to the parameters can be introduced

\[
(y_i - y_i(\theta + \delta \theta)) \approx y_i(\theta) + \left[ \frac{\partial y_i}{\partial \theta}(\theta) \right]_\theta \times \delta \theta \tag{4.18}
\]

Replacing Equation 4.18 in Equation 4.17, the expected value is described by

\[
E[J(\theta + \delta \theta)] = E[J(\theta)] + \delta \theta^T \times \left[ \sum_{i=1}^{N} \left( \frac{\partial y_i}{\partial \theta}(\theta) \right)^T \times Q_i^{-1} \times \left( \frac{\partial y_i}{\partial \theta}(\theta) \right) \right] \theta^T \tag{4.19}
\]

From Equation 4.19, the definition of the Fisher Information Matrix (FIM) can be obtained, thus

\[
FIM = \sum_{i=1}^{N} \left( \frac{\partial y_i}{\partial \theta}(\theta) \right)^T \times Q_i^{-1} \times \left( \frac{\partial y_i}{\partial \theta}(\theta) \right) \tag{4.20}
\]

The importance of the FIM is that it quantifies the information content of an experiment related to the model parameters.

In order to get a global minimum for the \(J(\theta)\), the difference between \(J(\theta)\) and \(J(\theta + \Delta \theta)\) must be maximized. The global minimum means that the function
$J(\theta)$ is not stopped in a small depression on the flat area of the plane. Instead of that the function goes to the deepest "pit" possible in the plane (see Figure 4.8) [43]

![Figure 4.8: Type of minimum values that can be obtained for the objective function $J(\theta)$](image)

In this case, the maximization of the difference between $J(\theta)$ and $J(\theta + \Delta \theta)$ can be achieved through maximizing the FIM. There are actually different optimal experimental design criteria based on FIM properties, e.g. A, D, and E optimal designs. Furthermore, the criterion is chosen according to the optimization problem. [55].

4.7 List of symbols

**Abbreviations**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODE</td>
<td>Ordinary differential equation</td>
</tr>
<tr>
<td>PDE</td>
<td>Partial differential equation</td>
</tr>
<tr>
<td>AE</td>
<td>Algebraic equation</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of means</td>
</tr>
<tr>
<td>FIM</td>
<td>Fisher information matrix</td>
</tr>
</tbody>
</table>

**Symbol**
### Mathematical methods for model analysis

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i$</td>
<td>$i^{th}$ output model</td>
</tr>
<tr>
<td>$S_{i,j}$</td>
<td>Sensitivity function related to an output $i$ and a parameter $j$</td>
</tr>
<tr>
<td>$\tilde{S}$</td>
<td>Normalized sensitivity function</td>
</tr>
<tr>
<td>$sc_i$</td>
<td>Measurement accuracy of $y_i$</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of experimental data points</td>
</tr>
<tr>
<td>$p$</td>
<td>Number of parameters in the model</td>
</tr>
<tr>
<td>$J$</td>
<td>Objective function - cost function</td>
</tr>
<tr>
<td>$y_{meas}$</td>
<td>Output model measured</td>
</tr>
<tr>
<td>$Cov$</td>
<td>Covariance matrix</td>
</tr>
<tr>
<td>$Q$</td>
<td>Covariance matrix of measurement errors</td>
</tr>
<tr>
<td>$N-p$</td>
<td>The degrees of freedom</td>
</tr>
<tr>
<td>$\frac{\partial y}{\partial \theta}$</td>
<td>Local sensitivity - Jacobian matrix</td>
</tr>
<tr>
<td>$(1-\alpha)$</td>
<td>Confidence level</td>
</tr>
<tr>
<td>diag</td>
<td>The diagonal values of a matrix</td>
</tr>
<tr>
<td>$Cor$</td>
<td>Correlation matrix</td>
</tr>
<tr>
<td>$m$</td>
<td>The sample size</td>
</tr>
<tr>
<td>$E$</td>
<td>Expected value - mean value</td>
</tr>
<tr>
<td>$V$</td>
<td>Variance</td>
</tr>
<tr>
<td>$EE$</td>
<td>Elementary effects</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of input factors</td>
</tr>
<tr>
<td>$r$</td>
<td>Number of samples</td>
</tr>
<tr>
<td>$F_i$</td>
<td>Distribution of $EE$ for the $i^{th}$ factor</td>
</tr>
<tr>
<td>$me$</td>
<td>Model evaluation cost</td>
</tr>
</tbody>
</table>

**Greek letters**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_i$</td>
<td>$i^{th}$ model parameter</td>
</tr>
<tr>
<td>$\xi$</td>
<td>Perturbation factor</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Sensitivity measure</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Collinearity index</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Eigenvalue of the matrix $\tilde{S}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>$\theta_{1-\alpha}$</td>
<td>Confidence interval of parameters</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Significance level</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>Optimal perturbation factor</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean value</td>
</tr>
</tbody>
</table>

**Subscripts**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>Index for the number of model outputs</td>
</tr>
</tbody>
</table>
4.7 List of symbols

\(j\) \hspace{1em} \text{Index for the number of model parameters}
\(k\) \hspace{1em} \text{Index for the number of experimental data points}
\(\text{min}\) \hspace{1em} \text{Minimum value}
\(\text{meas}\) \hspace{1em} \text{Measurement}

\textit{Superscripts}

\(\text{msqr}\) \hspace{1em} \text{Mean square root}
\(T\) \hspace{1em} \text{Transpose}
Mathematical methods for model analysis
Since the target of multi-enzyme in-pot processes is the combination of different enzymes in one reactor, it is then also necessary to have them work in an efficient and stable environment, preferably under the same pH, temperature and media conditions [8, 18, 80]. In reality, not all enzymes work exactly over the same operating range i.e. each enzyme acts best around its own optimal conditions [81]. This enzyme limitation is important to analyze in multi-enzyme processes because enzyme activity and stability are highly sensitive to operating conditions. In this way, enzyme penalty due to media incompatibility can be observed as a loss of activity, inhibitory and inactivation effects, deterioration of enzyme structure and loss of stability. These issues raise the question of how the multi-enzyme process should be modified in order to improve the productivity in which the correct cost-effective process balance is achieved, even though the activity of some enzymes is compromised in the operation.

Potential improvements of multi-enzyme processes can be much more easily detected and analyzed if a reliable mathematical model is available. Consequently mathematical models implemented in a mathematical software can be used as a relatively cheap tool to identify the type of modifications that are required to both overcome the multi-enzyme process limitations and to achieve the desired productivity of the process. In this way, several scenarios with virtual changes
Multi-enzyme process improvements

can be performed by simulation in order to find the precise and ideal process operation, reactor configuration or biocatalyst performance. In this manner, the virtual results may be targets for further researches in fields such as process engineering, enzyme immobilization and protein engineering. In the following sections, the mentioned fields are addressed in order to open further and promising opportunities for multi-enzyme in-pot processes.

5.1 Process Engineering

Process engineering mainly involves decisions related to reactor design, operating mode, process control, and optimization of processes. Nowadays all these areas are mostly supported through systematic computer-based methods which facilitate and speed up the implementation of new processes, [82]. Nonetheless, relevant decisions need to be taken still in a realistic way in order to achieve reliable models and thus feasible process implementations. However computational tools have been mainly developed for chemical processes, and hence much work is required to achieve a comparable level for bioprocesses.

In the case of multi-enzyme in-pot processes, the characteristics of each enzyme and their compatibility in the whole system play an important role in the selection and design of a proper reactor, as briefly mentioned in Chapter 2 Section 2.2. For instance, standard stirred tank reactors, reactors with concentration gradients and membrane reactors are commonly used to carry out the reactions using a mixture of enzymes at laboratory scale. In the same way, the enzyme format (soluble/immobilized) and the operating mode can open a vast number of innovative process configurations. In this way different possibilities can be exploited prior to the actual process implementation.

For multi-enzyme processes it is also important to achieve stable real time operation, consistent product quality and an optimal operation at pre-defined set points. Hence, process control is necessary to be developed. In this case, mathematical models may also be used as a basis for controller design [43, 56, 83].

In this thesis, further details of reactor design, process control and process intensification are discussed in the following Sections 5.1.1, 5.1.2 and 5.1.3.
5.1 Reactor design

For multi-enzyme mixtures, different reactor options can be envisaged according to the system characteristics [24, 58]. In general, there are two major classes of reactors. The first class consists of the stirred tank reactors (STR) where a homogeneous mixture of all the compounds is assumed. The second class contains reactors that are characterized by a concentration gradient through the reactor, e.g. plug flow reactors (PFR) and packed bed reactors (PBR) [43, 46]. Furthermore, a membrane bioreactor (MBR) can have characteristics of both classes, dependent upon the specific design. The main requirement for a membrane reactor is a semi-permeable membrane which allows the free passage of the substrates or products while retaining the enzymes and/or cofactors in the system [84].

The selection of a proper reactor should in principle first consider a large list of existing reactors. However, this list soon narrows down to only a few reactor types when both physical constraints of the desired process and the precise model scope are taken into consideration [24, 48, 85]. One important criterion for multi-enzyme in-pot reactors is how the combination of enzymes will be arranged in the system, i.e. the format of the enzymes such as soluble, immobilized or a combination of both formats [86]. The final set of decisions lie on the option that best handles the characteristics of each enzyme (i.e. kinetics, media compatibility, catalytic stability, operation conditions and cost) in order to keep the process at a high productivity level for as long a time as possible.

Advantages and disadvantages of some of the likely reactors are compiled in Table 5.1. For example, a mixture of soluble enzymes can be used as catalyst in a STR or a PFR in cases where enzymes are not too expensive, neglecting the need for recycling. The situation is a little different with a PBR, where different configurations of the enzymes in the reactor can be envisaged. Figure 5.1 shows some PBR schemes for different enzyme arrangements. For example, Figure 5.1d shows a reactor with two enzymes in which the first enzyme is immobilized and the second one is continuously fed in a soluble format.

One advantage shown for the MBR is the dosing of a reactant (see Table 5.1). This case is well illustrated in the bi-enzyme production of lactobionic acid (see also Chapter 7). In this case, a membrane reactor was used for the biocatalytic reaction and an oxygen-permeable membrane was used for bubble-free oxygenation of the reaction system which improved the stability of the enzymes [27, 87].
Table 5.1: Advantages and disadvantages of different reactors for multi-enzyme in-pot process

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR</td>
<td>Soluble &amp; immobilized enzymes(^1)</td>
<td>Possible inactivation of enzymes</td>
</tr>
<tr>
<td></td>
<td>Multi-phase media</td>
<td>Difficult enzyme recycle</td>
</tr>
<tr>
<td></td>
<td>Good control possible</td>
<td>High mechanical stress on enzymes</td>
</tr>
<tr>
<td></td>
<td>No complex model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simplicity of construction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Easy to clean</td>
<td></td>
</tr>
<tr>
<td>PFR</td>
<td>Soluble enzymes</td>
<td>No immobilized enzymes</td>
</tr>
<tr>
<td></td>
<td>High conversion</td>
<td>Single phase</td>
</tr>
<tr>
<td></td>
<td>Low mechanical stress on enzymes</td>
<td>Complex model</td>
</tr>
<tr>
<td>PBR</td>
<td>Soluble &amp; immobilized enzymes(^1)</td>
<td>Thermal gradient</td>
</tr>
<tr>
<td></td>
<td>Low mechanical stress on enzymes</td>
<td>Poor control</td>
</tr>
<tr>
<td></td>
<td>High conversion</td>
<td>Complex model</td>
</tr>
<tr>
<td></td>
<td>Enzyme recycle/separation/exchange</td>
<td>Mass transfer limitations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pressure drop possible</td>
</tr>
<tr>
<td>MBR</td>
<td>Soluble enzyme</td>
<td>Complex model</td>
</tr>
<tr>
<td></td>
<td>Low enzyme damage</td>
<td>Poor control</td>
</tr>
<tr>
<td></td>
<td>High conversion</td>
<td>Membrane fouling</td>
</tr>
<tr>
<td></td>
<td>In-situ separation</td>
<td>Flow rate restrictions</td>
</tr>
<tr>
<td></td>
<td>Retention of enzymes/Cofactors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dosing of a reactant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Compartmentalization</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Different enzyme configurations are possible by combining soluble and immobilized enzymes in different ways (see Figure 5.1)
5.1 Process Engineering

Figure 5.1: Immobilized enzyme configurations in a packed bed reactor for multi-enzyme in-pot processes for two enzymes. a) Two sequential beds, b) Alternate beds, c) Mixed enzyme bed, and d) One enzyme bed plus one soluble enzyme. Enzyme 1 (○) and Enzyme 2 (□)

5.1.2 Process control

As a further step in multi-enzyme process implementation, process control must be considered. Controllers are basically designed either to stabilize the process and keep it at given operating conditions which are expected to be optimum or stably reproduce an optimal process trajectory. However, additional and more robust characteristics of the controllers can be achieved if built based on mathematical models [83, 88]. In this case, process control can handle:

- a more efficient use of substrates, cofactors and enzymes
- consistent product quality in time
- rejecting disturbances in the operation
- damping model uncertainties and
- performing recursive parameter estimation to enable tracking of slow process as process trajectory changes.

Process control has an established theory which is applied in several engineering fields [83]. However implementation of automatic control has been a big challenge for both bioprocesses using whole cells (biotransformations) and isolated enzymes (biocatalysis) [43, 89]. Consequently, multi-enzyme processes
have many interesting issues to investigate related to process control like controllability, monitoring, and the use of observers.

In order to design and implement a control strategy, controlled variables in the process are typically selected following a hierarchy, as shown in Figure 5.2. In this case, process control is divided in two layers (i.e. regulatory and supervisory control) [57].

The regulatory control usually handles activities of low complexity. At this level, single-input single-output (SISO) controllers are used e.g. inventory control or a proportional-integral-derivative (PID) controller. The main objective is to locally control secondary measurements which are variables that are easy to measure and easy to control, e.g. temperature, pH, dissolved oxygen concentrations. These types of measurements are easily monitored online, furthermore a manipulated variable with fast and strong effect on the response is sufficient to maintain the desired variable under control. Controllers in this layer are commonly implemented for batch, semi-batch and continuous processes [57].

The supervisory layer usually handles more complex control activities to maintain the main process variables at the desired set point. In this layer, substrate, intermediate product and product concentrations are classified as main process variables [57]. However, the monitoring of these variables is not always a straightforward task in bioprocesses because they often require more sophisticated analytical instrumentation and are rarely implemented online. SISO or
5.1 Process Engineering

multi-input multi-output (MIMO) controllers such as inventory control [90, 91], model predictive control [92, 93], or adaptive control [94], can be implemented for semi-batch and continuous processes. Actually the mentioned controllers are model-based. Hence an enormous advantage is achieved when reliable mathematical models of multi-enzyme processes are formulated. These controllers attempt to correct possible bias in the response. It means that the controller not only keeps the variables at the reference values but also can handle, disturbances and uncertainties of the model. It means that the process can be lead to optimal operating conditions even if the model is not perfect.

Multi-enzyme processes often exhibit biocatalyst inhibition/inactivation due to high concentration of substrates or/and products. For substrate inhibition, concentration of the toxic compound must preferably be maintained at its lowest possible value in the media. Thus, controllers must be designed in such a way that they can receive continuously the signal which corresponds to the measurement of the substrate. In this way, the control action can take place by the actuators supplying just the required amount of substrate that is sufficient to promote the desired reactions. However, multi-enzyme processes also face similar problems with the process monitoring since the main variables are often expensive and difficult to measure online. Consequently, the implementation of process control in a conventional way is difficult i.e. process, sensor, controller (see Figure 5.3) [83].

![Figure 5.3: Block diagram for conventional feedback control strategy](image)

In order to overcome this limitation, mathematical models play again an useful role in the formulation of state estimators also known as 'soft-sensors or observers'. An observer is a virtual measurement of the desired variable which is used by the controller in order to obtain the required control action for the process [43]. The observer also uses the measurements available from the process in order to update the estimated state [95]. In this case, the process including the control loop and the observer is shown in Figure 5.4. The use of observers can definitely be an advantage for multi-enzyme processes, since a reliable model can be formulated following the methodology developed here (Chapter 3 and 4).
A commonly used observer is the Kalman filter which is formulated for linear systems. In a similar manner, the extended Kalman filter can be used as a state estimator for nonlinear systems. However for the solution of it, a linearization of the model around its estimated trajectory is then required. In that way states or variables, that cannot be physically measured, can be estimated ($\hat{x}$) by the observer using the available measurements ($y$) and the process inputs ($u$). Afterwards, the control action can take place for the desired controlled variable [71].

An ideal variable to control in multi-enzyme processes is the enzyme concentrations which is an intrinsic variable to the maximum reaction rate of an enzyme ($V_{max}$) catalyzed reaction under the given conditions. However in reality, $V_{max}$ is function of the process conditions. This means that $V_{max}$ changes if the initial conditions of the process change. Hence, a soft-sensor provides a feasible solution to estimate the enzyme concentrations. In this case, the influence of other process variables on the enzyme concentrations can be investigated and then included into the model structure. In this way, the virtual enzyme concentrations can be estimated to be supplied to the controller without the necessity of physically measuring this variable as a function of time (see Figure 5.4).

A further application using a closed loop process is recursive parameter estimation [43, 96]. This technique is interesting for multi-enzyme processes because, as discussed in Chapters 3 and 4, there are many parameters in the model that are difficult to determine with precision using batch experimental data. Hence, applying recursive parameter estimation, the value of the most uncertain parameters and their variation can be greatly improved since they can be re-estimated over the course of time taking into account the dynamics of the process.
5.1 Process Engineering

5.1.3 Process intensification

Process intensification is the term used to describe the means by which a chemical or biochemical process can be made drastically more efficient. Significant intensification of a process requires a holistic view in order to identify and eliminate the main transport resistances that limit the overall performance [97]. In that way, benefits as lower costs, energy efficient processes, inherently safer design and a means to sustainable development, can be achieved.

A multi-enzyme process is actually considered as a particular case of process intensification. In this case, characteristics such as reduction of process units, avoidance of purification of intermediates and development of green/sustainable process, are relevant for multi-enzyme processes. However, not all these processes are carried out under optimal conditions. Consequently, other alternatives can be explored in order to find more compact and efficient processes. As mentioned in the previous section, substrate and product inhibitions are important limitations in multi-enzyme processes. In this case the inhibitory or toxic effects of those compounds lead to ineffective use of the enzyme and/or promote unfavorable equilibria in the reactions. Consequently, toxic substrates must be constantly supplied in low amounts and toxic products need to be removed from the media as soon as they are formed.

*In-situ* product removal (ISPR) techniques can offer different possibilities in order to reduce the concentration of the toxic product in the media. In this case, the separation of the product can be achieved by methods such as adsorption, extraction, membrane separation or pervaporation. Hence the implementation of ISPR in a process is considered as an intensification of the process. In this way more intensified multi-enzyme processes can be achieved exploiting the benefits of these techniques. Further advantages of ISPR are to increase the half life of the enzyme, improve process efficiency, especially in the reaction conversion, and decrease downstream operation costs [98, 99].

One ISPR technique is based on the use of adsorbent resins [100]. In practice, the resins are able to adsorb the desired compound (e.g. the toxic product) due to their affinity. In the same way, the adsorbed product can easily be recovered by washing the resins with a proper solvent. In the process, the resins are often added into the reactor. This means that the reactor volume for the process is reduced which is a disadvantage. Consequently, the amount of resin must be added in an optimal way in order to minimize the volume occupied by the resins in the reactor volume. For this purpose, mathematical models can help in the design of the process operation.

Another example of process intensification is the integration of bioreactors with
a membrane separation process [73]. In this case certain components can be removed from the reactor depending on the membrane characteristics.

5.2 Enzyme immobilization technology

By enzyme immobilization, an enzyme is attached to an inert and insoluble material. In this case, the mobility of the enzyme is limited by applying a chemical or physical treatment. The main goal of enzyme immobilization is to facilitate the separation of the enzyme from the product such that the enzyme may be reused. Furthermore, enzyme stability is one of the characteristics that is improved by this technology. Consequently, enzymes can be resistant to changes in conditions such as pH or temperature for longer operating time [101]. In some cases, enzyme immobilization also achieves better enzyme performance, e.g. improvements in enzyme activity and compatibility which are characteristics desired for multi-enzyme processes. These cases occur when the immobilization technique makes the active side of enzymes more accessible and then the substrate-enzyme binding is easier. Further benefits are the ability to activate or stop the reactions rapidly by adding or removing the enzymes from the reaction solution (tea-bag process), easier downstream processing, no product contamination with the enzyme, and a considerable process cost reduction [102].

As mentioned, some enzyme characteristics are changed after the immobilization. This also means that the kinetics of the immobilized enzyme are modified. One reason for this is the mass transfer problems that are faced with immobilized enzymes. However the degree of influence depends on the utilized immobilization technique. Consequently, further kinetic studies must be carried out in order to estimate and validate the kinetic parameter values for the new format of the enzymes. Hence, kinetic parameters of free enzymes are unsuitable for the mathematical model of multi-enzyme processes where immobilized enzymes are involved.

For a multi-enzyme in-pot process, enzyme immobilization opens the opportunity to promote more innovative enzyme configurations in the reactor design. In this way, issues to be addressed are how the mixture of enzymes should be immobilized? Should it be done individually for each enzyme, or should several enzymes be immobilized on the same support? And, which immobilization methods are best to achieve the best characteristics of all the enzymes involved in the process?. In this case, Figure 5.5) shows a schematic drawing of potential enzyme configurations in a stirred tank reactor. A further scheme is illustrated in Figure 5.1 where different configurations of the immobilized enzymes are
organized in a packed bed reactor.

![Figure 5.5: Different immobilization schemes for multi-enzyme processes (2 enzymes) a) free enzymes, b) individual enzyme immobilized, c) one free and one immobilized d) both enzymes immobilized on the same support](image)

Material science will also have a role to play specifically in the introduction of novel materials with special properties that can be used for successful or improved immobilization of enzymes. An interesting example is the covalent immobilization of an enzyme-cofactor-enzyme system. Here, lactate dehydrogenase (LDH) (EC 1.1.1.27), glucose dehydrogenase (GDH) (EC 1.1.99.10), and cofactor NADH were incorporated into two porous silica glass supports. Effective regeneration cycles of NADH/NAD⁺ are observed and enzyme activities were improved when smaller pores were used. Thus, the nano-porous structure of the glass supports could enhance the molecular interactions among the immobilized enzymes and cofactor, thus improving the catalytic efficiency of the system [103, 104].

### 5.3 Protein Engineering

The main scope of protein engineering is based on developing useful and valuable proteins by improving the existing characteristics of a protein or generating new properties [105-108]. In fact, the most relevant biocatalyst properties that can be designed are the enzyme activity, selectivity and molecular stability. Precisely, the protein folding (structure) is engineered to offer thermodynamic stability as well as thermal and environmental stability, and the protein function which includes improvements in binding properties and catalysis performance and selectivity [16, 17].

Studies of microorganisms that live under extreme environmental conditions
have provided important insights about the structural stability of enzymes [80]. Hence, enzymes keep most of their characteristics outside of the cell. Taking advantage of that, microorganisms can be genetically modified in order to get enzymes with wider operational ranges in pH and temperature, as well as a higher level of tolerance at high media concentration. One example of enzyme improvements were developed for an esterase enzyme in which the thermostability was increased without compromising its catalytic activity at lower temperatures [109]. A further biocatalyst improvement was obtained for (S)-aminotransferase which is used in the production of a chiral amine. In this case not only the specific activity was improved from 5.9 to 1582.8 IU/g, but also the thermostability in that the enzyme was able to operate at temperatures greater then 50 °C for an extended period of time [110].

The above examples illustrate that an enzyme may be ‘designed’ until the ideal enzyme characteristics are achieved for a given process. Actually, this concept is well illustrated by Burton and coworkers, 2002 [80]. This paper showed the main biocatalyst properties that are desired to be improved in order to find the ideal biocatalyst for an optimal bioprocess application. Furthermore, it was discussed how those desired changes can be achieved by the available protein engineering techniques [80].

The idea of designing a biocatalyst is then to achieve specific characteristics according to the requirements of the process [80, 111]. In fact biocatalyst performance is the key factor to improve the overall yield in reactions and process cycle time. It means that having for example a high specific activity and a stable biocatalyst, a significant increase in productivity and decrease in process cost can be obtained [20, 112].

As mentioned, enzymes work best around their optimal condition. However in a multi-enzyme process all enzymes must act under the same process condition and then it is difficult to determine a suitable point where all enzymes can best match in the media. In case of a wrong operating conditions, a fast deterioration and loss of activity of some enzymes may be experienced. Consequently, biocatalysts must be replaced more often in the process.

Figure 5.6 shows the operability area that is generated when two enzymes have different optimal operating conditions. In this case, the question is which enzyme has priority in the process (so better conditions can be offered). The priority can be the cost, the stability, the availability, the activity, the purity, among others. If the priority is given to the enzyme 1, the performance of the enzyme 2 must be compromised. Likewise for the opposite case. Another option is to find an optimal point within the operating region. Thus both enzymes are not working under their optimal conditions but probably in a feasible environment for working together.
However, why is it required to compromise the performance of the enzymes in the process when protein engineering can modify them in order to make them work under the same optimal operating conditions? In this case, mathematical models can also play an interesting role in the identification of potential biocatalyst modifications e.g. the enzyme activity or tolerance to high media concentration. These and other types of characteristics can be analyzed virtually by simulating the process under different scenarios until the best option is found. In this case, clear goals can be provided to protein engineering in order to improve the process and render the implementation of multi-enzyme processes at an industrial level feasible. However, new experimental kinetic studies are required in order to re-calibrate the mathematical models, if enhanced enzymes are included in the multi-enzyme process.

5.4 List of symbols and nomenclature

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR</td>
<td>Stirred tank reactor</td>
</tr>
<tr>
<td>PFR</td>
<td>Plug flow reactor</td>
</tr>
<tr>
<td>PBR</td>
<td>Packed bed reactors</td>
</tr>
<tr>
<td>MBR</td>
<td>Membrane bioreactor</td>
</tr>
<tr>
<td>SISO</td>
<td>Single-input single-output</td>
</tr>
<tr>
<td>PID</td>
<td>Proportional-integral-derivative controller</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>$y$</td>
<td>Output model</td>
</tr>
<tr>
<td>$u$</td>
<td>Control action</td>
</tr>
<tr>
<td>$e$</td>
<td>Error between measurement and setpoint value</td>
</tr>
<tr>
<td>$\hat{x}$</td>
<td>State estimated</td>
</tr>
</tbody>
</table>

**Nomenclature**

$V_{max}$  Maximum reaction rate of an enzyme ($Ms^{-1}$)
Part II

Case studies
Chapter 6

Introduction to case studies

As mentioned at the beginning of this thesis, the main aim of the proposed methodology is to guide process model building in order to obtain improved models which can be used for process design, control, optimization, prediction and scale-up. In order to achieve this, a sequence of steps must be followed, as described in the methodology. A summary of the procedure is shown in Figure 6.1.

Figure 6.1 indicates that the model building procedure requires some initial information here that serves as input for the model. This information is mainly to be found in the scientific literature. From here, a preliminary model is required which can be taken either from the scientific literature directly or built based on the available information. Afterwards, some model analyses can be carried out in order to evaluate its reliability. In case the reliability is not met, the next step is to change either the model structure or to collect more (better) experimental data, if possible. From here, the model is re-evaluated until it is improved, as necessary. The final step is to ask if there are sufficient experimental data to validate the model. If not, more experiments are required. If positive, the model is ready to use in different applications.

In this thesis three case studies have been selected in order to build, test and verify the methodology developed here. In this way, valuable contributions have been obtained from each case study not only for improved process understanding.
Figure 6.1: General procedure scheme to achieve model improvements. Superscripts 1, 2, and 3 indicate the sequence of steps followed in each case study, also named with the same number.
but also to refine the requirements for each step in the methodology. The three case studies are:

- Case Study 1. Bi-enzyme production of lactobionic acid
- Case Study 2. Bi-enzyme production of N-acetyl-D-neuraminic acid
- Case Study 3. Tri-enzyme production of 1-phenylethylamine

The three case studies show different characteristics such as the model structure and complexity, and the information that was available in the first place. Figure 6.1 shows the sequence of steps that was followed for each case study.

For the first case study (see the sequence of steps indicated by the superscript 1 in Figure 6.1), the model input was quite substantial since most of the information was provided by Ghent University. It also included a preliminary model structure and parameters. In this way most effort was placed on how to develop, implement and study the different model analyses for the application of multi-enzyme processes. At this point in the thesis, this case study was highly important to structure the MATLAB codes and to evaluate the necessity of the techniques for model analysis in order to achieve a better understanding of the models, as well as the interpretation of the results in order to have a clear direction about how to achieve model improvements. This case study discusses at the end how the model can be improved (see Section 7.6). However, those suggestions were not performed since the focus for this case was on developing the methodology.

The second case study had a slightly different starting point and then further capabilities of the methodology were tested. In this case (see the sequence of steps indicated by the superscript 2 in Figure 6.1), the process information was found in the scientific literature. The preliminary model was partly obtained from literature and partly built by doing the work presented in this thesis. In this way, this case study was focused on testing the methodology and model analysis capabilities to handle model structure simplification. This task was highly important to verify the methodology since one of the hypotheses to be tested was that the preliminary models are often overparametrized for multi-enzyme processes, because they are built by a combination of kinetic expressions from single enzymes, as mentioned in Section 3.6.5. The hypothesis was proven for this case study since elimination of four kinetic parameters out of fifteen was possible without compromising the dynamic behavior of the process. However, in order to achieve a feasible application of this simplified model, a validation would be necessary.
In the third case study (see the sequence of steps indicated by the superscript 3 in Figure 6.1), the preliminary model was used to analyze potential process configurations in order to improve the productivity of the process. This evaluation was relevant to analyze, since the production of 1-phenylethylamine shows interesting reaction limitations which make it difficult to carry the reaction out under conventional conditions e.g. in a batch stirred tank reactor. The use of multiple enzymes for this process is already an improvement of the transamination although not sufficient to make the process feasible at an industrial level. Therefore what was shown in this case study is how dynamic simulation of the preliminary model can also be used as a screening tool to select potential process configurations that can be relevant for a more detailed analysis.

In order to give an overview of the second part of this thesis, Table 6.1 summarizes all the case studies, including the type of information that was found in the scientific literature and the calculations that were performed following the steps described by the methodology developed in this thesis.
Table 6.1: Compilation of the case studies

<table>
<thead>
<tr>
<th></th>
<th>Lactobionic acid</th>
<th>Neu5Ac</th>
<th>1-phenylethylamine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1. Modeling objective</strong></td>
<td>Testing Methodology</td>
<td>Model simplification</td>
<td>Understanding and operating mode</td>
</tr>
</tbody>
</table>

**Step 2. Information**

*About the reactions*

<table>
<thead>
<tr>
<th>Compounds involved</th>
<th>4 components</th>
<th>4 components</th>
<th>7 components</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 enzymes</td>
<td>2 enzymes</td>
<td>3 enzymes</td>
<td></td>
</tr>
<tr>
<td>1 redox mediator</td>
<td></td>
<td></td>
<td>1 cofactor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Lactose</th>
<th>GlcNAc</th>
<th>Acetophenone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>Pyruvate</td>
<td>L-alanine</td>
<td>GLucose</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intermediates</th>
<th>Lactobiono lactone</th>
<th>ManNAc</th>
<th>Pyruvate</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Products</th>
<th>Lactobionic acid</th>
<th>Neu5Ac</th>
<th>1-phenylethylamine</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Enzyme and mechanism</th>
<th>CDH: Substituted Epimerase: reversible</th>
<th>TAm: Substituted</th>
<th>LDH: Compulsory order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laccase: Substituted</td>
<td>Michaelis-Menten</td>
<td>GDH: Compulsory order</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aldolase: bi-uni</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-enzymatic reactions</th>
<th>Spontaneous hydrolysis</th>
<th>Not applicable</th>
<th>Not applicable</th>
</tr>
</thead>
</table>

|-------------------------|------------------------|------------------|---------------------|

<table>
<thead>
<tr>
<th>Reaction structure</th>
<th>Network</th>
<th>Cascade</th>
<th>Network</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Main interactions</th>
<th>Not reported</th>
<th>Competitive inhibition</th>
<th>Inhibition by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>by pyruvate and Neu5Ac</td>
<td>Acetophenone and PEA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>About the process</th>
<th>Batch</th>
<th>Fed-batch</th>
<th>Batch and Fed-batch</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Operating mode</th>
<th>Batch</th>
<th>Fed-batch</th>
<th>Batch and Fed-batch</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Type of reactor</th>
<th>Stirred tank reactor</th>
<th>Stirred tank reactor</th>
<th>Stirred tank reactor</th>
</tr>
</thead>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Step 3. Formulation</th>
<th>Lactobionic acid</th>
<th>Neu5Ac</th>
<th>1-phenylethylamine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservation equations</td>
<td>6 ODEs</td>
<td>5 ODEs</td>
<td>10 ODEs</td>
</tr>
<tr>
<td>Constitutive equations</td>
<td>2 enzymatic reaction rates</td>
<td>2 enzymatic reaction rates</td>
<td>3 enzymatic reaction rates</td>
</tr>
<tr>
<td>- CDH: 3 parameters</td>
<td>- Epimerase: 6 parameters</td>
<td>- TAM: 12 parameters</td>
<td></td>
</tr>
<tr>
<td>- Laccase: 3 parameters</td>
<td>- Aldolase: 9 parameters</td>
<td>- LDH: 4 parameters</td>
<td></td>
</tr>
<tr>
<td>- Hydrolysis reaction rate</td>
<td>- GDH: 4 parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- LBL: 1 parameter</td>
<td>- Oxygen: 2 parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total no. parameters</td>
<td>9</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Conditions</td>
<td>No</td>
<td>2 equilibrium constants</td>
<td>1 equilibrium constant</td>
</tr>
<tr>
<td>Step 4. Calculations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model implementation</td>
<td>Performed (Section 7.5.1)</td>
<td>Performed (Section 8.5.1)</td>
<td>Performed (Section 9.5.1)</td>
</tr>
<tr>
<td>Identifiability analysis</td>
<td>Performed (Section 7.5.2)</td>
<td>Not performed</td>
<td>Not performed</td>
</tr>
<tr>
<td>Parameter estimation</td>
<td>Performed (Section 7.5.3)</td>
<td>Not performed</td>
<td>Not performed</td>
</tr>
<tr>
<td>Uncertainty analysis</td>
<td>Performed (Section 7.5.4)</td>
<td>Not performed</td>
<td>Not performed</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>Performed (Section 7.5.5)</td>
<td>Performed (Section 8.5.2)</td>
<td>Performed (Section 9.5.2)</td>
</tr>
<tr>
<td>Optimal exp. design</td>
<td>Not performed</td>
<td>Not performed</td>
<td>Not performed</td>
</tr>
<tr>
<td>Step 5. Experimentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continued on next page</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lactobionic acid</td>
<td>Neu5Ac</td>
<td>1-phenylethylamine</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Boundary conditions</strong></td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>Progress curves</strong></td>
<td>Available [27]</td>
<td>Not available</td>
<td>Not available</td>
</tr>
<tr>
<td></td>
<td>Lactose</td>
<td>Oxygen</td>
<td>LBA</td>
</tr>
</tbody>
</table>
6.1 List of symbols

Abbreviations

GlcNAc  N-acetyl-D-glucosamine
ManNAc  N-acetyl-D-mannosamine
Neu5Ac  N-acetyl-D-neuraminic acid
ABTS  2,2’-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt
CDH  Cellobiose Dehydrogenase
TAm   Transaminase bulk
LDH   Lactate dehydrogenase
GDH   Glucose dehydrogenase
PEA   1-phenylethylamine
ODE   Ordinary differential equation
LBL   Lactobionolactone
LBA   Lactobionic acid
Case study 1: Bi-enzyme production of lactobionic acid

7.1 General description

Lactobionic acid (4-O-β-D-galactopyranosyl-D-gluconic acid) and its salts are compounds used in the production of high-value products, pharmaceutical and food technology applications. Production of lactobionic acid is mainly achieved by the oxidation of lactose (see Figure 7.1). Different processes have been reported in the scientific literature for lactobionic acid production, such as: chemical, electrochemical, catalytic, microbial, and enzymatic processes [118]. Some of the reported enzymes to carry out an enzymatic process are: lactose dehydrogenase (EC 1.1.1.27), hexose oxidase (EC 1.1.3.5), carbohydrate oxidase (EC 1.1.3.4) e.g. lactase oxidase, glucooligosaccharide oxidase, glucose-fructose oxidoreductase, and cellobiose dehydrogenase (CDH) [118]. For these enzymatic reactions, the application of molecular oxygen was not effective as a direct electron acceptor in the reaction. Consequently, the enzyme laccase was investigated as electron acceptor to re-oxidize the CDH [119]. A subsequent investigation [120] demonstrated that a combination of two enzymes (CDH and laccase) in the presence of a redox mediator has produced a successful application of a
multi-enzyme system carried out in a single reactor.

In this case, a complete and efficient conversion of lactose to lactobionic acid with no by-product has been achieved. Furthermore, the redox mediator could be continuously reoxidized due to the use of laccase for the regeneration of various electron acceptors used by the CDH. In this way, only the required amount of these compounds is added to the process with no additional by-product since the reaction between laccase and oxygen only generates water [27]. In this way, the bi-enzyme production of lactobionic acid is then a clear example of how two biocatalysts can work in a synergistic manner in order to generate a desired product, furthermore minimizing undesired parallel reactions, e.g. formation of hydrogen peroxide that inhibits the enzyme activity [118].

7.2 Step 1: Modeling objective

In this case study, the main goal is to analyze the reliability of the mathematical model of the process in order to evaluate both prediction quality of the model and identification of critical factors that affect the model outputs.

7.3 Step 2: Information

For this case study, the mathematical model is available in the scientific literature including the kinetic parameters of the multi-enzyme process. Furthermore, progress curves of the process were provided by Dr. Wouter Van Hecke and Prof. Herman Van Langenhove (Ghent University, Belgium). In this way, the information is organized and described here, according the modeling methodology, as follows:

7.3.1 Reaction considerations

- Figure 7.1 shows the general scheme for the production of lactobionic acid. This illustrates the different reactions involved in the process. The first enzyme cellobiose dehydrogenase (CDH) catalyzes the dehydrogenation of lactose to lactobionic-δ-lactone, which is spontaneously hydrolyzed to lactobionic acid. In this case, the double action of the redox mediator 2,2'-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) is exploited. In the first reaction, it acts as an electron acceptor regenerating the initial
oxidation state of the first enzyme (CDH). And in the second reaction, ABTS serves as electron donor to obtain the reduction of laccase, which is the second enzyme added to the system. The reduced state of laccase catalyzes the second reaction where oxygen (the co-substrate) is fully reduced to water [27, 120]. Furthermore, it was found that the enzyme laccase produced a reduced ABTS and stable monocation ABTS\(^+\) mixture with no evidence of the dication ABTS\(^{++}\) formation [121].

- Both enzymes involved in the process (CDH and laccase) follow the substituted-enzyme mechanism (see Equation A.10 in Appendix A). Some kinetic parameters for each enzyme are obtained from the reported literature [27, 120, 122]. This means that interaction due to the combination of enzymes was not taken into account in their studies. The reported values are compiled in Table 7.3 in the calculations step.

- Table 7.1 shows the interactions that exist between compounds when both enzymatic reactions are combined in a multi-enzyme process. In this case, lactose is the substrate (S) for CDH and no interaction (X) with laccase; lactobionic lactone is the intermediate product (IP) for CDH and no interaction (X) with laccase; lactobionic acid is the product (P) for CDH and no interaction (X) with laccase; ABTS is the product (P) for CDH and substrate (S) for laccase; ABTS\(^+\) is the substrate (S) for CDH and product (P) for laccase, and oxygen is the substrate (S) for laccase and no interaction with CDH.

### 7.3.2 Process considerations

- The bi-enzyme process is carried out in batch mode.
Table 7.1: Interaction matrix for the bi-enzyme production of lactobionic acid

<table>
<thead>
<tr>
<th>Compound</th>
<th>Enzyme</th>
<th>CDH</th>
<th>Laccase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactose</td>
<td></td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>Lactobionolacto</td>
<td>IP</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lactibionic acid</td>
<td>P</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ABTS</td>
<td>P</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>ABTS+</td>
<td>S</td>
<td></td>
<td>P</td>
</tr>
<tr>
<td>Oxygen</td>
<td>X</td>
<td></td>
<td>S</td>
</tr>
</tbody>
</table>

(S) substrate, (P) product, (X) no interaction (IP) intermediate product

- For this process a membrane bioreactor was used. The main purpose of this reactor was to provide bubble-free oxygenation. Furthermore, the mass transfer process was taken into account in the mathematical model [87].

- Initial conditions for all components in the process are also reported in literature. Furthermore, the values are compiled in Table 7.3.

- During the process the temperature was controlled at 30°C and pH was maintained at 3.9. Furthermore, concentrations of lactose, lactobionic acid and oxygen were measured during 6 hours. After that time, the lactose was completely consumed. The sampling time for lactose and lactobionic acid was 1 hour and the samples were measured by using HPLC. The dissolved oxygen measurements were recorded every 10 seconds.

7.4 Step 3: Formulation

7.4.1 Model assumptions

According to the information compiled in Step 2 and literature [27], the following assumptions can be extracted for the mathematical model.
7.4 Step 3: Formulation

Figure 7.2: Reactor scheme and balance volume of the process. (a) balance volume for the reactor (BSTR) and (b) balance volume for the enzymes (CDH and laccase)

- Initial kinetic parameter values are taken from literature
- Substrate and product inhibitions are neglected in the process
- pH and temperature are maintained constant during the operation.
- Perfect mixing in the reactor

7.4.2 Boundaries and balance volume

Figure 7.2 shows a scheme of the type of reactor and the balance volumes that are assumed for the reactor and both enzymes involved in the process.

7.4.3 Conservation equations

The stoichiometry of the bioreaction is illustrated in Figure 7.1. Based on that, a set of differential equations was formulated for a batch operating mode, including component balances for concentrations of substrates (lactose and oxygen), redox mediator (ABTS), intermediate products (lactobiono-lactone), and main product (lactobionic acid). Thus, the model for the system consists of six (6) differential equations that can be written in a matrix notation, as shown in Table 7.2
Table 7.2: Mass balances of the process represented by the stoichiometric matrix notation

<table>
<thead>
<tr>
<th>Reaction</th>
<th>$C_{Lact}$ (mM)</th>
<th>$C_{LBB}$ (mM)</th>
<th>$C_{LBA}$ (mM)</th>
<th>$C_{O2}$ (mM)</th>
<th>$C_{ABTS}$ (mM)</th>
<th>$C_{ABTS^+}$ (mM)</th>
<th>Kinetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td>$r_{CDH}$</td>
</tr>
<tr>
<td>Laccase</td>
<td>-1/2</td>
<td>-1</td>
<td>-2</td>
<td>2</td>
<td></td>
<td></td>
<td>$r_{lacc}$</td>
</tr>
<tr>
<td>Hydrolysis</td>
<td>-1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r_{hyd}$</td>
</tr>
<tr>
<td>Aeration</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r_{omt}$</td>
</tr>
</tbody>
</table>

lacc: laccase, hyd: hydrolysis, omt: oxygen mass transfer

An example, of how the matrix should be read, is shown with the balance of oxygen, thus,

$$\frac{dC_{O2}}{dt} = r_{omt} - \frac{1}{2}r_{lacc} \quad (7.1)$$

7.4.4 Constitutive equations

Figure 7.1 and Table 7.2 show that each enzyme (CDH and laccase) interacts mainly with two different substrates, and consequently two different products are formed. Furthermore, CDH and laccase catalyze the formation and release of a product before all the substrates have bound. This type of mechanism is called ping-pong bi-bi or substituted-enzyme [18, 62]. The general structure of the enzyme mechanism is illustrated in Figure A.8

In this case study, both enzymes follow the same type of mechanism. Hence, two coupled substituted-enzyme mechanisms are suggested to describe both enzymatic reactions, as shown in Figure 7.3.

Based on the double substituted-enzyme mechanism, the initial reaction rate for CDH may be derived as:

$$r_{CDH} = V_{CDH}^{max} \frac{C_{Lact} \cdot C_{ABTS^+}}{K_{M}^{lact} \cdot C_{ABTS^+} + K_{M}^{lact} \cdot C_{Lact} + C_{lact} \cdot C_{ABTS^+}} \quad (7.2)$$
And similarly the initial reaction rate for laccase may be derived as:

\[
    r_{\text{lacc}} = \frac{V_{\text{max}}}{K_M} \cdot \frac{C_{O_2} \cdot C_{\text{ABTS}}}{K_{M} \cdot C_{\text{ABTS}} + C_{O_2} + C_{O_2} \cdot C_{\text{ABTS}}} \quad (7.3)
\]

In both cases, the initial reaction rate has been assumed. This means that the presence of products is neglected. In the reaction rate expressions, \( K_M \) represents the Michaelis-Menten constant for each substrate involved in the enzymatic reactions and \( V_{\text{max}} \) represents the maximum reaction rate.

Furthermore, two sets of kinetics are involved in this process. One reaction rate due to the hydrolysis from lactobionolactone to lactobionic acid, which is described by:

\[
    r_{\text{hyd}} = K_{\text{hyd}} \cdot C_{\text{LBL}} \quad (7.4)
\]

where, the parameter \( K_{\text{hyd}} \) represents the hydrolysis constant. Furthermore, the mass transfer rate of oxygen is included in the model to describe the aeration process where oxygen migrates from the gas phase (air / \( O_2 \)) to the liquid.

\[
    r_{\text{omt}} = K_L a \left( C_{\text{out}} - C_{O_2} \right) \quad (7.5)
\]
Table 7.3: Initial operating conditions and parameters for the lactobionic process [27]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Units</th>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_0^{\text{lact}}$</td>
<td>50</td>
<td>mM</td>
<td>$V_{\text{CDH max}}^{\text{act}}$</td>
<td>26.76</td>
<td>mM/h</td>
</tr>
<tr>
<td>$C_0^{\text{LBL}}$</td>
<td>0</td>
<td>mM</td>
<td>$K_{M}^{\text{act}}$</td>
<td>2.4</td>
<td>mM</td>
</tr>
<tr>
<td>$C_0^{\text{LBA}}$</td>
<td>0</td>
<td>mM</td>
<td>$K_{M}^{\text{ABTS}^+}$</td>
<td>0.0004</td>
<td>mM</td>
</tr>
<tr>
<td>$C_0^{O_2}$</td>
<td>1.19</td>
<td>mM</td>
<td>$V_{\text{lacc max}}^{\text{max}}$</td>
<td>120</td>
<td>mM/h</td>
</tr>
<tr>
<td>$C_0^{\text{ABTS}}$</td>
<td>180</td>
<td>mM</td>
<td>$K_{M}^{O_2}$</td>
<td>0.41</td>
<td>mM</td>
</tr>
<tr>
<td>$C_0^{\text{ABTS}^+}$</td>
<td>0</td>
<td>mM</td>
<td>$K_{M}^{\text{ABTS}^+}$</td>
<td>0.033</td>
<td>mM</td>
</tr>
<tr>
<td>$K_La$</td>
<td>3.636</td>
<td>$h^{-1}$</td>
<td>$C_{O_2}^{\text{sat}}$</td>
<td>1.19</td>
<td>mM</td>
</tr>
<tr>
<td>$K_{\text{hyd}}$</td>
<td>0.7704</td>
<td>$h^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

where $K_{La}$ is the mass transfer coefficient and $C_{O_2}^{\text{sat}}$ is the saturation concentration of oxygen which is assumed constant in the media.

7.5 Step 4: Calculations

The information collected in the Steps 2 and 3 is used in order to implement and analyze the reliability of the model according the modeling objective. Initial conditions for the system and parameter values are reported in the scientific literature [27, 120, 122]. These values are summarized in Table 7.3.

7.5.1 Model implementation and dynamic simulation

Having defined the mass balances of the process and the kinetic parameter of the reaction rates, the set of differential equations can be solved to analyze the dynamics of the process. In this case, initial operating conditions and parameter values are required for the process simulation, as presented in Table 7.3. Additionally, progress curves are used in this step to compare the predicted model outputs with the corresponding experimental data points, as shown in Figure 7.4.
7.5 Step 4: Calculations

Figure 7.4: Dynamic simulation of the bi-enzyme production of lactobionic acid using parameters reported in the literature [27]. Experimental data (•) and model simulation (-).

119
In Figure 7.4, the dynamic behavior of the model follows the experimental progress curves. However, the experimental concentration of lactobionic acid is higher than the concentration predicted by the model. The deviation, which was also reported in the literature \[27\], was assigned to a measurement error in the lactobionic acid concentration. It was assumed that the concentration of lactobionolactone can disturb the real measurement of LBA.

However, a deviation can be expected from the actual model since the reported kinetic parameters were obtained for single enzyme experimental data. This can mean that the combination of enzymes, substrates and products in the same reactor can affect the general kinetic behavior of the system. Following the last hypothesis, the model needs to be re-calibrated using the progress curves of the process. For this case study, the kinetic parameters are estimated following the criteria described in Chapter 4, specifically in Section 4.3.

7.5.2 Identifiability analysis

Due to the complexity of the multi-enzyme processes, identifiability of the parameters involved must be verified. For this case, the identifiability analysis was performed by following the calculations as explained in Section 4.2, Chapter 4. For these calculations, the experimental information, the total number of model parameters, and initial parameter values are required as input (see Figure B.3). The suggested bi-enzyme model has 9 parameters of which only 7 could be identified from the available information. Hence 2 parameter values must be fixed to obtain a collinearity index no greater than 15, which is the threshold. In this case, \(C_{O_2}\) and \(K_{O_2}\) were fixed at their reported values. Consequently, the estimation of the other parameters is performed, as shown in Section 7.5.3.

7.5.3 Parameter estimation, confidence intervals and correlation matrix

The parameter estimation of the model is performed by minimizing the cost function, as described in Chapter 4, more specifically in Equation 4.5. The cost function formulation includes the standard deviation of measurements \(\sigma_k\). In this case, the standard deviation of each measurement was obtained by performing a linear regression of the experimental data specifically for those points where the process has the most constant behavior, specifically, between 14.1-39.5 mM for lactose, 5.9-36.5 mM and 0.36-1.03 mM. Figure 7.5 shows the fits in a parity plot between the experimental data and the linear model.
7.5 Step 4: Calculations

**Table 7.4**: Standard deviation values for the measured variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of points</th>
<th>Correlation coeff.</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactose</td>
<td>4</td>
<td>0.9993</td>
<td>0.25858 mM</td>
</tr>
<tr>
<td>LBA</td>
<td>5</td>
<td>0.9986</td>
<td>0.40632 mM</td>
</tr>
<tr>
<td>Oxygen</td>
<td>38</td>
<td>0.9989</td>
<td>0.0065898 mM</td>
</tr>
</tbody>
</table>

**Figure 7.5**: Parity plot for lactose, LBA and O$_2$ variables. (●) experimental data, (-) linear model fit

Furthermore, Table 7.4 shows the standard deviation values obtained from these regressions. Table 7.4 shows that the correlation coefficients are close to the unity. This increases the reliability of the standard deviation values of the measurements. Having these values, the weighted least square method can be applied for the parameter estimation (see Equation 4.5). In this case, the parameter values reported in literature are taken as the initial guess for the estimation routing (see Table 7.3). As a result, the kinetic parameters are estimated and the corresponding values are shown in Table 7.5. Furthermore, the estimated parameter values also includes the calculation of a 95% confidence interval of the parameters. Figure 7.6 compares the dynamic behavior between experimental data and the model outputs using the estimated parameters.

In Figure 7.6, a better fitting of lactose and oxygen is achieved when compared with the previous fit (see Figure 7.6). Furthermore, the lactobionic acid fit was improved, but still with a small deviation.

In Table 7.5, the parameters $K_{M}^{ABTS^-}$ and $K_{M}^{ABTS}$ have a large confidence band and in this case they can not really be considered to be reliably identifiable.
Figure 7.6: Comparison between experimental data and simulation of the system using the estimated parameters (see Table 7.5). (o) experimental data, (-) model.

**Case study 1: Bi-enzyme production of lactobionic acid**
7.5 Step 4: Calculations

Table 7.5: Estimated parameter values for the model including a 95% confidence interval

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{CDH_{max}}^{CDH}$</td>
<td>23.33 ± 16.4</td>
<td>mM/h</td>
</tr>
<tr>
<td>$K_{M}^{lect}$</td>
<td>1.27 ± 3.06</td>
<td>mM</td>
</tr>
<tr>
<td>$K_{M}^{ABTS}$</td>
<td>$4.1 \times 10^{-5}$ ± 0.09</td>
<td>mM</td>
</tr>
<tr>
<td>$V_{lacc_{max}}^{lacc}$</td>
<td>58.48 ± 34.7</td>
<td>mM/h</td>
</tr>
<tr>
<td>$K_{M}^{ABTS}$</td>
<td>$8.74 \times 10^{-3}$ ± 0.51</td>
<td>mM</td>
</tr>
<tr>
<td>$K_{L_a}$</td>
<td>3.84 ± 0.10</td>
<td>h$^{-1}$</td>
</tr>
<tr>
<td>$K_{hyd}$</td>
<td>0.655 ± 0.44</td>
<td>mM/h</td>
</tr>
</tbody>
</table>

Consequently, the correlation matrix is calculated by using Equation 4.8 in order to analyze the relationship between the estimated parameters. The correlation matrix is shown in Table 7.6.

From Table 7.6, a strong relationship between the parameters $K_{M}^{ABTS}$ and $V_{CDH_{max}}^{CDH}$ is observed. This indicates that any change in one parameter could be compensated by a similar change in other. This is the reason why it is difficult to find a unique solution for these parameters. Furthermore, these results also agree with the identifiability problems that were obtained previously. Consequently, further analysis of the parameter is required. The estimated parameters are then studied in more detail by applying uncertainty and sensitivity analyses.

7.5.4 Uncertainty analysis: Monte Carlo Method

In order to calculate the uncertainty of the model, the Monte Carlo method is performed by following the systematic steps described in Chapter 4, specifically in Section 4.4. In order to generate the parameter sampling, the Latin hypercube sampling (LHS) is applied. In this case, the following input information is required: number of samples (m=500), number of parameters (n = 7), correlation matrix and confidence intervals. Figure 7.7 shows the random sampling generated by applying LHS with correlation control. This means that the generated distribution is based on the correlation matrix.

Having the random sampling, the model evaluation can be performed for each set of parameters. Finally the uncertainty is represented by calculating the mean
**Case study 1: Bi-enzyme production of lactobionic acid**

Table 7.6: Correlation matrix of the parameter in the model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\theta_5$</th>
<th>$\theta_6$</th>
<th>$\theta_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{CDH}^{max}$</td>
<td>$\theta_1$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{lact}^{M}$</td>
<td>$\theta_2$</td>
<td>$-0.47$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{ABTS}^{M}$</td>
<td>$\theta_3$</td>
<td>0.85</td>
<td>$-0.71$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{lacc}^{max}$</td>
<td>$\theta_4$</td>
<td>0.29</td>
<td>0.13</td>
<td>$-0.08$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{ABTS}^{M}$</td>
<td>$\theta_5$</td>
<td>0.42</td>
<td>0.18</td>
<td>$-0.06$</td>
<td>0.83</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$K_{Lac}$</td>
<td>$\theta_6$</td>
<td>0.13</td>
<td>0.13</td>
<td>0.23</td>
<td>$-0.07$</td>
<td>$-0.22$</td>
<td>1</td>
</tr>
<tr>
<td>$K_{hyd}$</td>
<td>$\theta_7$</td>
<td>$-0.00$</td>
<td>$0.00$</td>
<td>$-0.00$</td>
<td>$-0.00$</td>
<td>$0.00$</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 7.7**: Random sampling using latin hypercube sampling (LHS) for 500 samples including correlation control.
7.5 Step 4: Calculations

value, and the 10th and 90th percentiles of the distribution of model outputs evaluated as function of time. Figure 7.8 shows the uncertainty in the model outputs for the measured variables i.e. concentrations of lactose, oxygen and lactobionic acid. Indeed, the larger the spread of the distribution of the output, the higher the uncertainty.

The uncertainty found in the estimated parameters is reflected by the uncertainty of the model outputs, as shown in Figure 7.8. In this case, the uncertainty for lactose is small while the lactobionic acid uncertainty is constantly increasing. In the same way, the oxygen uncertainty increases during the last stage of the reaction which also corresponds to the depletion of the substrate and then the oxygen concentration is rapidly increased, until the saturation level is achieved. This analysis indicates that the model can provide a reliable prediction for the variables as long as lactose is available in the process. In order to understand the main sources of the model output uncertainty, sensitivity analysis is performed in the next section.

7.5.5 Sensitivity analysis: Morris Method

In order to calculate the sensitivity of the model parameters, the Morris screening method is performed by following the systematic steps described in Chapter 4, specifically in Section 4.5. In order to generate the parameter sampling, the Morris sampling is applied. The input information is: 1) the number of parameters $n$ are the same as estimated, meaning 7 input factors, 2) the number of levels $p$ is chosen as 6, and 3) The optimal perturbation factor is $\Delta$ of 3/5 calculated by using Equation 4.13, and 4) the number of samples $r$, which are used to evaluate the elementary effects per factor, is taken as 30. Consequently, the final cost relies on 240 model evaluations corresponding to the same number of total random samples. As a result, the distribution function $F_i$ which is based on the calculated elementary effects can be evaluated. This result is better illustrated by histograms as shown in Figure 7.9 for certain factors contributing to the oxygen output.

Additionally, the Morris results can be visualized by comparing the mean and standard deviation of the distribution function $F_i$ of each factor as shown in Figure 7.10. Furthermore, the two solid lines help to interpret the type of effects that each factor produces in the outputs. This means that if a factor lies between the two lines the effect of this factor on the output can be considered as insignificant.

In a similar manner, a ranking of the most significant factors can be achieved according to the absolute mean value of the distribution function $F_i$. In this
Figure 7.8: Uncertainty of the Monte Carlo output represented by the mean value and 10th and 90th percentiles.
7.5 Step 4: Calculations

**Figure 7.9:** Distribution of the elementary effect of some parameters in the oxygen output for 240 model simulations

**Table 7.7:** Ranking of the factor according to the absolute mean value of the distribution $F_i$

<table>
<thead>
<tr>
<th>Rank</th>
<th>Lactose</th>
<th>LBA</th>
<th>$O_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V_{lacc,max}^\theta$</td>
<td>$K_{La}$</td>
<td>$V_{lacc,max}^\theta$</td>
</tr>
<tr>
<td>1</td>
<td>$K_{L}^{-0.557}$</td>
<td>0.424</td>
<td>$K_{L}^{0.078}$</td>
</tr>
<tr>
<td>2</td>
<td>$V_{CDH,max}^{\theta}$</td>
<td>$K_{L}^{0.374}$</td>
<td>$V_{lacc,max}^{\theta}$</td>
</tr>
<tr>
<td>3</td>
<td>$K_{ABTS}^{M}^{0.257}$</td>
<td>$K_{ABTS}^{M}^{0.058}$</td>
<td>$V_{CDH,max}^{\theta}$</td>
</tr>
<tr>
<td>4</td>
<td>$K_{ABTS}^{+}^{0.194}$</td>
<td>$V_{lacc,max}^{\theta}$</td>
<td>$K_{ABTS}^{M}^{0.051}$</td>
</tr>
<tr>
<td>5</td>
<td>$K_{lacc}^{M}^{0.078}$</td>
<td>$V_{CDH,max}^{\theta}$</td>
<td>$K_{lacc}^{M}^{-0.048}$</td>
</tr>
<tr>
<td>6</td>
<td>$K_{L}^{0.066}$</td>
<td>$K_{ABTS}^{M}^{+0.032}$</td>
<td>$K_{L}^{0.058}$</td>
</tr>
<tr>
<td>7</td>
<td>$K_{hyd}^{8.87 \times 10^{-10}}$</td>
<td>$K_{hyd}^{1.05 \times 10^{-8}}$</td>
<td>$K_{lacc}^{+0.037}$</td>
</tr>
</tbody>
</table>

...case, the ranking is shown in Table 7.7 where high mean values indicate a more significant effect on the model output.

From the results, parameters $V_{lacc\,max}^{\theta}$, $K_{L}^{\theta}$ and $K_{hyd}^{\theta}$ are identified as the most significant of the model. Therefore they are mainly responsible for the variance in the outputs of lactose, oxygen and lactobionic acid. From this analysis, a direct relation between the oxygen mass transfer coefficient and the oxygen output was expected, as well as the relation between the hydrolysis coefficient with lactobionic acid output, due to their relations in the mass balance equations. However, this analysis shows that the lactose behavior which is the substrate of the first enzymatic reaction (CDH) is also strongly affected by parameters from the second enzymatic reaction (laccase).
Case study 1: Bi-enzyme production of lactobionic acid

Figure 7.10: Evaluated mean and standard deviation of the distribution of elementary effects of the inputs
7.6 Conclusions

The suggested mathematical model can in general explain the process behavior. It was also demonstrated that the parameter estimation could improve the description of the oxygen and the lactose concentration, but at the same time the available experimental information was not sufficient to obtain a better description of the dynamics of the main product (lactobionic acid). 7 parameters were found to be identifiable based on the given data, but the kinetic parameters \( (K_M) \) for both oxidation states of the intermediate redox mediator ABTS are very small which physically means a really fast dynamics in the system. That effect could increase the uncertainty in those parameters obtaining as a result a large confidence interval and thus limitations in the identification of them.

From the uncertainty analysis, the uncertainty propagation for lactose is small, which indicates that the predicted value using the model may be reliable for further applications e.g. process control and prediction. In the same manner, the oxygen prediction is also good when there is sufficient lactose in the reactor because a lack of lactose decreases the consumption of oxygen in the whole reaction. Consequently, the oxygen concentration is increased too fast that the model cannot predict this change with sufficient accuracy.

The decomposition of the input uncertainty in the sensitivity analysis could show that the maximum rate of each enzyme \( (V_{max}) \) is highly relevant for the enzymatic reactions. Consequently, it indicates that the experimental efforts should be more focused on determining those parameters, if further experimental work is carried out. Additionally, the oxygen mass transfer phenomenon has been found equally important as the others. In this way, experimental data that can offer more information about such phenomena can be highly relevant to estimate the parameter that were fixed in this case study (e.g. \( C_{O_2}^{sat} \) the saturation concentration of oxygen).

The parameters \( V_{lacc}^{max}, K_La \) and \( K_{hyd} \) were identified as the main factors responsible for the variance in the outputs on lactose, oxygen and LBA, respectively.

The uncertainty propagation of lactose and oxygen was small which indicates that even if the factors \( V_{lacc}^{max} \) and \( K_La \) are significant, the model will have good predictive capability. Furthermore, these variables are reliable to be measured and based further calculations on them.

Furthermore, \( K_{hyd} \) was the main factor that produced the high range on the output uncertainty of LBA which indicates that a better experimental design must be formulated for the estimation of this parameter.
7.7 List of symbols and nomenclature

Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH</td>
<td>Cellobiose dehydrogenase</td>
</tr>
<tr>
<td>ABTS</td>
<td>2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt</td>
</tr>
<tr>
<td>ABTS+</td>
<td>2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt cation radical</td>
</tr>
<tr>
<td>LHS</td>
<td>Latin hypercube sampling</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
</tbody>
</table>

Symbol

- A: Substrate
- B: Substrate
- P: Product
- Q: Product
- n: Number of input factors
- p: Number of levels
- r: Number of samples
- $F_i$: Distribution of elementary effects

Nomenclature

- $V_{\text{max}}$: Maximum initial velocity of an enzyme ($mM h^{-1}$)
- $K_M$: Michaelis-Menten constant ($mM$)
- $K_{l,a}$: Volumetric mass transfer coefficient ($h^{-1}$)
- $K_{\text{hyd}}$: Hydrolysis constant ($h^{-1}$)
- $C^0$: Initial concentration of any species ($mM$)
- $C$: Concentration of any species ($mM$)
- r: Reaction rate ($mM h^{-1}$)

Greek letters

- $\Delta$: Optimal perturbation factor

Subscripts
7.7 List of symbols and nomenclature

\[\begin{align*}
lact & \quad \text{Lactose} \\
LBL & \quad \text{Lactobiolactone} \\
LBA & \quad \text{Lactobionic acid} \\
O_2 & \quad \text{Oxygen} \\
ABTS & \quad \text{Reduced redox intermediate} \\
ABTS^+ & \quad \text{Oxidized redox intermediate} \\
omt & \quad \text{Oxygen mass transfer}
\end{align*}\]

**Superscripts**

\[\begin{align*}
CDH & \quad \text{Cellobiose dehydrogenase} \\
lacc & \quad \text{Laccase} \\
ABTS^+ & \quad \text{Oxidized redox mediator} \\
ABTS & \quad \text{Reduced redox mediator} \\
sat & \quad \text{Saturation}
\end{align*}\]
Case Study 2: Bi-enzyme production of N-acetyl-D-neuraminic acid

8.1 General description

The intermediate N-acetyl-D-neuraminic acid (Neu5Ac) is used in pharmaceutical industry for the production of the valuable product zanamivir which is an anti-viral drug used to prevent and treat influenza (also marketed under the trade name Relenza) [123 125]. A further motivation is that one of the production methods is based on the combination of two enzymes in one reactor which is then a clear example of a multi-enzyme in-pot process. Consequently, it is also interesting to analyze the reliability of the existing mathematical model of this process which can be used for further model-based applications e.g. process optimization and control.

The synthesis of Neu5Ac has been achieved by applying different methods, basically in two sequential reactions. The primary method applied industrially is the enzymatic process in which the action of two enzymes is exploited in a single reactor [113]. Another option for the production of Neu5Ac is the implementation of a chemoenzymatic process. In that case, the first reaction is
an alkaline-catalyzed epimerization [5, 126, 127]. However the use of metal hydroxides hinders the integration of both reaction steps in a single reactor. More recent approaches have been reported in the use of a novel whole cell system which was developed for the production of NeuAc from N-acetyl-D-glucosamine (GlcNAc) and pyruvate by recombinant whole cells [128, 129].

8.2 Step 1. Modeling objective

The objective of this case study is to analyze the influence in the dynamic behavior of the outputs (concentrations of components) using simplified mathematical model due to the elimination of parameters from the original model based on sensitivity analysis.

8.3 Step 2. Information

For this case study, the mathematical model is available in the scientific literature including the kinetic parameters of the multi-enzyme process. Furthermore, experimental process conditions are also reported [113]. In this way, the information is organized and described here, according the modeling methodology, as follows:

8.3.1 Reaction considerations

- Figure 8.1 shows the general scheme for the production of Neu5Ac. This illustrates the full cascade reactions involved in the process. The first enzyme (epimerase, E.C. 5.1.3.8) uses N-acetyl-D-glucosamine (GlcNAc) as a substrate to be transformed into N-acetyl-D-mannosamine (ManNAc). For the subsequent reaction, pyruvate is required as second substrate. In this manner, the transformation of ManNAc to the desired product (Neu5Ac) is induced by using a second enzyme (aldolase, E.C. 4.1.3.3). However the second reaction is limited by the thermodynamic equilibrium, and then an excess of pyruvate is required to shift the reaction to the product side. The problem with that approach is faced in the downstream process in which the separation of Neu5Ac from the pyruvate becomes difficult because both compounds have similar properties and are highly water soluble [113, 130, 131].
Figure 8.1: Synthesis of Neu5Ac from GlcNAc in two cascade reactions; A: N-acetyl-D-glucosamine (GlcNAc); B: N-acetyl-D-manosamine (ManNAc); C: pyruvate (Pyr); D: N-acetyl-D-neuraminic acid (Neu5Ac); Enzymes: N-acylgulcosamine-2-epimerase, N-acetylneuraminic acid aldolase
Table 8.1: Interaction matrix for the multienzymatic production of N-acetyl-D-neuraminic acid

<table>
<thead>
<tr>
<th>Compound</th>
<th>Enzyme</th>
<th>Epimerase</th>
<th>Aldolase</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc</td>
<td></td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>ManNAc</td>
<td></td>
<td>P</td>
<td>S-I</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>I</td>
<td>S-I</td>
<td>P</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>I</td>
<td>P</td>
<td></td>
</tr>
</tbody>
</table>

(S) substrate, (P) product, (X) no interaction
(I) inhibition, (S-I) substrate inhibition

- **Reaction mechanism for the single enzymes:** Reported kinetic parameter values were independently estimated for both reactions by initial rate measurements and progress curve analysis. Furthermore, most parameter values were reported with a standard deviation between 10 and 15% due to the analytical method for the measurements i.e. HPLC.

- **Epimerase:** this enzyme is reported to follow a reversible Michaelis-Menten mechanism (see Appendix A). Furthermore, inhibition terms by the compounds in the second reaction were also included in the model i.e. pyruvate and Neu5Ac. The inhibition was described as competitive in that case. Table 8.2 shows the reported parameter values for epimerase.

- **Aldolase:** this enzyme is reported to follow an ordered bi-unim mechanism. Furthermore, the equilibrium constant was calculated using the Haldane relationship. Table 8.3 shows the kinetic parameter values for aldolase in this process.

- **Interaction matrix:** Taking into account the multi-enzyme system, the interactions between compounds and enzymes are shown in Table 8.1. In this case, the most relevant interactions are the inhibitory effects of the epimerase at high concentration of pyruvate and Neu5Ac. Furthermore, the aldolase shows substrate inhibition from both ManNAc and pyruvate.

### 8.3.2 Process considerations

The reactions were carried out in batch operation during the first 990 minutes. After this time GlcNAc and pyruvate were fed simultaneously, and then the fed-
8.4 Step 3. Formulation

8.4.1 Model assumptions

According to the information compiled in the step 2 and the literature [113], the following assumptions can be extracted for the mathematical model.

- Kinetic parameter values are taken from literature and then they are assumed as the true values for further analyses.
- It is assumed that only components of the second reaction affect the first enzyme and not the other way around. This assumption means that inhibition of pyruvate and Neu5Ac is taken into account in the reaction rate of epimerase.
- pH and temperature are maintained constant during the operation.
- Since the type of reactor is not specified in the literature, a stirred tank reactor is assumed for this case study.

8.4.2 Boundaries and balance volume

Figure 8.2 shows a scheme of the type of reactor and the balance volumes that are assumed for the reactor and both enzymes involved in the process.

8.4.3 Conservation equations

The formulation of the model is given by the following set of differential equations. It includes the component balances for the concentration of substrates (GluNAc and Pyruvate), intermediate product (ManNAc), and desired product (Neu5Ac). Furthermore, the volume change model is included in the model to take into account the dilution rate due to the fed-batch operation.

batch operation starts. pH was maintained constant during the process at 7.5 and temperature at 25°C. The process was carried out for 75 hours. Table 8.4 shows the initial operating conditions which are used to perform the simulations of the mathematical model and further model analyses [113].
Figure 8.2: Reactor scheme and balance volume of the process. (a) balance volume for the reactor (fed-batch STR) and (b) balance volume for the enzymes (epimerase-aldolase)

\[
\frac{dC_{\text{GlcNAc}}}{dt} = \left( \frac{C_{\text{fd}}^{\text{GlcNAc}} - C_{\text{GlcNAc}}}{V} \right) + r_{\text{epi}} - r_{\text{ald}} \quad (8.1)
\]

\[
\frac{dC_{\text{ManNAc}}}{dt} = \frac{C_{\text{ManNAc}} \cdot F}{V} - r_{\text{ald}} \quad (8.2)
\]

\[
\frac{dC_{\text{Pyr}}}{dt} = \left( \frac{C_{\text{fd}}^{\text{Pyr}} - C_{\text{Pyr}}}{V} \right) - r_{\text{ald}} \quad (8.3)
\]

\[
\frac{dC_{\text{Neu5Ac}}}{dt} = \frac{C_{\text{Neu5Ac}} \cdot F}{V} - r_{\text{ald}} \quad (8.4)
\]

\[
\frac{dV}{dt} = F \quad (8.5)
\]

8.4.4 Constitutive equations

Based on the reverse Michaelis-Menten mechanism including inhibitions of components from the second reaction, the reaction for epimerase may be derived as:

\[
r_{\text{epi}} = \frac{E_{\text{epi}} \left( \frac{A_{\text{GlcNAc}} C_{\text{GlcNAc}}}{K_{\text{GlcNAc}}} - \frac{A_{\text{ManNAc}} C_{\text{ManNAc}}}{K_{\text{ManNAc}}} \right) C_{\text{GlcNAc}}}{1 + \frac{C_{\text{GlcNAc}}}{K_{\text{Gld}}^{\text{GlcNAc}}} + \frac{C_{\text{ManNAc}}}{K_{\text{Gld}}^{\text{ManNAc}}} + \frac{C_{\text{Pyr}}}{K_{\text{Gld}}^{\text{Pyr}}} + \frac{C_{\text{Neu5Ac}}}{K_{\text{i}}^{\text{Neu5Ac}}}} \quad (8.6)
\]

Based on an ordered bi-uni mechanism, the reaction rate for aldolase is derived as:
8.5 Step 4. Calculations

\[
    r_{\text{ald}} = \frac{E_{\text{ald}} \left( \frac{A_f \cdot C_{\text{ManNAc}} \cdot C_{Pyr}}{K_{v_{\text{ald}}} \cdot K_{M_{\text{ManNAc}}}^M} - \frac{A_r \cdot C_{\text{ManNAc}}}{K_{v_{\text{ald}}} \cdot K_{M_{\text{ManNAc}}}^M} \right)}{1 + \frac{C_{Pyr}}{K_{v_{\text{ald}}}} + \frac{K_{v_{\text{ald}}} \cdot C_{\text{ManNAc}}}{K_{M_{\text{ManNAc}}}^M} + \frac{C_{\text{ManNAc}} \cdot C_{Pyr}}{K_{M_{\text{ManNAc}}}^M} + \ldots}
\]

\[\text{(8.7)}\]

In all cases, \(K_M\) is the Michaelis-Menten constants and \(K_i\) is the inhibition constant.

8.4.5 Conditions

Both reaction rates include reversibility, and then the equilibrium constant for each reaction can be described by the Haldane relationship, thus

For epimerase:

\[
    K^\text{epi}_{eq} = \frac{A^\text{GlcNAc}_u \cdot K^\text{ManNAc}_M}{A^\text{ManNAc}_M \cdot K^\text{GlcNAc}_u}
\]

\[\text{(8.8)}\]

For aldolase:

\[
    K^\text{ald}_{eq} = \frac{A^\text{f} \cdot K^\text{Neu5Ac}_M}{A^\text{f}_r \cdot K^\text{Pyruv}_r \cdot K^\text{ManNAc}_M}
\]

\[\text{(8.9)}\]

8.5 Step 4. Calculations

The information collected in the Steps 2 and 3 is used to implement and analyze the reliability of the model according the modeling objective. As mentioned in Step 2, kinetic parameter values are reported in the scientific literature for each enzyme mechanism [113]. These values are summarized in Tables 8.2 for epimerase and 8.3 for aldolase.

Furthermore, the initial conditions for the process are also reported and shown in Table 8.4 [113].
Table 8.2: Kinetic parameter values for epimerase taken from the literature [113]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{GlcNac}^v$</td>
<td>$(5.31 \pm 0.74) \times 10^{-7}$ U/l</td>
</tr>
<tr>
<td>$K_{GlcNac}^v$</td>
<td>$(1.76 \pm 0.26) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$A_{ManNac}^v$</td>
<td>$(1.16 \pm 0.33) \times 10^{-5}$ U/l</td>
</tr>
<tr>
<td>$K_{ManNac,e}^v$</td>
<td>$(9.93 \pm 1.14) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$K_{Pyr,e}^v$</td>
<td>$0.146 \pm 0.019$ M</td>
</tr>
<tr>
<td>$K_{Neu5Ac,e}^v$</td>
<td>$0.719 \pm 0.158$ M</td>
</tr>
</tbody>
</table>

Table 8.3: Kinetic parameter values for aldolase taken from the literature [113]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_f^v$</td>
<td>$(7.51 \pm 1.11) \times 10^{-8}$ U/l</td>
</tr>
<tr>
<td>$K_{Pyr}^v$</td>
<td>$(8.49 \pm 1.06) \times 10^{-3}$ M</td>
</tr>
<tr>
<td>$K_{ManNac}^v$</td>
<td>$(1.31 \pm 0.15) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$A_r^v$</td>
<td>$(1.05 \pm 0.05) \times 10^{-7}$ U/l</td>
</tr>
<tr>
<td>$K_{Neu5Ac}^v$</td>
<td>$(4.26 \pm 0.8) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$K_{Pyr}^v$</td>
<td>$(9.41 \pm 0.91) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$K_{ManNac}^v$</td>
<td>$(1.19 \pm 0.91) \times 10^{-2}$ M</td>
</tr>
<tr>
<td>$K_{eq,exp}$</td>
<td>$28.7 \pm 10.4$ M</td>
</tr>
<tr>
<td>$K_{eq,th}$</td>
<td>$27.4 \pm 8.1$ M</td>
</tr>
</tbody>
</table>
8.5 Step 4. Calculations

Table 8.4: Initial operating and feeding conditions for the multi-enzyme process [113]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epimerase activity</td>
<td>150 U</td>
</tr>
<tr>
<td>Aldolase activity</td>
<td>2400 U</td>
</tr>
<tr>
<td>$V^0$</td>
<td>0.1 L</td>
</tr>
<tr>
<td>$C_{GlcNAc}$</td>
<td>0.3 M</td>
</tr>
<tr>
<td>$C_{PyrAc}$</td>
<td>0.38 M</td>
</tr>
<tr>
<td>$C_{GlcNAc}^{fd}$</td>
<td>1 M</td>
</tr>
<tr>
<td>$C_{PyrAc}^{fd}$</td>
<td>0.9 M</td>
</tr>
<tr>
<td>$F$</td>
<td>$2 \times 10^{-5}$ l/min</td>
</tr>
</tbody>
</table>

For this case study, identifiability, parameter estimation and uncertainty analysis were not performed since these analyses are unnecessary to achieve for the model objective. Furthermore, progress curves of the process are not available. Therefore, sensitivity analysis is performed in order to identify insignificant kinetic parameters that are possible to eliminate from the mathematical model in order to make it simpler and more reliable.

8.5.1 Model implementation and dynamic simulation

In order to performed the dynamic simulation of the process, the set of differential equations described in Step 3 are solved. The input to the solver are the process model, the initial conditions, the simulation time and the kinetic parameters values (see Figure B.2). The outputs of this process model are the dynamic behavior of all component concentrations, as shown in Figure 8.3.

Figure 8.3 shows the dynamic behavior of each component concentration in the reactor during the process. The point where the feeding is started is clearly seen in the process behavior at $t = 990\text{min}$. In this manner, the concentration of the desired product is increased due to the supply of substrate in time. In the same way, the concentration of the intermediate product (ManNAc) is indeed kept at a low concentration as desired during the whole process operation.
8.5.2 Sensitivity analysis

The Morris Screening method was performed in order to evaluate parameter reliability of the reported model structure (see Section 4.5 for a description of the method). The total number of factors to be evaluated was $n = 13$ where 6 parameters were from the epimerase reaction and 7 from the aldolase reaction. Since confidence intervals were available for each parameter, they were used to limit the sampling space in the Morris sampling. In order to generate the Morris sampling, the number of levels is selected as $p = 6$, and the number of repetitions for the $EE_i$ is $r = 30$. Using Equation 4.13, the optimal perturbation factor is calculated as $\Delta = 0.6$. In order to analyze the elementary effect for each model output, the distribution is obtained after running a total number of model evaluations of $me = 420$ which was calculated using Equation 4.14.

Furthermore, the standard deviation $\sigma$ and the mean value $\mu$ of each elementary effect were calculated and plotted (Figure 8.4). In the same manner, the absolute mean value was calculated in order to rank the most significant parameters in the mathematical model (see Table 8.5).

The analysis indicates that the same 4 parameters in the model have a low effect on all the model outputs at the evaluated conditions. The parameters are
Table 8.5: Ranking of significant parameters for each model output based on the absolute mean value of the elementary effects

<table>
<thead>
<tr>
<th>Rank</th>
<th>GlcNAc</th>
<th>ManNAc</th>
<th>Pyruvate</th>
<th>Neu5Ac</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta$</td>
<td>$\mu$</td>
<td>$\theta$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>1</td>
<td>$A_v^{\text{ManNAc}}$</td>
<td>0.5772</td>
<td>$K_M^{\text{Neu5Ac}}$</td>
<td>0.5554</td>
</tr>
<tr>
<td>2</td>
<td>$K_M^{\text{Neu5Ac}}$</td>
<td>0.3391</td>
<td>$A_i^{\text{Pyr}}$</td>
<td>0.4026</td>
</tr>
<tr>
<td>3</td>
<td>$A_v^{\text{GlcNAc}}$</td>
<td>0.3361</td>
<td>$K_i^{\text{Pyr}}$</td>
<td>0.3818</td>
</tr>
<tr>
<td>4</td>
<td>$A_i^{\text{Pyr}}$</td>
<td>0.3153</td>
<td>$K_M^{\text{ManNAc}}$</td>
<td>0.3501</td>
</tr>
<tr>
<td>5</td>
<td>$K_i^{\text{GlcNAc}}$</td>
<td>0.3136</td>
<td>$A_v^{\text{ManNAc}}$</td>
<td>0.3255</td>
</tr>
<tr>
<td>6</td>
<td>$K_i^{\text{Pyr}}$</td>
<td>0.2425</td>
<td>$K_M^{\text{GlcNAc}}$</td>
<td>0.1493</td>
</tr>
<tr>
<td>7</td>
<td>$K_M^{\text{ManNAc,e}}$</td>
<td>0.2272</td>
<td>$A_v^{\text{GlcNAc}}$</td>
<td>0.1215</td>
</tr>
<tr>
<td>8</td>
<td>$K_M^{\text{ManNAc}}$</td>
<td>0.2235</td>
<td>$K_M^{\text{ManNAc,e}}$</td>
<td>0.1311</td>
</tr>
<tr>
<td>9</td>
<td>$A_v^{\text{GlcNAc}}$</td>
<td>0.0831</td>
<td>$K_M^{\text{ManNAc}}$</td>
<td>0.1001</td>
</tr>
<tr>
<td>10</td>
<td>$K_i^{\text{ManNAc}}$</td>
<td>0.0834</td>
<td>$K_M^{\text{ManNAc}}$</td>
<td>0.0346</td>
</tr>
<tr>
<td>11</td>
<td>$K_i^{\text{Pyr,e}}$</td>
<td>0.0148</td>
<td>$K_i^{\text{Pyr,e}}$</td>
<td>0.0331</td>
</tr>
<tr>
<td>12</td>
<td>$K_M^{\text{Pyr}}$</td>
<td>0.0021</td>
<td>$K_M^{\text{Pyr}}$</td>
<td>0.0031</td>
</tr>
<tr>
<td>13</td>
<td>$K_i^{\text{Neu5Ac,e}}$</td>
<td>0.0005</td>
<td>$K_i^{\text{Neu5Ac,e}}$</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
specifically $K_{\text{ManNAc}}$, $K_{\text{Pyr}}, e$, $K_{\text{Pyr}}$ and $K_{\text{Neu5Ac}, e}$ (see last 4 parameters in Table 8.5). Hence these parameters do not contribute significantly to the dynamics of the model which therefore can be simplified as a result. Consequently, the model structure needs to be modified and it is then necessary to return to Step 3 in the methodology, as shown in Figure 3.1.

8.6 Step 3. (Re) Formulation

According to the results obtained in Step 4, the model structure is modified by removing the parameters. In this case, structural simplification of the model is done in the reaction rate of each enzyme, specifically Equations 8.6 and 8.7. In order to identify each model modification, the model structures are named as presented in Table 8.6.

Once the modified models are available, they are ready to be implemented in order to numerically and graphically analyze the effects that the simplifications have in comparison with the original model.
8.7 Step 4. Calculations

Table 8.6: Parameters that are eliminated for each simplified model structure

<table>
<thead>
<tr>
<th>Model</th>
<th>Eliminated parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>$K_{Neu5Ac,e}$</td>
</tr>
<tr>
<td>Model 2</td>
<td>$K_{Neu5Ac,e}$, $K_{Pyr,e}$</td>
</tr>
<tr>
<td>Model 3</td>
<td>$K_{Neu5Ac,e}$, $K_{Pyr,e}$, $K_{ManNAc}$</td>
</tr>
<tr>
<td>Model 4</td>
<td>$K_{Neu5Ac,e}$, $K_{Pyr,e}$, $K_{ManNAc}$, $K_{Pyr}$</td>
</tr>
</tbody>
</table>

Table 8.7: Output model comparison between original model and the simplified models (see Table 8.6)

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of eliminated parameters</th>
<th>Deviation(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1</td>
<td>0.7800</td>
</tr>
<tr>
<td>Model 2</td>
<td>2</td>
<td>3.5341</td>
</tr>
<tr>
<td>Model 3</td>
<td>3</td>
<td>3.5832</td>
</tr>
<tr>
<td>Model 4</td>
<td>4</td>
<td>4.3600</td>
</tr>
</tbody>
</table>

8.7 Step 4. Calculations

In this step, graphical and numerical comparisons between the original model and the simplified models are performed in order to quantify the deviation from the original model. Numerical results are shown in Table 8.7.

The deviation is calculated as the average value of the errors between the modified and the original models, presented as a percentage. Table 8.7 shows that even though maximum four parameters are removed from the original model, the simplified models do not show differences higher than 5%. It demonstrates that the insignificant parameters in this case can be removed from the model without producing big differences in the model prediction accuracy.

For the graphical comparison, the concentration of the desired product Neu5Ac is taken as model output. Figure 8.5 shows the plots obtained from the dynamic simulation of each modified model.

Figure 8.5 shows that all simplified models can well follow the dynamics of the original process. Furthermore, the deviation between different model prediction...
Figure 8.5: Comparison between dynamic simulations of the modified models

is again really small. At the right-bottom of the same figure, an extension of the marked area is amplified in order to show the difference between the model simulation of the simplified models.

8.8 Step 5. Experimentation

For the purpose of this case study, experiments were not performed. However, results obtained from sensitivity analysis can be used as a basis to formulate an optimal experimental design. In that manner, relevant experiments can be carried out focused on the identification of significant parameters. It will open the option to re-estimate the parameters in order to validate the simplified model.

8.9 Conclusions

A description for model formulation in the application of multi-enzymatic processes was shown. Since complex models are expected due to the interaction
of compounds and enzymes, this methodology could be used to simplify the
structure of the models in a systematic manner based on sensitivity analysis,
as exemplified with the Neu5Ac case study. It was also demonstrated that
the original model was overparametrized and a feasible simplification could be
performed without losing reliability of the model. It was also found that the
simplified models (see Table 8.7) still showed a good performance with a devi-
ation of less than 5% from the original model. Actually, it was demonstrated
based on sensitivity analysis that the inhibitory effects in the first reaction due
to compounds of the second reaction are insignificant, specifically $K_{Neu5Ac,e}$
and $K_{Pyr,e}$, and therefore that this assumption can be removed when working
with a simplified model. However for a validation of the model, parameters
need to be re-calibrated using experimental data and the new model. In fact,
optimal experimental design can then be applied in order to get experimental
data that are as information-rich as possible to focus laboratory effort on the
most significant parameters in the model as indicated in the sensitivity analysis.

8.10 List of symbols and nomenclature

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-D-glucosamine</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetyl-D-manosamine</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>N-acetyl-D-neuraminic acid</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>STR</td>
<td>Stirred tank reactor</td>
</tr>
</tbody>
</table>

**Nomenclature**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_v$</td>
<td>Enzyme activity ($U/l$)</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis-Menten constant ($M$)</td>
</tr>
<tr>
<td>$K_i$</td>
<td>Inhibition constant ($M$)</td>
</tr>
<tr>
<td>$K_{eq}$</td>
<td>Equilibrium constant ($M$)</td>
</tr>
<tr>
<td>$C^0$</td>
<td>Initial concentration ($M$)</td>
</tr>
<tr>
<td>$C$</td>
<td>Concentration ($M$)</td>
</tr>
<tr>
<td>$F$</td>
<td>Input flow rate ($l/min$)</td>
</tr>
</tbody>
</table>

**Subscripts**
### Case Study 2: Bi-enzyme production of Neu5Ac

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-D-glucosamine</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetyl-D-mannosamine</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>N-acetyl-D-neuraminic acid</td>
</tr>
<tr>
<td>Pyr</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>theo</td>
<td>Theoretical calculation</td>
</tr>
<tr>
<td>exp</td>
<td>Experimental data</td>
</tr>
<tr>
<td>epi</td>
<td>Epimerase</td>
</tr>
<tr>
<td>ald</td>
<td>Aldolase</td>
</tr>
<tr>
<td>A</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>B</td>
<td>ManNAc</td>
</tr>
<tr>
<td>C</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>D</td>
<td>Neu5Ac</td>
</tr>
</tbody>
</table>

**Superscripts**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-D-glucosamine</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetyl-D-mannosamine</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>N-acetyl-D-neuraminic acid</td>
</tr>
<tr>
<td>Pyr</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>f</td>
<td>Forward direction</td>
</tr>
<tr>
<td>r</td>
<td>Reverse direction</td>
</tr>
<tr>
<td>ManNAc,e</td>
<td>N-acetyl-D-mannosamine</td>
</tr>
<tr>
<td>Neu5Ac,e</td>
<td>N-acetyl-D-neuraminic acid</td>
</tr>
<tr>
<td>fd</td>
<td>Feeding</td>
</tr>
</tbody>
</table>
Chapter 9

Case Study 3. Tri-enzyme production of 1-phenylethylamine

9.1 General description

Optically active chiral amines are used as active compounds in several drug products due to their pharmacological properties. However, the enzymatic synthesis of amines is highly limited by the equilibrium constant which generally favors the reverse reaction. On the other hand, transaminases, which are used as the main enzymes in this transformation, typically show substrate and product inhibition [114]. Hence, one strategy that may help to reduce these reaction limitations is to combine the action of three enzymes in a single reactor [132]. Experimental evidence of this implementation has been provided in literature which promote a better understanding of the nature of each reaction/enzyme involved in the process.
9.2 Step 1: Model objective

Build a mathematical model of the process in order to analyze different process configurations to maximize the productivity of the process. The promising configuration is further analyzed in order to identify the parameters that are still significant in the model.

9.3 Step 2: Information

For this case study, the model can be built with the information that is found in the literature. In this way, the model can also be used to analyze the dynamic behavior of the process. Consequently promising process configurations can be further analyzed and subsequently implemented experimentally. In the scientific literature, the enzyme mechanisms were identified for all enzymes involved in the process. In the same manner, some kinetic parameters were found for each enzyme. In these cases, the kinetic studies have been performed for the individual enzymes where the conditions for the enzyme characterization were different for each enzyme [114-116]. The multi-enzyme process also has been tested at laboratory scale. In this case, the studies have been more focused on the productivity and compatibility of the enzymes [117, 132, 133]. This means that neither kinetic studies of the full system have been fully developed nor a mathematical model of this multi-enzyme process. Progress curves of this multi-enzyme process are not available in the literature.

The available information is organized and described here, according the modeling methodology, as follows:

9.3.1 Reaction considerations

The reaction mechanisms and kinetic parameters for the single enzymes were obtained from the literature:

- The optically pure chiral product 1-phenylethylamine is synthesized from acetophenone by using ω-transaminase (E.C.2.6.1.X) as biocatalyst. Furthermore L-alanine is required as amino donor which is, in the same reaction, transformed to pyruvate. However, the transaminase reaction is limited by the equilibrium constant that generally favors the ketone (acetophenone), and transaminases typically suffer significant inhibition by
9.3 Step 2: Information

Figure 9.1: Reaction scheme for three-enzyme process in the production of L-phenylethylamine (PEA). (a) Acetophenone, (b) L-phenylethylamine, (c) L-alanine, (d) Pyruvate, (e) Lactate, (g) Glucose and (h) Gluconolactone

The ketone and keto acid by-product [114, 134, 135]. In order to shift the equilibrium to the product side, pyruvate is removed by a second enzyme, lactate dehydrogenase (LDH, EC 1.1.1.27). In that manner, pyruvate is converted to lactate. However, the LDH reaction is NADH-dependent. Consequently, a third enzyme, glucose dehydrogenase (GDH, EC 1.1.1.47), is also added to the system in order to promote the NADH regeneration. In that manner, the enzyme GDH catalyzes the dehydrogenation of glucose to gluconolactone which spontaneously hydrolyses to gluconic acid [132]. Figure 9.1 illustrates the reaction network.

- ω-Transaminase: this enzyme follows the substituted mechanism for binding substrates and releasing products. From literature [114], the kinetic parameters for this enzyme have been obtained including substrate inhibition (i.e., acetophenone and L-alanine). Parameter values are shown in Table 9.3.

- LDH and GDH: these enzymes follow the compulsory order mechanism for binding substrates and releasing products (see Figure A.6). From the literature, the kinetic parameters for LDH have been obtained for the initial rate reaction [115] and their values are shown in Table 9.4. In the same way, the kinetic parameters for GDH have been obtained for the initial rate reaction [116], and their values are shown in Table 9.5.
Case Study 3. Tri-enzyme production of 1-phenylethylamine

Table 9.1: Interaction matrix for the multi-enzyme production of 1-phenylethylamine

<table>
<thead>
<tr>
<th>Compound/Enzyme</th>
<th>Transaminase</th>
<th>LDH</th>
<th>GDH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetophenone</td>
<td>S - I</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1-phenylethylamine</td>
<td>P - I</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>L-alanine</td>
<td>S</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>P</td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>NADH</td>
<td>X</td>
<td>S</td>
<td>P</td>
</tr>
<tr>
<td>Lactate</td>
<td>X</td>
<td>P</td>
<td>X</td>
</tr>
<tr>
<td>NAD+</td>
<td>X</td>
<td>P</td>
<td>S</td>
</tr>
<tr>
<td>Glucose</td>
<td>X</td>
<td>X</td>
<td>S</td>
</tr>
<tr>
<td>Gluconolactone</td>
<td>X</td>
<td>X</td>
<td>P</td>
</tr>
</tbody>
</table>

(S) substrate, (X) no interaction,
(P) product, (S-I) substrate inhibition

- Interaction matrix: Taking into account the whole multi-enzyme process, Table 9.1 shows the interactions between compounds and enzymes that are expected for this multi-enzyme process according to the literature.

9.3.2 Process considerations

Some cases have been reported in the scientific literature for the full system. Such information is here collected in terms of the process conditions [114, 132, 133].

Process conditions:

- pH: 7
- Temperature: 30°C
- Operation mode: batch-fed-batch
- Process time: 48h

Initial operating conditions:
9.4 Step 3: Formulation

The idea here is to analyze different scenarios of the process in order to select the configuration where the productivity of PEA is improved. Consequently four scenarios are analyzed, thus:

- **Scenario 1**: The mixture of the three enzymes is considered in batch operation.

- **Scenario 2**: The mixture of the three enzymes is considered in batch operation. Furthermore, *in-situ* PEA removal is evaluated in this case.

- **Scenario 3**: The mixture of the three enzymes is considered in fed-batch operation. In this case, acetophenone and L-alanine are constantly fed to the process.

- **Scenario 4**: The mixture of the three enzymes is considered in fed-batch operation, similar to Scenario 3. Furthermore, *in-situ* PEA removal is also evaluated.

As mentioned, the combination of three enzyme has been used as a solution for moving the equilibrium of the transamination to the amine side. However, the main product concentration is still too low to be a promising implementation at industrial level. Consequently, the other scenarios are included in order to find a configuration that can achieve better productivity of the process. Furthermore, the enzyme activity of the transaminase is increased virtually to achieve relevant conversions. In this case, *in-situ* product removal (ISPR) is included in the evaluation since product inhibition can be avoided and the unfavorable equilibrium can be enhanced (see Section 5.1.3) [98, 99]. Another limitation of this process are the substrate inhibitions (i.e. acetophenone and L-alanine) that are reported in the literature [114]. Consequently, fed-batch operation mode is included in the analysis since the controlled dosing of substrates avoids the enzyme intoxication in the process.
9.4.1 Model assumptions

- Kinetic parameters are taken from literature and then they are assumed correct for the preliminary analysis of the process.

- The multi-enzyme model is built by the combination of the single enzyme kinetics, and then only the interactions shown in Table 9.1 are taken into account.

- For those scenarios with in-situ product removal, the removal of PEA is assumed to be as fast as it is produced in the process. Furthermore polymeric resins are assumed as a reliable ISPR technique [99, 117].

- Changes in the reactor volume are taken into account only when substrates are supplied (fed-batch cases). This means that the volume of the PEA removed is neglected.

- pH and temperature are assumed to be constant in the process.

- The transaminase activity is improved virtually 200 times in this case study in order to achieved relevant amount of PEA for an industrial level application.

9.4.2 Boundaries and balance volume

Figure 9.2 show the schemes of the type of reactor and the balance volumes that are assumed for the reactor and the three enzymes involved in the process in the four scenarios.

9.4.3 Conservation equations

The stoichiometry of this multi-enzyme process is illustrated in Figure 9.1. Based on that, the set of differential equations is formulated for each scenario, including component balances for concentrations of substrates (acetophenone, L-alanine, glucose), co-enzyme (NADH), intermediate product (pyruvate), and products (PEA, lactone, gluconolactone). Thus, each model for the system consists of nine (9) differential equations when batch operation is to be simulated and ten (10) differential equations when fed-batch operation is to be simulated because of the additional total volume balance in the reactor.
Table 9.2: Mass balances for all variables in the process for the different process configurations

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{dC_{Ace}}{dt} ) = ( r_{TA} )</td>
<td>( r_{TA} )</td>
<td>( \frac{(C_{Ace}^{I_i} - Ace)F}{V} + r_{TA} )</td>
<td>( \frac{(C_{Ace}^{I_i} - Ace)F}{V} + r_{TA} )</td>
</tr>
<tr>
<td>( \frac{dC_{Ala}}{dt} ) = ( r_{TA} )</td>
<td>( r_{TA} )</td>
<td>( \frac{(C_{Ala}^{I_i} - Ala)F}{V} + r_{TA} )</td>
<td>( \frac{(C_{Ala}^{I_i} - Ala)F}{V} + r_{TA} )</td>
</tr>
<tr>
<td>( \frac{dC_{PEA}}{dt} ) = ( -r_{TA} - r_{LDH} - 0.9r_{TA} )</td>
<td>( -r_{TA} - r_{LDH} - 0.9r_{TA} )</td>
<td>( -r_{TA} - r_{LDH} - 0.9r_{TA} )</td>
<td>( -r_{TA} - r_{LDH} - 0.9r_{TA} )</td>
</tr>
<tr>
<td>( \frac{dC_{Glu}}{dt} ) = ( r_{GDH} - r_{LDH} )</td>
<td>( r_{GDH} - r_{LDH} )</td>
<td>( r_{GDH} - r_{LDH} )</td>
<td>( r_{GDH} - r_{LDH} )</td>
</tr>
<tr>
<td>( \frac{dV}{dt} ) = ( F )</td>
<td>( F )</td>
<td>( F )</td>
<td>( F )</td>
</tr>
</tbody>
</table>
9.4.4 Constitutive equations

Based on the substituted-enzyme mechanism including substrate inhibitions, the reaction rate for transaminase may be derived as:

\[
r_{TAm} = \frac{VfCA}{K_f K_m} \left( 1 + \frac{CA}{K_s} \right) + \frac{V^*CP}{K_i K_m} \left( 1 + \frac{CA}{K_s} \right) + \frac{CQ}{K_q K_m} + \cdots
\]

\[
\cdots + \frac{CA}{K_f K_m} + \frac{CP}{K_i K_m} \left( 1 + \frac{CA}{K_s} \right) + \frac{CQ}{K_q K_m} + \cdots
\]

(9.1)

Where \( A \) is related to L-alanine, \( B \) is related to acetophenone, \( P \) is related to pyruvate and \( Q \) is related to PEA.

Based on the compulsory order mechanism, the initial reaction rates for LDH and GDH are derived as:

\[
r_{LDH} = r_{GDH} = \frac{VCAC}{K_m C_A + K_m C_B + CA C_B}
\]

(9.2)

For the LDH reaction, \( A \) is related to NADH, \( B \) is related to pyruvate, \( P \) is related to lactate and \( Q \) is related to NAD+.
9.5 Step 4: Calculations

The information collected in Steps 2 and 3 is used to implement and analyze the reliability of the model according to the modeling objective. As mentioned in Step 2, kinetic parameter values are reported in the scientific literature for each enzyme mechanism [114 116]. These values are summarized in Tables 9.3 for the transaminase, 9.4 for the LDH, and 9.5 for the GDH.

Identifiability, parameter estimation and uncertainty analysis are not performed for this case study experimental data is not available. Therefore, sensitivity analysis is performed for the most promising process configuration in order to identify relevant parameters that need further experimental research in order to validate the suggested model.
Table 9.4: Kinetic parameter values for LDH from literature [115]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{LDH}$</td>
<td>1.5445 mM/min</td>
</tr>
<tr>
<td>$K_{NADH}^M$</td>
<td>$2.4991 \times 10^{-1}$ mM</td>
</tr>
<tr>
<td>$K_{Pyr}^M$</td>
<td>2.0303 mM</td>
</tr>
<tr>
<td>$K_i^{NADH}$</td>
<td>$2.2288 \times 10^{-3}$ mM</td>
</tr>
<tr>
<td>T</td>
<td>37 °C</td>
</tr>
<tr>
<td>pH</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 9.5: Kinetic parameter values for GDH from literature [116]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{GDH}^G$</td>
<td>$1.26 \pm 0.08$ mM/min</td>
</tr>
<tr>
<td>$K_{NAD+}^G$</td>
<td>$3.17 \pm 0.32$ μM</td>
</tr>
<tr>
<td>$K_{Gluc}^G$</td>
<td>$4.56 \pm 0.41$ mM</td>
</tr>
<tr>
<td>$K_i^{NAD+}$</td>
<td>$6.21 \pm 0.40$ μM</td>
</tr>
<tr>
<td>$E_{0}$</td>
<td>0.027 μM</td>
</tr>
<tr>
<td>T</td>
<td>37 °C</td>
</tr>
</tbody>
</table>
9.5 Step 4: Calculations

9.5.1 Model implementation and dynamic simulation

Dynamic process simulation is performed with the information presented in Step 2 (Section 9.3) and Step 3 (Section 9.4). Furthermore, the influence of substrate concentrations (acetophenone and L-alanine) on the reaction rate is analyzed for the transamination due to substrate inhibition (see Equation 9.1) [114]. At the reported conditions, initial concentrations of acetophenone and L-alanine are too low in the process for a promising industrial application. Consequently, substrate concentration ranges have been increased in order to find more relevant conditions to operate the process.

For both substrates the inhibition can be observed in Figure 9.3. It is analyzed that if both substrate concentrations are in a ratio of $\text{Ala}/\text{Ace} = 6$, then the archived reaction rates are higher than $0.02 \text{mM/min}$. In order to also obtain a relevant productivity at industrial level, acetophenone must be added at the highest concentration possible. In this case, the solubility limitation of acetophenone is at $50 \text{mM}$ [135]. Consequently, the selected acetophenone concentration is $48 \text{mM}$ to keep a safety range since the model does not include the influence of the solubility. According to Figure 9.3 and keeping the mentioned concentration ratio, the L-alanine concentration is taken as $288 \text{mM}$ for the process in order to achieve a reaction rate of $0.024 \text{mM/min}$ which is a high value according to the enzyme characteristics.

The dynamic simulation of the different scenarios for the process is compared in order to evaluate the achieved benefits of each configuration (see Figure 9.4). In the same manner, the used initial and operating conditions for each scenario are summarized in Tables 9.6 and 9.7.

Figure 9.3: Influence of reaction rate for transaminase at different substrate concentrations
### Table 9.6: Initial conditions for all scenarios used for the dynamic process simulation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{0}^{Aec}$</td>
<td>48 mM</td>
</tr>
<tr>
<td>$C_{0}^{Ala}$</td>
<td>288 mM</td>
</tr>
<tr>
<td>$C_{0}^{PEA}$</td>
<td>0</td>
</tr>
<tr>
<td>$C_{0}^{Pyr}$</td>
<td>0</td>
</tr>
<tr>
<td>$C_{0}^{NADH}$</td>
<td>0</td>
</tr>
<tr>
<td>$C_{0}^{Lact}$</td>
<td>0</td>
</tr>
<tr>
<td>$C_{0}^{NAD+}$</td>
<td>0.1 mM</td>
</tr>
<tr>
<td>$C_{0}^{Gluc}$</td>
<td>240 mM</td>
</tr>
<tr>
<td>$C_{0}^{Glt}$</td>
<td>0</td>
</tr>
<tr>
<td>$V_{0}^{*}$</td>
<td>95 l</td>
</tr>
</tbody>
</table>

[*] Initial volume condition is only valid for fed-batch operation i.e. Scenarios 3 and 4

---

### Table 9.7: Operating conditions for the dynamic simulation of the process in fed-batch operation (Scenarios 3 and 4)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{fd}^{Aec}$</td>
<td>500 mM 2300 mM</td>
</tr>
<tr>
<td>$C_{fd}^{Ala}$</td>
<td>750 mM 2530 mM</td>
</tr>
<tr>
<td>$F$</td>
<td>for $t &lt; 24h$ for $t &lt; 34h$</td>
</tr>
<tr>
<td>$F$</td>
<td>$F = 0.2 \text{ l/h}$ $F = 0.2 \text{ l/h}$</td>
</tr>
</tbody>
</table>
From the different process configurations, progressive improvements are expected for the process. In Figure 9.4, it is observed that Scenario 2 achieves full conversion of the reaction after 10 hours according to the limiting substrate (i.e. acetophenone). In this case, the reaction rate is improved due to the main product removal (i.e. PEA) which really shows a positive effect increasing the reaction rate in the full process and moving the equilibrium to the product side.

On the other hand, it is interesting to observe that Scenario 3 shows no improvements at all when compared with Scenario 1. This is due to the process conditions that were applied, since both substrates are kept at the optimal ratio in the process and then no inhibitory effects are observed at those conditions. However, Scenario 4 shows a remarkable improvement assuming that the product is successfully removed and the process is not limited by substrates since they are provided for the defined period.

For the scenarios operated in fed-batch mode, the feeding flow rate is stopped after a time (see Table 9.7) in order to consume the rest of the acetophenone present in the media. In that manner, the downstream processes are facilitated.

Evaluating the suggested scenarios, Scenario 4 is selected as the most promising configuration that can reach applicable concentrations of the product to be implemented at industrial level. Therefore, the reliability of all parameters in the Scenario 4 model is analyzed by applying sensitivity analysis.
9.5.2 Sensitivity analysis

The Scenario 4 shows a significant improvement in the production of the desired product. Therefore, it is the configuration selected to analyze the significance of the parameters in the model. In that case, sensitivity analysis is performed using the Morris method (See Chapter 4 Section 4.5).

The number of factors (parameters) to be evaluated is in total $n = 19$ of which 11 parameters are originating from the transamination (see Table 9.3), 4 parameters from the LDG (see Table 9.4), and 4 parameters more from GDH (see Table 9.5). Since confidence intervals are not available for the parameters, these values are assumed to have a deviation of +/-5% from the original values.

In order to generate the Morris sampling, the number of levels is selected as $p = 6$, and the number of samples are taken as $r = 10$. Applying Equation 4.13, the optimal perturbation factor is $\Delta = 0.6$. Applying Equation 4.14, the method required $me = 200$ model simulations for each sampling point.

From the simulation, the elementary effects are obtained by applying Equation 4.15, and thus the mean and standard deviation of each distribution $F_i$ are visualized graphically in order to identify the significant parameters in the model, as shown in Figure 9.5.

Figure 9.5 shows the plot for the acetophenone output. Other model outputs are not shown here since they have shown similar trends in the result. In that manner, parameters from the transamination reaction are quickly identified as the most relevant ones for the evaluated model.

9.6 Conclusions

Kinetic parameters for the enzymes LDH and GDH have been simplified to the initial reaction rate expression in this case study due to the lack of parameter values from the literature. It would be interesting to investigate the influence of products in each reaction in order to evaluate the necessity of including them in the whole reaction rate expression, as described in Equation 9.2. In that case, a full characterization of the enzyme kinetics is required in order to have a more reliable description of the enzyme dynamics.

Substrate inhibition in the transamination has been reported in the literature and it was taken into account in the reaction rate model ($r_{TAm}$). However, the
reported concentration ranges were $1 - 10 \text{mM}$ for acetophenone and $10 - 50 \text{mM}$ for L-alanine which are considered low for an industrial application. For this case, those substrate concentrations have been increased without validating the prediction power of the model. Consequently the further analysis here developed are useful to show how the model can be exploited in the evaluation of different conditions. However to validate the obtained results, the process has to be carried out at the conditions explored here.

From the sensitivity analysis, it can be concluded that there are many parameters that are irrelevant for the configuration and conditions selected. It is well known that results from these simulations must be validated in order to achieve a reliable implementation at an industrial application. However similar behaviors and trends, as simulated, can be expected in the process due to the reaction characteristics.

9.7 List of symbols and nomenclature

Abbreviations

LDH Lactose dehydrogenase
Case Study 3. Tri-enzyme production of 1-phenylethylamine

GDH  Glucose dehydrogenase
PEA  1-phenylethylamine
ISPR  in-situ product removal

Symbols

\( n \)  Input factors
\( p \)  Number of levels
\( r \)  Number of samples
\( mc \)  Model evaluation cost
\( F_i \)  Distribution of elementary effects

Nomenclature

\( V \)  Maximum initial velocity (mM/min)
\( K_M \)  Michaelis-Menten constant (mM)
\( K_{si} \)  Substrate inhibition constant (mM)
\( K_i \)  Inhibition constant (mM)
\( E_0 \)  Initial enzyme concentration (\( \mu \)M)
\( C \)  Concentration (mM)
\( C^0 \)  Initial concentration (mM)

Greek letters

\( \Delta \)  Optimal perturbation factor

Subscripts

\( TAM \) \( \omega \)-transaminase
LDH  Lactate dehydrogenase
GDH  Glucose dehydrogenase
Ace  Acetophenone
Ala  L-alanine
PEA  1-phenylethylamine
Pyr  Pyruvate
NADH  Cofactor
Lact  Lactate
9.7 List of symbols and nomenclature

\[ NAD^+ \quad \text{Cofactor} \]
\[ Gluc \quad \text{Glucose} \]
\[ Glc \quad \text{Gluconolactone} \]

**Superscripts**

\[ Ace \quad \text{Acetophenone} \]
\[ Ala \quad \text{L-alanine} \]
\[ PEA \quad \text{1-phenylethylamine} \]
\[ Pyr \quad \text{Pyruvate} \]
\[ Gluc \quad \text{Glucose} \]
\[ NADH \quad \text{Cofactor} \]
\[ NAD^+ \quad \text{Cofactor} \]
\[ f \quad \text{Forward direction} \]
\[ r \quad \text{Reverse direction} \]
\[ fd \quad \text{Feeding} \]
Case Study 3. Tri-enzyme production of 1-phenylethylamine
The main goal of this thesis has been to establish a methodological framework that helps in model-building for multi-enzyme processes. The final methodology is able to guide not only model formulation (i.e., mass balances based on first engineering principles) but also the mathematical implementation of the model and further techniques for model quality evaluation. This methodology has been developed by extracting the particular characteristics of multi-enzyme processes presented in the literature and the development of three relevant case studies which have also been used to tune and validate the methodology.

The general overview of the methodological framework has been explained in the introduction to the thesis. However, each step has been covered in more detail in Chapter 3. The methodology has been developed in a way that models can realistically represent the dynamic behavior of multi-enzyme ex-vivo processes, while keeping a reasonable degree of mathematical complexity, as explained in Chapters 3 and 4, and shown in the case studies.

The difference between this methodology compared to others is that the particular features of multi-enzyme ex-vivo processes have been introduced in each step (e.g., relevant modeling considerations and the decision-making in the modeling procedure), the incorporation of state-of-the-art methods which provide background and insight into their application for model development purposes, the step-by-step procedure which includes the required tools and methods to de-
Develop and implement multi-enzyme process models, and the possibility to re-use knowledge, experience and insight from case study to case study. In this way, the methodology provides a systematic methodology for modeling, including a structure, guidance, documentation and support to the modeller.

Step 2 of the methodology (information) is considered as one of the most important steps in the methodology because organized and precise information about the problem not only facilitates modeling development but does also contribute with a more realistic analysis of the results. It has been experienced that the compilation of information is actually the step that consumes most of the time when applying the methodology (without taking experimental work into account). Consequently, special attention has been paid on developing this step emphasizing the information that has been found as the most relevant for the building and analysis of multi-enzyme process models. In this way, the different considerations that are required to understand the reactions and process features have also been highlighted. Furthermore, the interaction matrix has been suggested in order to identify the type of interactions that exist among compounds and enzymes in a multi-enzyme process. This information is useful to keep a clear idea in terms of kinetic parameters that the full model structure should take into account.

Due to the full multi-enzyme process model complexity, a decomposition of the modeling problem has been suggested. Consequently reaction rate expressions are first developed for single enzymes and afterwards the full mathematical model is complemented with transfer phenomena, reactor type and operating mode. In order to evaluate the reliability of all parameters in the model, a global sensitivity analysis has been suggested. In that manner, results are used to prioritize further analysis for significant parameters or to consider model simplifications of insignificant parameters, all with statistical support, as illustrated in Chapter 8.

In Step 4 of the methodology (Calculations), different mathematical methods have been suggested to solve and analyze the model structures. In the same way, most of these methods have been tested in the case studies developed here. The selected quantitative mathematical methods have been found as the most appropriate for the type of mathematical problems that the modeller can face with multi-enzyme processes. However, in the scientific literature there are several methods available that can also be used for the same or more specific problems that have not been covered in this thesis. Optimal experimental design has been proposed as a method to prioritize and emphasize primary measurements. Knowing the positive impact that this step can have in the development of multi-enzyme process models, a close collaboration has been started with the BioMath research center at Ghent University in Belgium in order to develop this further. Hence the third case study (the production of 1-phenylethylamine) is currently
a study jointly with BioMath (Ghent) for further analysis since the complexity of the process model has been too high and the available experimental data have been insufficient even for partial model validation.

This methodological framework was implemented as a MATLAB toolbox which incorporates the required methods and tools to carry out the methodological approach. The development of the MATLAB codes has been based on the previous knowledge of the author and on the basis of course material which has been used as a support in implementation of the particular case studies. In this manner, the MATLAB codes have been modified and adapted in a way that enable management of the available information for each case study. Several MATLAB functions have been programmed in a general manner, so they can be used in other case studies without modifying the original code. In that way, they can be used as a basis for building a computational tool for analyzing many types of multi-enzyme process.

Different types of limitations for the implementation of multi-enzyme processes, such as enzyme stability, substrate or product inhibition, and media compatibility among others, have been mentioned throughout the thesis. However, Chapter 5 has discussed in particular how these limitations can be analyzed on a ‘virtual’ basis using mathematical models in order to find the required improvements for either the process configuration or the enzyme performance until a desired productivity of a given process is achieved. Likewise, goals for potential improvements of multi-enzyme processes can be formulated by employing tools that originate from fields such as process engineering, enzyme immobilization and protein engineering.

From the literature review, a classification has been suggested for multi-enzyme processes. This classification has been found relevant to present in this thesis not only to clarify ambiguous definitions for multi-enzyme processes, but also to focus the research effort on the concepts required to understand a specific group of multi-enzyme ex-vivo processes. Furthermore, the advantages and disadvantages of both in-vivo and ex-vivo processes have been discussed specifically when carried out in single and multiple reactor(s). In the same way, special emphasis has been placed on the operability of both process groups with respect to cell/biocatalyst constraints, reaction constraints, controllability, and monitoring among others.

It also has been discussed that multi-enzyme processes are mostly applied for in-situ cofactor regeneration and seen as an innovative alternative for effectively recycling the cofactors, which are generally expensive. Although cofactor limitation seems to be solved, in reality cofactors are also found highly sensitive to the media conditions (degradation/denaturalization problems). Hence cofactor stability also needs to be analyzed in order to ensure a proper multi-enzyme
process application.

The methodological framework has been developed and tested for different case studies. The case studies have been selected according to the availability of information in the literature, as well as the reliability of the multi-enzyme process not only with respect to the degree of product importance but also with respect to the type of reaction structure that is generated due to the enzyme combination. In this case, the main drawback has been the poor quality/quantity of experimental data reported in the scientific literature suitable for modeling e.g. clearly defined operating conditions, enzyme concentrations or progress curves. Consequently, identifiability problems were faced in the development of the case studies. Since limited experimental data was available, the most reliable kinetic parameters have been overcome by fixing them at their reported value. In this way, only identifiable parameters were required to be estimated.

For the case studies developed here, conclusions according to the specific model objective have been presented in the corresponding chapters. However, in general, each case study has contributed to a better understanding of multi-enzyme processes and thus in the development of the methodology. Furthermore, in the three cases it has been imperative also to be aware of the complexity of these processes especially for model calibration where not only the model structure formulation is important, but also the experimental data available for the process. It has been learnt that the modeler has a huge advantage when the desired case study is implemented and available data been provided from experimentation in the laboratory, because the modeling procedure is incremental and a number of iterations in the methodology are required before a model can be validated for the defined model objective. In the same way, it has been experienced that help from expertise in the laboratory is highly valuable since it offers information that forms a key for understanding the process in an efficient manner.

Finally, many of the multi-enzyme process modeling challenges have extensively been discussed and solved in this thesis. There are some others that have been opened for further discussion such as the necessity of better experimental data acquisition and implementation of optimal experimental design. Consequently, Chapter 11 presents some recommendations for future directions of multi-enzyme process modeling. In this way, multi-enzyme processes can become a powerful technology at an industrial level in the near future.
Chapter 11

Suggestions for future work

During the development of this thesis several exciting challenges have been addressed. Likewise, there are some further topics and issues that can be interesting to study in order to improve the methods and tools that were already developed here. Consequently a description of the new challenges that have been opened by the work done in this thesis are summarized below in order to guide and advise the work that can be carried out in the future.

Parameter identifiability is a significant issue for multi-enzyme process models. Assuming that a model structure is correct, and there is a lack of good experimental information, two practices are recommended in order to improve the information quality suitable for modeling: 1) the first recommendation refers to the evaluation and implementation of optimal experimental design in the laboratory, so that experimental work only focuses on what is really relevant for the process and its model for a specific purpose, and 2) the second suggestion is to use standardized databases for enzymatic processes.

The generation of a database for bioprocesses promotes a standardization of how experimental data must be reported. Furthermore it facilitates the information search in the literature which can be quite extensive for some cases. The minimum and essential information that can be included in a database is the progress curves of the process, characteristics of the enzymes involved in the process (e.g. enzyme mechanism and kinetic parameters), and clear initial
Suggestions for future work

conditions of the process (e.g. substrate and enzyme concentrations).

A procedure for optimal experimental design has been briefly introduced in this methodology. In the same way, the general conclusion between all the case studies developed here has been the need for better experimental information. Hence further developments in the implementation of optimal experimental design for either these case studies or other suitable cases are recommended. In this way, the methodology can be improved by including the practical knowledge that is achieved when implementing a real process.

In a similar manner, a significant contribution to the methodological framework was obtained from what was learnt in the case studies. The methodological framework was made generic and enabled accommodation of model developments for process control and optimization objectives. Here, some case studies with these types of application or objective are desirable to support the modification of the developed methodology.

Working with raw experimental data it is common to have limitations due to missing or outlying observations. The treatment of these limitations are relevant in order to have reliable experimental data (progress curves). In the same way, too many data points do not always mean more process information. For these cases, further studies on experimental design and data analysis can give sufficient support to know how to treat the experimental data in order to obtain and keep as much information as required and avoiding redundancy in the information. This issue can definitely offer a huge input to the methodology developed here.

From the literature and experience in the laboratory, some enzymes and compounds are found highly sensitive to degradation in the media, mainly by changes in temperature and pH. In modeling, a good enzyme stability is often assumed. In reality, this assumption is realistic only for short periods of time. This means that after a given time, enzyme degradation can be quite significant to such an extent that efficiency of the process can be highly affected. In order to ensure a successful multi-enzyme process application, further studies on enzyme stability and influence of temperature and pH on the enzymes should be carried out in order to take this information into account in the model formulation and process scale-up.

As mentioned in the conclusions, one of the most time consuming steps in the methodology is Step 2 (Information). It has furthermore been experienced with the third case study that the modeling development is remarkably more efficient when working together with expert people in the laboratory. That is why it is highly recommended to have a close collaboration between the modeler and people in the laboratory, if possible. In the end both sides can mutually benefit in order to achieve more challenging goals, faster process development and perhaps
opening new perspectives.

Another interesting step for future work is the development of a computational tool which can be built based on the computational methods that are already implemented for generic cases. It is known that well developed computational tools provide a faster work-flow and accelerate scientific advance in the development of any process. In this thesis, MATLAB codes have been generated to solve the different case studies. Furthermore, some of the functions have been developed in a generic form. Consequently they can be taken as the basis to develop them at a more advanced programming level. In this way, what it is really important is to create a user-friendly interface that allows assembly of models for a desired multi-enzyme process in a flexible manner.

The idea of developing a modeling methodology is not only to get a mathematical description of the process but also to build a way to achieve the implementation of real multi-enzyme in pot processes. From the multi-enzyme process review, it was realized that only a few cases have been implemented at pilot-plant or industrial scale. Considering this limitation in a positive manner, it probably can be one of the most fertile areas for future work on multi-enzyme in-pot processes. In this way, process engineering plays an important role, especially in the development of reactors that can best handle the combination of enzymes under similar media conditions. Another relevant point to discuss is the evaluation of the transition from batch to continuous operation in order to increase productivity using this approach. In the same manner at pilot plant or industrial scale, issues as process monitoring and process control become very important as well. Therefore, further research should be undertaken in those areas also, specifically to identify the types of measurements that can be obtained for production scale implementation (i.e. off-line, on-line or with soft-sensors). In the same manner, controllability of the process and suitable control structures must be evaluated in order to keep the operation at optimal conditions.

Enzyme immobilization shows many advantages in the application of multi-enzyme processes (e.g. enzyme recycling and stability). One promising application is the utilization of packed bed reactors. In these cases, different configurations can be envisaged according to the number of immobilized enzymes. The most accurate mathematical model formulation for these processes includes partial differential equations. A brief description of the PDE solution has been mentioned in this thesis. Hence, this can be taken as a starting point for the development of a more detailed procedure for the discretization methods that can best fit such multi-enzyme process models.

In this work, multi-enzyme processes have been studied to build reliable mathematical models. Nowadays the combination of bio- and chemo-catalysts in a process is feasible. Furthermore such implementation can improve conventional
processes. It is believed that some modeling similarities with multi-enzyme processes can be found in the implementation of these processes. Consequently, the methodological framework developed here can be re-usable for most of the steps in the applicability of bio- and chemo-catalized processes. In this case, the media compatibility of conditions could be a major issue to be analyzed.
Appendix A

Reactions rate expression of one and two substrates

This appendix compiles the reaction rate expressions for some of the enzyme mechanisms that are reported in the scientific literature [18, 62]. An enzyme mechanism explains how a single enzyme binds the substrate(s) and how the product(s) is released. This appendix includes the reaction rate expressions that are developed for enzymes that involve one and two substrates in their mechanisms. These reaction rate expressions are considered as the most important constitutive equations for the mathematical model formulation of multi-enzyme processes, as mentioned in Step 3 of the methodology (Section 3.5). Particular derivations of the reaction rate expressions are obtained when the enzymes are influenced by substrate(s) and product(s) (e.g. enzyme inhibition). These cases can be derived from steady-state kinetics using methods such as King-Altman or Wong-Hanes which are well explained in the literature [18, 62].

Enzyme mechanisms when one substrate is involved:

- The irreversible Michaelis-menten mechanism is illustrated in Figure A.1, and its corresponding reaction rate expression is derived as:

\[ r = \frac{V C_A}{K_m + C_A} \]  

(A.1)
Reactions rate expression of one and two substrates

\[ E + A \xleftrightarrow{} EA \xrightarrow{} E + P \]

**Figure A.1**: Irreversible single enzyme mechanism

- The reversible Michaelis-Menten mechanism is illustrated in Figure A.2, and its corresponding reaction rate expression is derived as:

\[ r = \frac{Vf_C_A - Vr_C_P}{1 + \frac{C_A}{K_m} + \frac{C_P}{K_m}} \]  

(A.2)

\[ E + A \xleftrightarrow{} EA \xleftrightarrow{} EP \xleftrightarrow{} E + P \]

**Figure A.2**: Reversible single enzyme mechanism

- The single enzyme mechanism including competitive inhibition is illustrated in Figure A.3, and its corresponding reaction rate expression is derived as:

\[ r = \frac{V_C_A}{K_m \left(1 + \frac{C_i}{K_{ic}}\right) + C_A} \]  

(A.3)

**Figure A.3**: Single enzyme mechanism with competitive inhibition

- The single enzyme mechanism including essential activation is illustrated in Figure A.4, and its corresponding reaction rate expression is derived as:

\[ r = \frac{V_C_A}{K_m \left(1 + \frac{K_x}{K_{ec}}\right) + C_A} \]  

(A.4)
The single enzyme mechanism including substrate inhibition is illustrated in Figure A.5, and its corresponding reaction rate expression is derived as:

$$r = \frac{VC_A}{K_m + C_A \left(1 + \frac{C_A}{K_{si}}\right)} \quad (A.5)$$

Enzyme mechanisms when two substrates are involved:

- The compulsory-order ternary-complex mechanisms is illustrated in Figure A.6, and its corresponding reaction rate expression is derived as:

$$r = \frac{V' C_A C_B}{K^{a}_i K^m K^{p}_i} - \frac{V' C_P C_Q}{K^{m}_i K^p} \left(1 + \frac{C_A}{K_{ai}} + \frac{C_B}{K_{bi}} + \frac{C_P}{K_{pi}} + \frac{C_Q}{K_{qi}} + \frac{C_A C_B}{K^{a}_i K^{b}_i K^m} + \frac{C_A C_P}{K^{a}_i K^p K^{m}_i} + \frac{C_A C_Q}{K^{a}_i K^q K^{m}_i} + \cdots \right) \quad (A.6)$$
In absence of products, Equation A.6 is reduced to the initial rate expression, thus:

\[
r = \frac{V C_A V_B}{K_a A K_B B + K_m A C_A + K_m B C_B + C_A C_B}
\] (A.7)

• The random-order ternary-complex mechanism is illustrated in Figure A.7, and its corresponding reaction rate expression is derived as:

\[
r = \frac{V C_A C_B - V C_P C_Q}{K_m K_N}
\]

(A.8)

In absence of products, Equation A.8 is reduced to the initial rate expression, thus:

\[
r = \frac{V C_A C_B}{K_a A K_B B + K_m A C_A + C_B K_m + C_A C_B}
\] (A.9)

• The substituted-enzyme mechanism is illustrated in Figure A.8, and its
corresponding reaction rate expression is derived as:

\[
r = \frac{V_f C_A C_B}{K^*_f K_m} + \frac{V_f C_P C_Q}{K^*_f K_m} - \frac{V_f C_P C_Q}{K^*_f K_m}
\]

(A.10)

**Figure A.8:** Reaction of two substrates - Substituted-enzyme mechanism

In absence of products, Equation A.10 is reduced to the initial rate expression, thus:

\[
r = \frac{V C_A C_B}{K^*_b C_A + K^*_m C_B + C_A C_B}
\]

(A.11)

### A.1 Nomenclature

**Nomenclature**

- **\( V \)**: Maximum initial velocity \((M/s)\)
- **\( C \)**: Concentration \((M)\)
- **\( K_M \)**: Michaelis-Menten constant \((M)\)
- **\( K_{si} \)**: Substrate inhibition constant \((M)\)
- **\( K_i \)**: Inhibition constant \((M)\)
- **\( K_{ic} \)**: Competitive inhibition constant \((M)\)
- **\( K_x \)**: Activation constant \((M)\)

**Subscripts**

- **\( A \)**: Substrate A
- **\( B \)**: Substrate B
- **\( P \)**: Product P
Reactions rate expression of one and two substrates

\[ Q \quad \text{Product Q} \]
\[ i \quad \text{Inhibitor} \]
\[ x \quad \text{Activator} \]

**Superscripts**

\[ f \quad \text{Forward direction} \]
\[ r \quad \text{Reverse direction} \]
\[ a \quad \text{Substrate A} \]
\[ b \quad \text{Substrate B} \]
\[ p \quad \text{Product P} \]
\[ q \quad \text{Product Q} \]
Appendix B

General programming schemes used to implement Step 4 (Calculations) in MATLAB

This appendix compiles the flow-sheets that describe the sequence of steps that were used for the implementation of the codes in MATLAB. In Figure B.1, the main call script is described. This offers the different options that can be selected for the calculations cited in Chapter 4. Option 1 is selected to perform the dynamic simulation of the model. In this case, the required input information is shown as well as the output of each block in order to solve the set of ordinary differential equations (see Figure B.2). Option 2 is selected to analyze the identifiability of parameters based on the collinearity index, as described in Chapter 4. Figure B.3 shows the required input information and the outputs of each calculation. Option 3 is selected to estimate the model parameter based on available progress curves of the process. Furthermore, the required input information for the calculation of the confidence intervals and the correlation matrix is shown (see Figure B.4). Option 4 is selected to performed uncertainty analysis. In this case, certain information calculated in other options is required as input for the new calculations, as shown in Figure B.5. Option 5 is selected to performed the sensitivity analysis. Figure B.6 shows the main
General programming schemes implemented in MATLAB

calculations required for this analysis.
Figure B.1: Sequence of steps that are used in the MATLAB implementation corresponding to the main call script.
Figure B.2: Sequence of steps that are used in the MATLAB implementation corresponding to the dynamic simulation of the mathematical models formulated following the methodology developed here.
**Figure B.3:** Sequence of steps that are used in the MATLAB implementation corresponding to identifiability analysis.
General programming schemes implemented in MATLAB

Figure B.4: Sequence of steps that are used in the MATLAB implementation corresponding to parameter estimation of the mathematical models.
Figure B.5: Sequence of steps that are used in the MATLAB implementation corresponding to uncertainty analysis of the mathematical models.
General programming schemes implemented in MATLAB
Bibliography


