Guilherme Duarte Camarinha Nunes

Bachelor of Science in Chemical and Biochemical Engineering

Effect of growth conditions on the activity of Michael hydratase from *Rhodococcus rhodochrous* ATCC 17895

Dissertation to obtain the Degree of Master of Science in
Chemical and Biochemical Engineering

Supervisor:	Prof. Dr. Ulf Hanefeld, Full Professor, TU Delft
Co-supervisors:	Hanna Busch, Ph.D. Student, TU Delft Prof. Dr. Rui Manuel Freitas Oliveira, Professor Associado com Agregação, FCT-UNL
Chairperson:	Prof. Dr. Mário Fernando José Eusébio, Professor Associado com Agregação, FCT-UNL
Member of the Examination Committe:	Dr. Maria Filomena Andrade de Freitas, Investigador Principal UCIBIO, FCT-UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

TU Delft
Delft University of Technology

September, 2017

Effect of growth conditions on the activity of Michael hydratase (Mhy) from *Rhodococcus rhodochrous* ATCC 17895

Copyright © Guilherme Duarte Camarinha Nunes, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

The Faculdade de Ciências e Tecnologia and the Universidade Nova de Lisboa have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

Growth is a process.

You gain, you lose, you cry, you grow, you learn, you try, you fail but you keep trying.

Time is the real luxury.

Acknowledgments

I would like to start by acknowledging my laboratory and daily supervisor, Hanna Busch, for giving me the opportunity to join the “Mission Mhy” and help in her Ph.D. project with the contribution of the present work. It was really challenging to work in a field that I was not used to, and her support with all the protocols and methods, and her knowledge, patience and effort were essential to achieve the goals of the project.

I would also like to extend my acknowledge to the project supervisor, Prof. Dr. Ulf Hanefeld, for accepting me in the biocatalysis group, for all the advices and recommendations given on the group meetings and for being always supportive.

A special thanks to Dr. Ir. Peter-Leon Hagedoorn and Dr. Carolin Bisterfeld for the wise recommendations and discussion of results during the project meetings. To Luuk Mestrom, also my gratitude for the recommendations. To Kasper Bergsma, for helping in everything at the laboratory and for being a good and responsible colleague, my gratitude.

An extended thanks to all the biocatalysis group and office colleagues for receiving me in such an open and friendly way and for the great work environment at TU Delft.

I would also like to thank some friends I made in Delft, namely Bianca Vijai, Leonor Prata and Tiago Silva, for their support and friendship.

To the human beings who gave me the opportunity to do my master thesis abroad, my beloved parents, I could not be more grateful! Together with them, I would also like to thank the rest of my wonderful family for the patience, support and encouragement during the most difficult times. Grácio, Carvalho, Camarinha and Nunes family, I am certain that without you this would not be possible.

Last but not least, I want to thank all my friends from Portugal who were always by my side in every moment, with special gratitude for two of them. Margarida Pinto and Gonçalo Marcelo, you are the perfect definition of friendship.

“A todos vós, o meu mais sincero obrigado!”

Abstract

The stereoselective water addition to double bonds is chemically hard to accomplish. However, the use of a novel Michael hydratase from *Rhodococcus rhodochrous* ATCC 17895 gives a promising alternative route. The aim of this study is to reveal the dependency of the enzymatic activity on the culture medium and a subsequent optimisation.

Focus was given to the composition of the microorganism's culture medium due to the fact that the literature protocol from 1998 was never changed in this regard. Moreover, the use of iron sulphate and magnesium sulphate in unusually high amounts has drawn the attention to an investigation.

By using a "one-variable-at-a-time" (OVAT) strategy, it was proven that the addition of iron sulphate is redundant which enhances the practical treatment of the whole cells.

Via a Design of Experiments (DoE) with the support of the software Design Expert v 7.7.0 (Stat Ease, Minneapolis, MN), optimisation of the culture medium parameters' glucose, yeast extract, peptone and magnesium sulphate was achieved. It was concluded that the addition of magnesium sulphate to the culture medium has no added value. The specific amounts of 6.59 g/L for glucose, 1.84 g/L for yeast extract and 9.20 g/L for peptone provided the best result in a restricted interval of study. Due to this culture medium optimisation, the enzymatic activity was improved by three times compared to the activity provided by the standard culture medium reported in literature.

Transferring the shake-flask optimised medium compositions to a fermentor led to large amount of biomass with high Michael hydratase activity thereby saving a substantial amount of growth time.

The results of this study provided a significant increase in specific Michael hydratase activity and thereby securing a sufficient amount of active whole-cells needed in the desired Mhy (Michael hydratase) isolation process.

Keywords: Biocatalysis, enzyme, Michael hydratase, optimisation, culture medium, design of experiments

Resumo

A adição esteroselectiva de água a ligações duplas é quimicamente difícil de se realizar, no entanto o uso de uma "Michael hydratase" presente em células de *Rhodococcus rhodochrous* ATCC 17895 fornece uma alternativa promissora. O objetivo deste estudo é desvendar a dependencia da atividade enzimática no meio de cultura e subsequentemente optimisá-lo.

Este estudo focou-se na composição do meio de cultura, por este seguir um protocolo de 1998 que nunca foi modificado. Além disso, o uso de sulfato de magnésio e sulfato de ferro em quantidades invulgarmente elevadas despertaram a atenção para uma investigação.

Usando uma estratégia de alteração de uma variável de cada vez, provou-se que a adição de sulfato de ferro ao meio de cultura é redundante, o que permitiu melhorar o tratamento prático das células.

Através de um Desenho de experiências (DoE), recorrendo-se ao programa Design Expert v 7.7.0 (Stat Ease, Minneapolis, MN), optimisou-se os seguintes parâmetros do meio de cultura: glucose, extrato de levedura, peptona e sulfato de magnésio. Concluiu-se que a adição de sulfato de magnésio ao meio de cultura não proporciona qualquer vantagem e que os valores específicos de 6,59 g/L para a glucose, 1,84 g/L para o extracto de levedura e 9,20 g/L para a peptona possibilitam o melhor resultado dentro de um intervalo de estudo restrito. Assim, devido a esta otimização, a atividade enzimática foi melhorada três vezes em comparação com a proporcionada pelo meio de cultura descrito na literatura.

Num fermentador testou-se o meio de cultura que foi otimizado em Erlenmeyers sob agitação, tendo-se obtido maior quantidade de biomassa com elevada atividade da "Michael hydratase", o que permitiu diminuir o tempo de crescimento.

Os resultados deste estudo proporcionaram uma melhoria da atividade da enzima "Michael hydratase", assegurando que uma quantidade suficiente de células com elevada atividade enzimática possa ser utilizada no processo de isolamento desta enzima.

Palavras-chave: Biocatálise, Michael hydratase, optimização, meio de cultura, desenho de experiêcias

List of contents

List of Figures

List of Tables

Acronyms and Abbreviations

ANOVA – Analysis of variance

ATCC – American Type Culture Collection

BC –Before Christ

CCD – Central Composite Design

DCM – Dichloromethane

DI water – Deionised water

DMSO – Dimethyl sulfoxide

DoE – Design of Experiments

DSM – Deutsche Sammlung von Mikroorganismen und Zellkulturen

EtOAc - Ethyl acetate

FID – Flame ionisation detector

GC – Gas chromatography

KPi-buffer – Potassium phosphate buffer

Mhy – Michael hydratase

NMR – Nuclear magnetic resonance

OD_{600} – Optical density (600 nm)

PE – Petroleum ether

R&D – Research and Development

RSM – Response Surface Methodology

R. rhodochrous – *Rhodococcus rhodochrous*

rcf – relative centrifugal force (units, x g)

rpm – revolutions per minute (units, min^{-1})

slpm – standard litter per minute

t/a – tons annual (tons per year)

CHAPTER 1

Introduction

1.1 Biocatalysis

Biocatalysis is the genuine definition of catalysis in living (biological) systems. Using a natural catalyst, such as enzymes, chemical transformations on organic compounds can be performed.[1]

Looking back to the oldest records of brewing, 6000 BC (Before Christ), the Sumerians and Babylonians were the pioneers in using the effects of the microorganisms to produce alcoholic beverages from barley. With a lack of knowledge on bioprocesses, this was the first commercial application of biocatalysis, being the procedures closely connected with the cultural history of mankind.[2] Nowadays, biocatalysis as part of the biotechnology field, is understood as the integrated application of engineering and natural sciences targeting the technical use of organisms, cells or parts thereof.[3]

In the biotechnology field, production processes can be distinguished between so called bioconversions (sometimes also called biotransformations) and the fermentation processes. A bioconversion is an enzyme- or cell-catalysed reaction of defined starting material(s) to yield defined product(s), often with high yields and enantioselectivities. Usually this is a one-step reaction in chemical production processes (e.g. during the production of optically active products and intermediates).[3]

Fermentation - from the Latin word *fermentare*, denoted as to leaven or to brew - is not limited to the anaerobic fermentative metabolism. From a biotechnologist point of view, it means the fermentative production of renewable raw materials by living microorganisms, through an entire metabolic pathway, not just one single enzymatic step. In this case, the product accumulates in the fermentation broth and, after a set of reactions, the by-products accumulation is higher when compared to a single bioconversion. Figure 1.1 highlights the difference between a fermentation procedure and a biotransformation.[3]

On the past years, in organic chemistry, biocatalysis has become more and more interesting for academic and industrial synthesis, providing a complement to the chemocatalytic processes. [4]

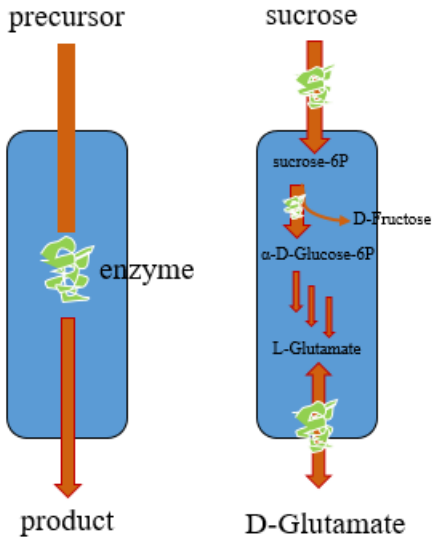


Figure 1.1 Difference between biotechnological processes: bioconversion (left) and fermentation (right). With bioconversion, only one synthesis step is carried out with a biocatalyst. The industrial production of glutamate is a successful fermentative process.[3]

In bioconversions, enzymes are utilised in order to speed up chemical reactions, replacing chemical catalysts, therefore improving the efficiency of a wide variety of industrial processes. The most important requirements for a catalyst in technical processes are selectivity, activity and stability. There are no catalysts as active as enzymes. Owing to their high specific activities, enzymes can be used in very small ratios relative to the substrate, about 0.0001-0.001 % (catalyst/substrate ratio), quite lower than in chemical catalysis (0.1-1 %).[3]

Selectivity also plays a fundamental role. Chemo-, regio-, and stereo-selectivity are terms used when the starting materials are prochiral and the products are stereoisomers produced in unequal amounts.[6] Due to their chirality, enzymes often allow the synthesis of chiral compounds with a possible enantiomeric excess higher than 99%, making them significantly superior to classical chemocatalysts with respect to their stereoselectivity.[3]

On the other hand, enzymes sometimes provide a frequent disadvantage on the biotransformation processes, which is the lack of stability. Therefore, the costs of catalyst production can play an important role in the economy of a biocatalytic procedure. Hence, inexpensive and reproducible production of the corresponding enzymes is an important success factor of industrial bioconversion.[3]

Chemical processes frequently run under high pressures and high temperatures. On the contrary, enzyme-catalysed processes usually work at low temperature and moderate pH. In addition, bioconversions often allow an economical use of material, which, for the chemical industry, means saving in terms of energy, raw materials as well as the avoidance of waste making them the most environmentally-friendly solution for industrial manufacturing.[3]

1.2 Commercial application of enzymes

The global market for industrial enzymes is expected to reach nearly $6.2 billion by 2020.[2] Examples of applications are the production of food and beverage, in addition to household items, including detergents and cosmetics. Products such as laundry detergents contain enzymes that aid the removal of stains and enable low-temperature washing. Ethanol and biofuels like biodiesel or bioethanol as well as concrete, leather and textiles, where the enzymes are not part of the end-product, are other kinds of successful applications.[2, 6-7]

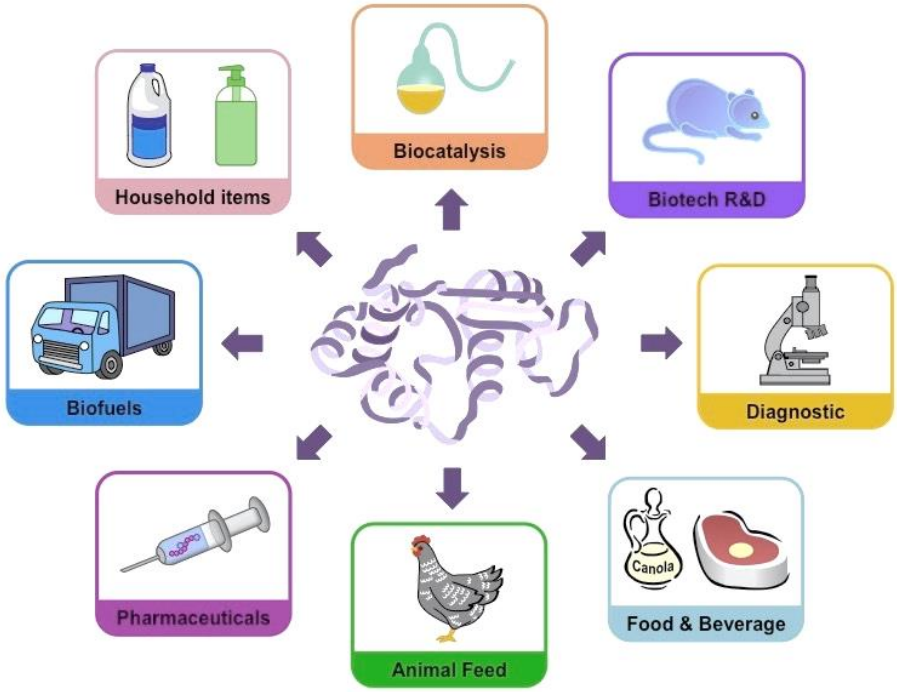


Figure 1.2 Broad range of applications of enzymes. They are used on daily-basis products such as household items, food and beverage and animal feed as well as for biotechnology research and development, pharmaceutical and fine chemical products, and also diagnostic of diseases.[6]

Additionally, enzyme-catalysed processes have been successfully applied in the production of fine chemicals and active pharmaceutical ingredients. The main focus of these processes is the achievement of chemical transformations for which efficient and sustainable solutions do not exist. The demand for new methodologies leads the pharmaceutical industry to the forefront of most of the recent researches and challenges.[4]

1.3 Research and development of biocatalysts

Very often, biotechnologists are confronted with two main challenges – the identification of products whose production by an enzymatic route is advantageous and the development of a process in the shortest possible time and with the minimum of resources.[3]

The development begins when the substrate and targeted molecules are known. Subsequently, it is necessary to identify a suitable biocatalyst and perform the synthesis of the starting material for the enzymatic step as well as the downstream processing. In doing so, a complete procedure combining classical chemical and enzymatic steps is implemented.[3]

Before starting an entire new research, one can search among commercial available enzymes in order to find the desired biocatalyst. The knowledge of the catalyst's mechanism may be helpful during the selection of enzymes, because often they have a broader application then what their name suggests.[2]

If the desired enzyme is not commercially available, one can also test microorganisms from strain collections like the American Type Culture Collection (ATCC) or Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSM). For centuries, living organisms played a fundamental role in providing enzymes with a broad range of applications on common products (*e.g.* bread, wine, vinegar). They are the most used source for industrial enzymes due to not only their easy availability and fast growth, but also the possibility in R&D (Research and Development) of being enriched and screened according to enzyme activity.[2-3]

Moreover, with the advent of recombinant DNA technology and protein engineering, a microbe can be manipulated and cultured in large quantities for elevated enzyme production and scientific development. The basic principles of direct evolution (i.e., mutation, selection and recombination) can be exploited in the laboratory in order to improve issues such as thermostability, substrate specificity, enantiomer selectivity and stability.[3]

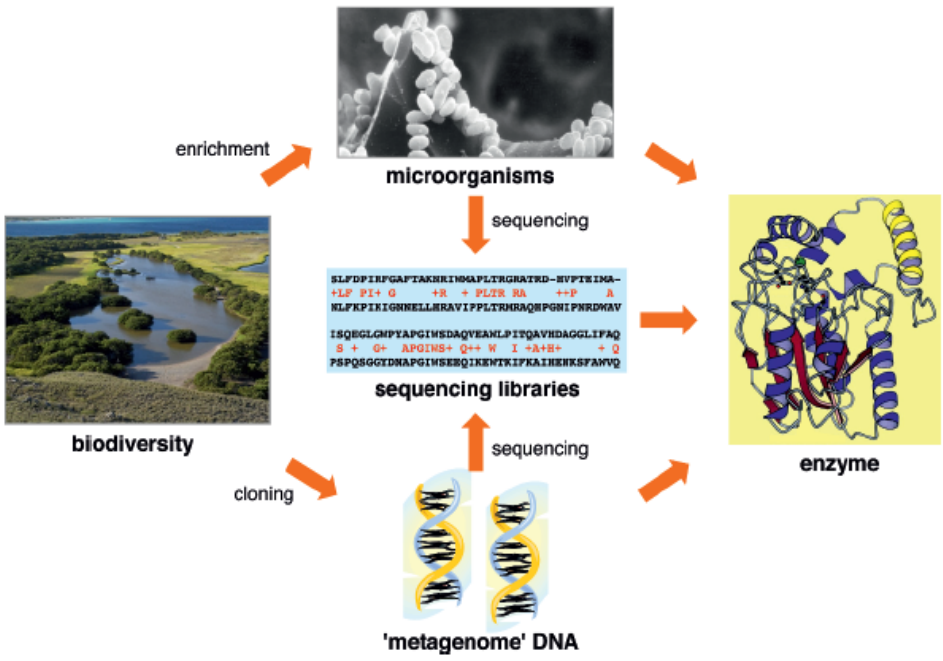


Figure 1.3 Enrichment, cloning and sequencing of microorganisms to yield access to novel biocatalysts. 'Metagenome' DNA is referred to the DNA that is to be isolated from an environmental sample.[3]

Another common issue is that wild-type strains, as isolated from nature, rarely produce enough of the enzyme. A solution is gene cloning and expression in a suitable host strain (*e.g. Escherichia coli*, *Bacillus subtilis*, *Pichia pastoris* or *Aspergillus*) in order to make the enzyme profitable and available in sufficient amounts. After that, the best fermentation conditions for the strain have to be determined. Therefore, factors such as the optimal medium, pH, feeding profiles, as well as aeration and stirring speed need to be optimised.[3]

The development requirements in the area of biocatalysis lie in the fast and effective access of new enzymes with desired characteristics. In particular, interest is focused on the use of enzymes for chemically difficult industrial reactions, like the water addition to carbon-carbon double bonds or the sophisticated regioselective introduction of oxygen with the help of oxygenases.[3]

1.4 Water addition to carbon-carbon double bonds

The water addition to (or hydration of) carbon-carbon double bonds is a reaction of great interest among the organic chemists, as water is considered an unreactive molecule.[8-10] In this case, it is used as nucleophile and solvent, allowing a route to the synthesis of alcohols or hydroxy carbonyl compounds depending on the mechanistics of the reactions.[10]

Mechanistically, two groups of hydration reactions can be distinguished: electrophilic addition and Michael addition. The first happens on isolated C=C bonds, following the rule of Markovnikov. Here, a H_2O is added as a nucleophile after a protonation of the isolated unpolarised C=C bond.[8] Besides determining the regiochemistry of the reaction, it also indicates the absence of regioselectivity if the C=C is in a linear carbon chain.[10] In this case, the reaction yields alcohols. The second happens in the presence of α,β-unsaturated (Michael) acceptors. Here the carbon-carbon double bound is polarised by an electron withdrawing group, as for instance, in ketones, aldehydes, lactones, carboxylic acids, thioesters or phosphate groups.[8-9] One example of the Michael-type is the water addition to α,β-unsaturated ketones in order to form β-hydroxy ketones, which are compounds of high concern for the synthesis of structural motifs in natural products.[9] Figure 1.4 exhibits a general scheme of the electrophilic and Michael water addition to C=C bonds.[8, 10]

The water addition (to C=C bonds) is, thermodynamically, an equilibrium reaction in which the equilibrium could either lie on the side of the alkene as for the electrophilic addition reaction or on the side of the alcohol, where a great interest lies.[8-10] However, this statement needs to be confirmed for every reagent as the thermodynamics of a reaction are always depending on the starting material.[10]

In chemical processes, the electrophilic addition of water to double bonds has been tried, however chemists experienced difficulties, mainly in the activation of water as nucleophile.[10] Only a few methods for the hydration of alkenes are reported, using acid catalysts and harsh reaction conditions. The production of *tert*-butanol, ethanol and similar alcohols are examples of acid-catalysed water

additions, as well as the described hydration of propene.[8, 10] Most of these processes deal with the same issues, as for instance high temperatures and pressures in addition to large recycle streams and unfavourable equilibria towards the alcohol. Also reported in the textbooks, the presence of other nucleophiles needs to be carefully avoided due to the protonation of water and consequently loss of its nucleophilic character when the double bond is acid induced.[8]

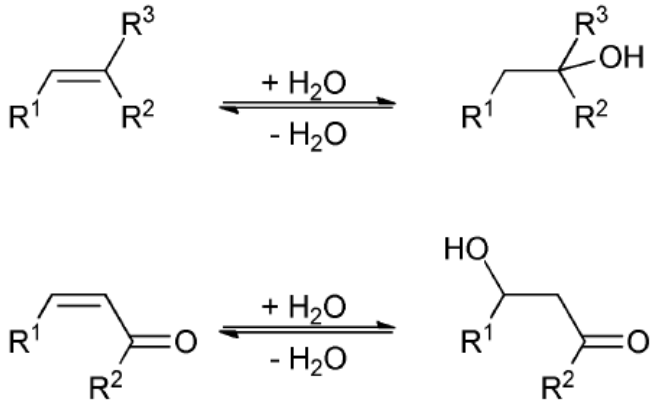


Figure 1.4 Water addition to isolated C=C bonds, following the rule of Markovnikov (above). Michael addition to α,β-unsaturated carbonyl compounds (bottom).[8, 10]

On the other hand, the addition of water proceeds more readily in Michael-type additions, and can be either acid or base catalysed, activating either the α,β-unsaturated carbonyl compound or the nucleophile, water. An example of a base-catalysed Michael addition is the conversion of *cyclo*-hex-2-enone to 3-hydroxy-*cyclo*-hexanone, using proteinogenic α-amino acids. In base-catalysed Michael additions, the conditions are, in comparison to the acid catalysed electrophilic water additions, very mild, however the reactions are also limited by the equilibrium.[8]

Consequently, only very few water addition reactions to C=C double bonds have been performed on industrial scale, and most of them were discontinued due to their inefficiency.

1.4.1 Enzyme-catalysed water addition to C=C bonds

In contrast to the chemical catalysis, enzyme-catalysed water addition on isolated or conjugated C=C double bonds is an essential reaction in nature. Most of the known enzymes are part of the primary metabolism but can also be found in the secondary metabolism to a smaller extent.[10, 11] For instance, an essential primary metabolic reaction is the addition of water to fumaric acid or to aconitic acid, both being part of the citric acid cycle.[10-11] Those reactions of the primary metabolism allow the synthesis of chiral products with high yields and enantiomeric excess.[8-10]

Synthesis of primary, secondary and tertiary alcohols given by such challenging reactions is of great attentiveness in preparative organic chemistry. The enzymes that catalyse the addition of a

molecule to double bonds are classified as lyases. In the case of the water addition to double bonds, those enzymes are called hydratases or hydro-lyases.[10]

Some hydratases are already applied in large-scale biotransformations, both for the electron-rich water addition to isolated C=C bonds and polarised electron-poor Michael-type addition, providing high yields and purities. An example of the first is the oleate hydratase from a bacterial strain for the production of γ-dodecalactone which is known as an essential flavour compound in whiskey. For the Michael-type additions, an industrialised enzyme is the fumarase, used in the production of enantiopure *(S)*-malate, the second-most widely used acidulant on the food industry (2500 t/a).[9-10]

Despite being successfully used on industrial scale for a specific production, those enzymes as well as many others not described in the present work have a limited substrate range.[10] This is a phenomenon typical for enzymes of the primary metabolism where perfect substrate specificity is required, mainly to ensure that within the cell no undesired side-reactions occur. Consequently, their practical applicability is very limited.[9-10]

1.4.2 Michael hydratase from *Rhodococcus rhodochrous*

A few years ago, several studies showed that a broader flexibility in the substrate spectrum for hydratases is possible. A novel Michael hydratase expressed in wild-type strains of *Rhodococcus rhodochrous* ATCC 17895 was found. This remarkable hydratase, as the name suggests, performs Michael-type additions of water to C=C bonds and is capable to accept substrates that are not part of the primary metabolism.[11] It was first reported in 1998 when Holland & Gu[12] used the whole cells of *R. rhodochrous* ATCC 17895 to prepare (*R*)-3-hydroxy-3alkylbutanolides by hydration of 3-alkyl-2-butenolides. This research opened a new approach for the preparation of hydroxylactones, useful chiral starting materials for organic synthesis.[12]

Recently, the research on this topic was continued by Chen *et al.*[11] who studied the substrate scope and limitations of the Michael hydratase in *R. rhodochrous* ATCC 17895 and screened several *Rhodococcus* strains as promising biocatalysts for the enantioselective Michael addition of water to a variety of α,β-unsatured carbonyl compounds.[11] Additionally, biotransformation conditions (time, pH, substrate concentration and temperature) of 3-methylfuran-2(*5H*)-one into (*S*)-3-hydroxy-3-methylfuranone catalysed by the putative Michael hydratase from *R. rhodochrous* ATCC 17895 were optimised and mechanistic and recyclability studies were accomplished.[11]

The efficient acceptance of a broad range of substrates from the Michael hydratase, as well as its presence in several *Rhodococcus* strains was proven.[11] Ongoing research concentrates on the isolation and characterisation of this valuable enzyme. With this, the toolbox of available hydratases with potential applicability in the synthesis of regio- and stereoselective alcohols can be expanded.

Last but not least is that the structural knowledge of this enzyme could help chemists in the design of new synthetic catalysts (artificial enzymes) with a high substrate acceptance.

1.5 Aim of the project

Within the scope of recent research attempting the isolation and characterisation of the known Michael hydratase from *Rhodococcus rhodochrous* ATCC 17895 it was found that, on the contrary of what is reported in the literature, the protocol for the substrate synthesis leads to the formation of (*E*)-4-hydroxy-3-methylbut-2-enoic acid **1** instead of 3-methylfuran-2(*5H*)-one. Due to this fact the substrate scope of the promising enzyme had to be re-evaluated (unpublished data). Figure 1.5 shows the Michael addition of water catalysed by whole cells of *R. rhodochrous* ATCC 17895 with the correct substrate, synthesised according to the literature protocol.[12]

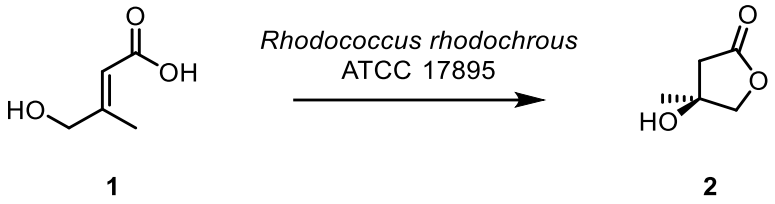


Figure 1.5 Reaction equation of the addition of water catalysed by Michael hydratase present in *Rhodococcus rhodochrous* ATCC 17895 using (*E*)-4-hydroxy-3-methylbut-2-enoic acid (1) as model substrate to give 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one (2).

On the other hand, the protocol from 1998 used for the growth of the microorganism was never modified. There are compounds whose use in the culture medium aroused curiosity. As for instance, the iron sulphate and magnesium sulphate are added in unusually high amounts.[12] Moreover, the iron sulphate is a problem in the OD_{600} analysis as its precipitation causes turbidity in the culture medium which interferes with the absorbance measurements (unpublished data). As a proof, figure D.1 in Appendix D evidences the precipitation of iron sulphate on the centrifuge tubes after the washing step described in section 3.6.

The low yields recently obtained on the biotransformation of (*E*)-4-hydroxy-3-methylbut-2-enoic acid **1,** that is used as a model substrate to check the Michael hydratase enzymatic activity, evidence a low expression of the enzyme, even with the biotransformation conditions already optimised (unpublished data).

All of this complicates the process of isolation of the enzyme, therefore there is a need of an extra study focused on the improvement of the Mhy activity and on the reveal of the role of the magnesium sulphate and iron sulphate on the culture medium and, particularly, on the Mhy activity.

Focus will be given to the composition of the culture medium, as it was never a motif of study. Culture medium formulations contain hundreds of ingredients in water solutions such as minerals, vitamins, peptones, amino acids, meat- and yeast- extracts or hydrosylates, inhibitors and proteins. Some

of these ingredients may be critical for cell growth or product synthesis, others may be toxic at certain levels, and many may be involved in complex interactions in the same or competing pathways within the cell. Moreover, an intrinsic relation between protein-encoding genes and the composition of the culture medium may exist. Only investigations in this regard will evidence these possibilities.

The reported culture medium used to grow *Rhodococcus rhodochrous* ATCC 17895 is considered as undefined. This is due to the fact that it is constituted by complex ingredients such as yeast extract and peptone, which consist of a mixture of several chemical species in unknown proportions, providing vitamins and essential nutrients.

Thus, the proposed investigation for the present thesis will focus on the culture medium in order to study its influence on the Mhy enzymatic activity. In case of a clearly evident influence, it will focus on the optimisation of the culture medium composition in order to enhance the enzymatic activity and improve the old protocol that has been applied in the current work of the isolation of the enzyme.

CHAPTER 2

Design of experiments (DoE)

2.1 Introduction

Statistics play a fundamental role in quality management, allowing a clear planning and evaluation of the characteristics of a certain product as well as the control and improvement of the respective process of production.[13]

Every productive process is nothing but a system in which the inputs are converted in outputs, as described in Figure 2.1. The inputs can be controllable factors such as (in a biochemical processes) temperature, pH, pressure and chemical compounds, or non-controllable factors such as environmental conditions, human error or variability of the feedstock. Generally, the outputs (also called responses) are products whose intrinsic characteristics are known as the *quality characteristics* that must satisfy customers.[13]

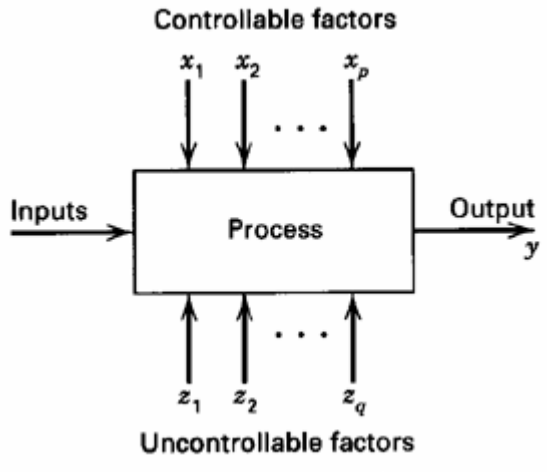


Figure 2.1 Scheme of a productive process.[13]

Among several of the statistical techniques developed or adapted by investigators and quality management professionals, highlight is given to the Design of Experiments (DoE).

DoE is a powerful technique, which is extensively used for engineering problem solving, applied to the collection and interpretation of data with the ultimate goal of generating increased knowledge of an existing process (or explore new processes) and ultimately achieve improvements in product quality and process efficiency.[14] As a set of guidelines in multiple industries, including the biotech industry, it supports an approach to product development that begins with predefined objectives and emphasises product and process understanding and process control, backed by sound science and quality risk management.[15]

The first approaches to this statistical method were started in the UK, 1920s by Sir Ronald. A. Fisher, applied to agriculture. On the following decades progresses were accomplished including a connection to fields of knowledge as algebra and the main investigations were developed in agriculture and chemical industry. Immediately following World War II the first industrial era marked another resurgence in the use of DoE. It was at this time that Box and Wilson (1951) wrote the key paper in response surface designs thinking of the output as a response function and trying to find the optimum conditions for this function. Late 1970s-1990 were marked by quality improvement initiatives in many companies with the Japanese industry in the forefront of two main concepts, a Total Quality Management (TQM) and a Continuous Quality Improvement (CQI), which are management techniques that have come out of this statistical quality revolution. The modern era, beginning circa 1990, when economic competitiveness and globalisation was driving all sectors of the economy to be more competitive.[16]

The DoE emerges as an alternative to the classical One-Variable-At-a-Time (OVAT) strategy in which the experimenter varies one variable at a time while keeping other factors fixed. This approach may produce false/misleading optimum conditions for the process, as it does not account for the possible joint effect of different variables on the process response. Furthermore this inefficient methodology demands the use of a large quantity of resources for incomplete insights on the process and relies strongly on guesswork, experience and intuition that can be fallible even for the most experienced engineer.[14]

2.2 Concepts of DoE methodology

Before introducing the DoE methodology, it is important to describe some important concepts involved:[17]

- **Factors -** Are the process inputs that are thought or known to affect the outputs. They can either be controlled or uncontrolled; (described before, but highlighted here)
- **Responses** – One or more measurable variables that describe the outcome of the process. These are also named as the process outputs; (described before, but highlighted here)

- **Model or Transfer Function -** It is the polynomial equation that fits to the experimental data. The complexity of the model used is dependent on the experimental design and should be determined by statistical significance tests combined with some process understanding;
- **Experimental Design** – It is the set of experiments that are performed. Different designs will fit different models in its complexity;
- **'Levels' of a factor** – Are different values of a factor at which the experiments must be carried out. The factor pH, for example, can be investigated at five levels: 4, 5, 6, 7 and 8 in the optimisation of a spectrophotometric method.
- **Central Points** – These are often the only replicated experiments and are located in the centre of the design. They provide needed information on the process variation and reproducibility.
- **Resolution** – It indicates what type of effects can be studied with each experimental design, depending of the number of factors and runs. Resolution II designs offer some support for linear main effects while resolution V or higher can quantitatively study second degree interactions.
- **Residual (ε)** – Is the difference between the calculated and experimental result for a determinate set of conditions. A good mathematical model fitted to experimental data must present low residuals values.
- **Noise** – It is the set of non-controllable factors that can affect a process, as for instance environmental conditions, human error or variability of the feedstock (described before). Noise is responsible for most of the variability in the response and the DoE increases the robustness of the response with regard to the noise.

2.3 Process models for DoE

A process model of the 'black box' type described in Figure 2.1 , with several discrete or continuous input controllable factors and one or more measured output responses, is built. The output responses are assumed continuous and the experimental data are used to derive an empirical (approximation) model linking the outputs and inputs.[18]

The most common empirical models fit to the experimental data take either a linear or a quadratic form. For its application, it is necessary that the responses (y) obtained are well fitted to the Equation 2.1,

Equation 2.1 Linear function that describes an empirical model.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon$$

where k is the number of variables, β_0 is the constant term, β_i represents the coefficients of the linear parameters, X_i represents the variables and ε is the residual associated to the experiments (experimental error).[19]

Two-level factorial designs are used in the estimation of first-order effects, but they fail when addition effects, such as second-order effects are significant. So, a central point in two-level factorial designs can be used to evaluate the curvature. The next level of the polynomial model should contain additional terms, which describe the interaction between the different factors. This way, a model for a second-order interaction presents the terms showed in Equation 2.2,

Equation 2.2 First order quadratic model.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{1 \le i \le j}^{k} \beta_{ij} X_i X_j + \varepsilon$$

where β_{ij} represents the coefficients of the quadratic parameter.[19]

In order to determine a critical point (maximum, minimum or saddle), it is necessary for the polynomial function to contain quadratic terms according to the Equation 2.3,

Equation 2.3 Second-order quadratic model.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{1 \le i \le j}^{k} \beta_{ij} X_i X_j + \varepsilon$$

where β_{ii} represents the coefficients of the quadratic parameter. This type of model is typically used in response surface methodology (RSM) *DoE*'s with suspected curvature.[19]

The matrix notation of the model is given in Equation 2.4.[20]

Equation 2.4 Matrix notation of the Model.

$$y = X\beta + \varepsilon$$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & x_{1k} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & x_{nk} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}}_{\varepsilon}$$

The system of equations given above is solved using the method of least squares (MLS). Method of least squares is a multiple regression technique and this method can be summarised as follows.

In MLS, it is assumed that random errors are identically distributed with zero mean and a common unknown variance and they are independent of each other. The criterion for choosing the β_i estimates is that they should minimise the sum of the squares of the residuals (SS_{res}).[20]

- **Evaluation of the fitted model**

The mathematical model found after fitting the function to the data can sometimes not satisfactorily describe the experimental domain studied. The more reliable way to evaluate the quality of the model fitted is by the application of analysis of variance (ANOVA). The central idea of ANOVA is to compare the variation due to the treatment (change in the combination of variable levels) with the variation due to random errors inherent to the measurements of the generated responses.[21] From this comparison, it is possible to evaluate the significance of the regression used to foresee responses considering the sources of experimental variance.

In ANOVA, the evaluation of data set variation is made by studying its dispersion. The evaluation of the deviation (d_i) that each observation (y_i) or its replicates (y_{ij}) present in relation to the media ($\bar{y}$), or, more precisely, the square of this deviation, is presented in Equation 2.5.[19]

Equation 2.5 Deviation from an observation in relation to the media.

$$d_i^2 = \left(y_{ij} - \bar{y}\right)^2$$

The sum of the square for all observation deviations in relation to the media is called the total sum of the squares (SS_{tot}); it can be dismembered in the sum of the squares due to the fitted mathematical model, that is, due to regression (SS_{reg}) and in the sum of the squares due to residuals generated by the model (SS_{res}) as shown in Equation 2.6.[19]

Equation 2.6 Relation between the total sum of squares, sum of the squares due to regression and sum of the squares due to residuals.

$$SS_{tot} = SS_{reg} + SS_{res}$$

As replicates of the central point are made, it is possible to estimate the pure error associated with repetitions. Thus, the sum of the square for residuals can be dismembered into two more parcels: the sum of the square due to pure (SS_{pe}) error and the sum of the square due the lack of fit (SS_{lof}) as shown in Equation 2.7.[19]

Equation 2.7 Relation between the sum of the square due to residuals, sum of square due to pure error and the sum of the square due to lack of fit.

$$SS_{res} = SS_{pe} + SS_{lof}$$

When the division of the sum of the square for each source of variation (total, regression, residual, lack of fit, and pure error) is made by its respective numbers of degrees of freedom (df), the ‘mean of the square’ (MS) are obtained. The numbers of degree of freedom for these sources of variation are

calculated by the expressions presented in the third column of Table 2.1, where p represents the number of coefficients of the mathematical model, n represents the number of total observations, and m represents the numbers of levels used in the investigation. Equations related to the source of variations for the calculation of SSs and MSs are also presented in Table 2.1.[22-23]

Table 2.1 Analysis of variance (ANOVA) for fitted mathematical model to an experimental data set using multiple regression.[19, 24]

Variation source	Sum of the square	Degree of freedom (df)	Media of the square
Regression	$SS_{reg} = \sum_{i}^{m}\sum_{j}^{n_1}(\widehat{y_i} - \bar{y})^2$	$p - 1$	$MS_{reg} = \frac{SS_{reg}}{p-1}$
Residuals	$SS_{res} = \sum_{i}^{m}\sum_{j}^{n_1}(y_{ij} - \widehat{y_i})^2$	$n - p$	$MS_{res} = \frac{SS_{reg}}{n-p}$
Lack of fit	$SS_{lof} = \sum_{i}^{m}\sum_{j}^{n_1}(\widehat{y_i} - \bar{y}_i)^2$	$m - p$	$MS_{lof} = \frac{SS_{lof}}{m-p}$
Pure error	$SS_{pe} = \sum_{i}^{m}\sum_{j}^{n_i}(y_{ij} - \bar{y}_i)^2$	$n - m$	$MS_{pe} = \frac{SS_{pe}}{n-p}$
Total	$SS_{tot} = \sum_{i}^{m}\sum_{j}^{n_i}(y_{ij} - \bar{y})^2$	$n - 1$	-

where: n_i, number of observations; m, total number of levels in the design; p, number of parameter of model; $\hat{y}_i$ estimated value by the model for the level i; $\bar{y}$, overall media; y_{ij} replicates performed in each individual levels; , $\bar{y}_i$ media of replicates performed in the same set of experimental conditions.

The significance of regression can be evaluated by the ratio between the mean of the square of regression (MS_{reg}) and the media of the square of residuals (MS_{res}) and by comparing these variation sources using the Fisher distribution (F-test), taking into account its respective degrees of freedom associated to regression (df_{reg}) and to residual (df_{res}) variances, as displayed in Equation 2.8.[19]

Equation 2.8 Comparison between the ration of the mean of the square of regression/mean of the square of residuals with the F-test taking into account the degrees of freedom of the regression and of the residuals.

$$\frac{MS_{reg}}{MS_{res}} \approx F_{df_{reg},\, df_{res}}$$

Thus, a statistically significant value for this ratio must be higher than the tabulated value for F. This is an indication that the mathematical model is well fitted to the experimental data. Another way to evaluate the model is the 'Lack of fit test'. If the mathematical model is well fitted to the experimental data, MS_{lof} should reflect only the random errors inherent to the system. Additionally, MS_{pe} is also an estimate of these random errors, and it is assumed that these two values are not statistically different. This is the key idea of the lack of fit test. It is possible to use the F distribution to evaluate if there is some statistical difference between these two media, in the same way that the significance of regression was verified. Equation 2.9 displays this relation,

Equation 2.9 Comparison between the ratio of the mean of the square of lack of fit/mean of the square of pure error with the F-test taking into account the degrees of freedom of the lack of fit and of the pure error.

$$\frac{MS_{lof}}{MS_{pe}} \approx F_{df_{lof},\, df_{pe}}$$

where, df_{lof} and df_{pe} are, respectively, the degree of freedom associated with the lack of fit and the pure error variances. If this ratio is higher than the tabulated value of F, it is concluded that there is evidence of a lack of fit and that the model needs to be improved. However, if the value is lower than the tabulated value, the model fitness can be considered satisfactory. To apply a lack of fit test, the experimental design must be performed with authentic repetitions at least in its central point.[19]

In short, a model will be well fitted to the experimental data if it presents a significant regression and a non-significant lack of fit. In other words, the major part of variation observation must be described by the equation of regression, and the remainder of the variation will certainly be due to the residuals. Most variation related to residuals is due to pure error (random fluctuation of measurements) and not to the lack of fit, which is directly related to the model quality.[19]

The visual inspection of the residual graphs can also generate valuable information about the model suitability. Thus, if the mathematical model is well fitted, its graph of residuals presents a behaviour that suggests a normal distribution. If the model generates larger residuals, it is not adequate to make precise inferences about the data behaviour in the studied experimental area. Moreover, if the model needs some other term, the residual graph will present a behaviour that indicates the kind of term that must be added to the model.[23]

2.4 Optimisation using DoE: response surface methodology (RSM)

2.4.1 Literature review

Response surface methodology (RSM) is the most popular DoE optimisation method used in recent years.[20] Some examples of the RSM applications performed for optimisation of biochemical processes involving enzymes are hydrolysis of pectic substrates, enzymatic synthesis of fatty esters, lipase-catalysed incorporation of docosahexanenoic acid (DHA) into borage oil, alkaline protease production from *Bacillus mojavensis* in a bioreactor, butylgalactoside synthesis by-galactosidase from *Aspergillus oryzae*, biotransformation of 2-phenyllethanol to phenylacetaldehyde in a two-phase fed-batch system, lipase catalysed esterification reactions, pectinase usage in pretreatment of mosambi juice for clarification, phytase production by *Pichia anomala* and determination of reaction parameters for damaged starch assay.[25-33]

Besides that, RSM is also the most widely used method in culture media optimisation for the synthesis of a specific product or for enhancement of enzymatic activity from a microorganism. Optimisation of a fermentation medium for stereoinversion of (*S*)-1-phenyl-1,2-ethanediol from

Candida parapsilosis CCTCC M203011, optimisation of culture conditions for the production of xylanase in submerge fermentation by *Penicillium citrinum*, optimisation of critical medium components for phenazine-1-carboxylic acid production by *Pseudomonas* sp. M-18Q, optimisation of α-amylase production by *Bacillus* sp., media optimisation for biosurfactant production by *Rhodococcus erythropolis* MTCC 2794 and cholesterol oxidase production by *Rhodococcusequi* no. 23 are a few examples.[34-39]

The ability to search for an optimum condition from a relatively small number of experiments and the ability to interpret the interactive effects among input factors are some attractive features of RSM. One drawback of RSM is that it is mainly restricted to quadratic nonlinear correlation, whereas biological process may show more complex nonlinear dependencies.[38]

2.4.2 RSM methodology

In bioprocessing, RSM methodology can be applied in three stages: Screening, which involves identification of factors that have significant effect on the process; Optimisation, that involves prediction of the response surfaces and finding the optimal set points (i.e. optimal operating conditions); and Robustness Testing that verifies the process response to small variations in the input parameters (this test was not applied in the present work).[17]

The starting point for the selection of the type of design is the definition of the experimental objective (response to be maximised or minimised), selection of factors that affect the response and range for their variation (levels), in which a 'level' is referred to a specific setting of the factor being tested and a 'run' is a the combination of factor levels to whose effect on the response we want to assess. The selection of what runs to test, within the range of the selected factors, depends on the type of design employed.

Some computer packages offer optimal designs based on the special criteria and input from the user. These designs differ from one other with respect to their selection of experimental points, number of runs and blocks. After selection of the design, the model equation is defined and coefficients of the model equation are predicted. The model used in RSM is generally a full quadratic equation as Equation 2.3 or the diminished form of this equation.

- **Screening**

Chemical and biochemical processes are affected by numerous parameters. Because it is not possible to identify the effects of all parameters, it is necessary to select the parameters that have major effects. Screening experiments are useful to identify the independent parameters. Factorial designs may be used for this purpose. After identification of the important parameters, the direction in which improvements lie is determined and the levels of the parameters are identified.[20]

To investigate a small number of process inputs (nº of $X < 5$), full factorial designs are the most common choice. In this type of designs, the number of experiments increases rapidly with the number of input factors (k). Equation 2.10 displays this relation.[40]

Equation 2.10 Relation between the number of experiments and the number of input factors and central points added.

$$nº\ of\ experiments = 2^k + nº\ central\ points$$

As each factor is analysed in its highest and lowest limit, the design presents the 0 levels (central points).Figure 2.2 exhibits the experimental designs for 2 and 3 factors using a full factorial design. This type of designs offers good support for linear effects and all interaction effects but cannot explain non-linear cause effect responses.[40]

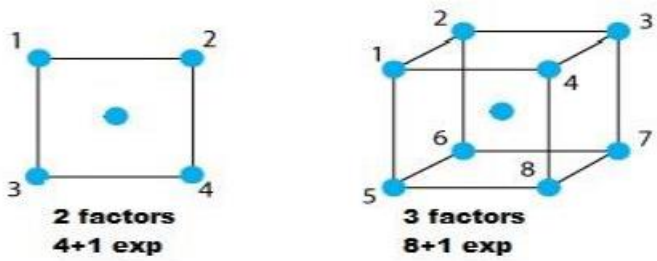


Figure 2.2 Full factorial experimental design with 2 input factors and 3 input factors. The number of central points in the shown designs is 1.

To study processes with more than five factors, fractional factorial (FF) designs are often used. When studying processes with more than 3 factors it becomes more complex and not easily describable as the dimensions required for the visualisation equals the number of factors. A fractional factorial design still allows for the identification of main effects without acquiring the detailed information provided by a full factorial design, therefore the non-linear and interaction effects cannot be assessed properly. Given its advantages and limitations, fractional factorial design can be useful in robustness tests or initial screening of many factors using relatively few experiments.[18, 40]

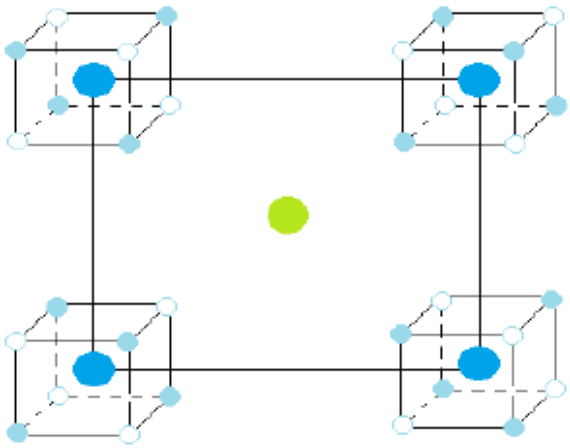


Figure 2.3 Fractional factorial experimental design with 5 input factors with one central point.

- **Optimisation**

In order to add interaction effects and the second degree effects, which allows for the modelling of a full response surface, additional data points are required to quantify the curvature. By adding these extra points to full factorial designs or fractional factorial designs, central composite designs are obtained. (Figure 2.4)[40]

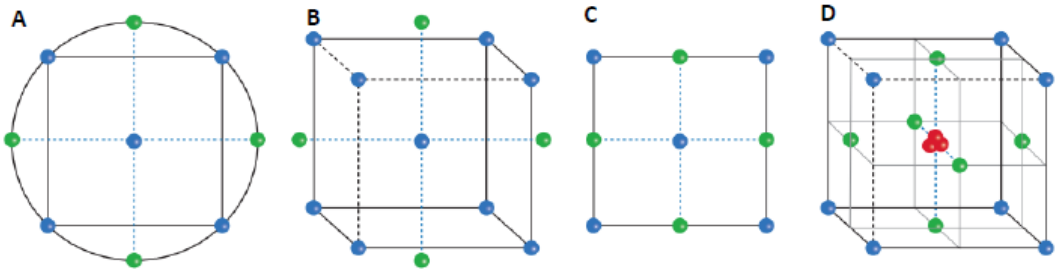


Figure 2.4 Central composite experimental designs: (A) Two factor central composite circumscribed (CCC); (B) Three factor CCC; (C) Two factor composite face-centred (CCF) design; (D) Three factor CCF design.

In each figure it is possible to see the factorial design points at blue, topped with the "star" points at green which are points that allow to study the second degree response effects as a result of the variation of one factor, always passing through the central points.

The response is usually represented graphically in three-dimensions or through a X_i, X_j plane with drawn curves of constant response, called contour plot, as displayed in Figure 2.5. This allows a visualisation of the response surface and its shape.[14, 18, 41]

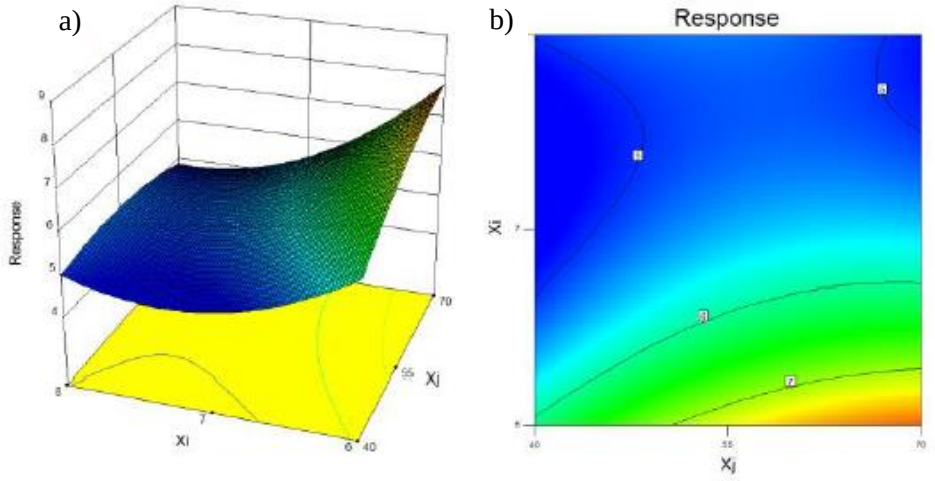


Figure 2.5 Generic example of Response Surface visualisation: a) Three-dimensional graph; b) Contour plot.

The 3D shows the main interaction, confirmed by an ANOVA, between the variables X_i and X_j, and its influence on the response.

Thus, the optimisation is performed when finding optimal X_i and X_j values that maximise or minimise the response, according to the purpose of the optimisation.

CHAPTER 3

Materials and methods

3.1 Chemicals

Hydroxyacetone was purchased from Alfa Aesar. Peptone and yeast extract were bought from BD Becton, Dickinson and Company. Sodium hydroxide was purchased from J.T. Baker while glucose was bought from Fluka Analytical. Petroleum ether (PE) was purchased from VWR International and distilled before utilisation. All the other commercial chemicals not mentioned were purchased from Sigma Aldrich.

3.2 Analytical methods

3.2.1 Gas chromatography

GC (Gas chromatography) analysis of the bioconversions of *(E)*-4-hydroxy-3-methylbut-2-enoic acid were performed with a Shimadzu type GC-2010 Plus equipped with a CP Wax 52 CB column (50 m x 0.53 mm x 2.0 μm) using N_2 as carrier gas. The following conditions were used for the nonchiral separation using direct injection: injector 280 °C, detector (FID - Flame ionisation detector) 300 °C, column flow rate 1.43 mL/min, linear velocity 37.0 cm/sec, pressure 98.5 kPa, temperature program: start at 90 °C, hold time 3 min, rate 5 °C/min to 250 °C hold time 1 min. The retention times of 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one **2** and 4-methylfuran-2(5H)-one **7** are shown in one example gas chromatogram in Appendix A.

3.2.2 NMR spectroscopy

^{1}H and ^{13}C NMR (Nuclear magnetic resonance) spectra were recorded on an Agilent 400 (400 MHz and 100 MHz respectively) instrument and were internally referenced to residual solvent signals. Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = duplet, dd = double duplet, t = triplet, m = multiplet), integration and assignment. Data for ^{13}C NMR are reported in terms of chemical shift.

3.3 Statistical software

The design of experiments (DoE) for the screening experiments and for the response surface methodology experiments were developed using the software Design Expert ver. 7.0.0 (Stat-Ease Inc, Minneapolis, MN).

For both designs, the concentrations of product (g/L) were used as responses which were analysed using the analysis of variance (ANOVA) combined with the Fisher's test to evaluate if a given variable has a significant effect for a confident interval of 5 % ($p < 0.05$).

3.3.1 Screening experiments

A 2^{4-1} factorial fractional design with 1 generator (p=1), resolution IV, was used in order to study the effect of four numeric factors: glucose (A), peptone (B), yeast extract (C) and magnesium sulphate heptahydrate (D) as well as their respective interactions. Using one replicate, one block and four central points, a matrix of 12 experiments (eight factorial points plus four central points) was developed, choosing empirically numeric values for the two levels (+, -) of the four factors (A, B, C, D). The central points contain the factors at reference levels (reference culture medium composition). The factorial points were randomised by the software, generating together with the central points a matrix of experiments. Table 3.1 shows the matrix of experiments in terms of coded factors (-, 0, +). The compositions of each culture media are specified on the subchapter "Composition of the culture medium".

Table 3.1 Matrix of experiments in terms of coded factors for the screening experiments.[24]

	Glucose A	**Yeast extract B**	**Peptone C**	**$MgSO_4$ * 7 H_20 D**
1	0	0	0	0
2	+	-	-	+
3	-	-	-	-
4	0	0	0	0
5	0	0	0	0
6	+	+	-	-
7	+	+	+	+
8	+	-	+	-
9	-	+	+	-
10	-	-	+	+
11	-	+	-	+
12	0	0	0	0

Level '+' represents the high level, which means that for a given variable, a value above the standard is used for the composition of the culture medium. Level '-' represents the low level, which

means that for a given variable, a value below the standard is used for the composition of the culture medium. The rows with all variables at level '0' represent the reference culture medium composition, described in section 3.4.1.

3.3.2 Response surface methodology experiments

A full circumscribed central composite design (CCD) was developed. The numeric factors chosen were glucose (A), peptone (B), yeast extract (C). Each numeric factor was varied over 5 levels: plus and minus alpha (axial points), plus and minus (factorial points) and a central point. The total number of runs were 20, consisting of six repeated central points, six axial points and eight factorial points. As well as in the screening experiments, the numeric values for the levels of the factorial points and axial points were chosen empirically. Also the central points contain the factors at reference levels (reference culture medium composition) and the factorial points were randomised by the software generating together with the axial points and the central points a matrix of experiments. Table 3.2 displays the matrix of experiments in terms of coded factors (-α, -, 0, +, +α). The compositions of each culture media are quantified on the subchapter Composition of the culture medium.

Table 3.2 Matrix of experiments in terms of coded factors for the response surface methodology experiments.[24]

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Glucose (g/L)	0	0	+α	-	-	+	+	0	-	0	0	+	0	0	0	0	+	-α	0	-
Yeast extract (g/L)	0	0	0	-	+	-	+	0	-	+α	0	-	0	-α	0	0	+	0	0	+
Peptone (g/L)	0	0	0	-	-	+	-	+α	+	0	0	-	- α	0	0	0	+	0	0	+

Levels '+' and '-' are the high and low levels respectively, of the factorial points. Levels '+α' and '-α' are the high and low levels, respectively, of the axial points. For a given variable, level '0' represents the value of the standard culture medium, described in section 3.4.1.

3.4 Microorganisms and culture conditions

The strain *Rhodococcus rhodochrous* ATCC 17895 was purchased as a lyophilised culture from ATCC (American Type Culture Collection, Manassas, USA). Glycerol and DMSO (Dimethyl sulfoxide) stocks of *R. rhodochrous* were stored at -80 °C.

For each cultivation the stock was plated on agar plates containing Nutrient Agar (BD 213000) (23 g/L) autoclaved at 121 °C. The plate was incubated for 72 hours at 26 °C.

3.4.1 Composition of the culture medium

The composition of the culture medium was changed taking into account the aim of this project, so it is stated in this subchapter the initial (standard) conditions reported in the literature, as well as the changed conditions used for the different experiments developed.[12] The following incubation conditions were maintained for all the experiments subsequently described: growth time of 72 h, 180 rpm of orbital shaking and temperature of 26 ºC.

- **Literature (standard) culture medium[12]**

The standard culture medium used for cultivation contained: Solution A (980 mL DI water) with glucose (15 g), peptone (5 g), yeast extract (1 g), 7 mM K_2HPO_4 (1.2 g), 3 mM KH_2PO_4 (0.4 g) with a final pH of 7.2 and autoclaved at 110 °C; Solution B (10 mL DI water) with 406 mM $MgSO_4 * 7\ H_2O$ (1 g) and filter sterilised; Solution C (10 mL DI) with 198 mM $FeSO_4 * 7\ H_2O$ (0.5 g) and filter sterilised.[12] Solutions A, B and C were mixed before inoculation in 5 L Erlenmeyer flasks. For inoculation in 2 L Erlenmeyer flasks, solutions A, B and C were halved.

Overnight, a pre-culture consisting of solutions A, B, C (50 mL, 500 µL, and 500 µL, respectively) and a pipet tip of the plated bacteria, was incubated. Afterwards, 2 mL of the pre-culture was added to the culture medium in 5 L Erlenmeyer flasks and/or halved for inoculation in 2 L Erlenmeyer flasks.

- **Experiments with different concentrations of $FeSO_4 * 7\ H_2O$**

In this experiment, the iron sulphate concentration was changed, keeping the pre-culture unchanged. Four different culture media were prepared by changing the solution C. Table 3.3 displays the concentration of $FeSO_4 * 7\ H_20$ in the different culture media.

Table 3.3 Different culture media according to different concentrations (mM) of iron sulphate.

Culture Media	Concentration of $FeSO_4 * 7 H_2O$ on the culture medium (mM)
1	0
2	99
3	198
4	396

- **Experiments without $FeSO_4 * 7 H_20$**

The total absence of iron sulphate heptahydrate in the pre-culture as well as in the main culture medium was also studied. In this case, no solution C was required and DI water was added in replacement to make up the volume of both (pre-) and main culture medium described before.

- **Screening experiments**

The composition of the culture medium was changed according to the statistical optimisation described in **3.3.1**. Furthermore, the iron sulphate heptahydrate (Solution C) was not used in these experiments on both (pre-) and main culture media. The same quantity in DI water was therefore added to solution A in replacement, to reach the final volume of the (pre-) and main culture media described in the literature culture medium.

Thus, twelve culture media were prepared varying the compositions in glucose, yeast extract, peptone and magnesium sulphate heptahydrate according to the statistical software output.

Table 3.4 displays the composition of solution A and solution B of each culture media. Culture media numbers 1, 4, 5 and 12 have equal compositions, corresponding to a reference culture medium (central points). The reference culture medium has the same composition of the standard culture medium from literature, except the fact that it does not have solution C.

The composition in phosphate buffer salts (KH_2PO_4 and K_2HPO_4) was maintained equal to the standard culture medium for all the culture media prepared.

Table 3.4 Culture media composition for the 2^3 factorial fraction design, screening experiments.[24]

	Solution A			Solution B
	Glucose (g/L)	**Yeast extract (g/L)**	**Peptone (g/L)**	**$MgSO_4$ * 7 H_20 (mM)**
1	15	1	5	406
2	30	0.5	2.5	812
3	7.5	0.5	2.5	0
4	15	1	5	406
5	15	1	5	406
6	30	2	2.5	0
7	30	2	10	812
8	30	0.5	10	0
9	7.5	2	10	0
10	7.5	0.5	10	812
11	7.5	2	2.5	812
12	15	1	5	406

Compositions randomised by the statistical software except for the central points (reference culture media composition). For further information of this method, check section 3.3.

- **Response surface methodology experiment**

In this batch of experiments, twenty culture media were prepared according to the output of the statistical software.

Table 3.5 shows the composition on glucose, yeast extract and peptone of the twenty culture media prepared. Culture media numbers 1, 2, 11, 15, 16 and 19 have equal compositions, corresponding to a reference culture medium (central points); culture media numbers 3, 8, 10, 13, 14 and 18 are axial points and the remaining 8 culture media are factorial points.

Table 3.5 Culture media composition for the central composite design, response surface

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Glucose (g/L)	15	15	23.41	10	10	20	20	15	10	15	15	20	15	15	15	15	20	6.59	15	10
Yeast extract (g/L)	1	1	1	0.5	1.5	0.5	1.5	1	0.5	1.84	1	0.5	1	0.16	1	1	1.5	15	1	1.5
Peptone (g/L)	5	5	5	2.5	2.5	7.5	2.5	9.20	7.5	5	5	2.5	0.80	5	5	5	7.5	1	5	7.5

Compositions of factorial points were randomised by the statistical software. For further information of this method, check section 3.3.

These experiments were developed in the absence of both magnesium sulphate heptahydrate and iron sulphate heptahydrate on both (pre-) and main culture media. Therefore, solutions B and C described on the literature standard culture medium were substituted by the same quantity in DI water respectively, to make up the volume of the (pre-) and main culture media. The culture medium used as reference has the same composition of the literature standard culture medium, except the fact that it does not have solutions B and C.

Similarly to the screening experiments, the composition in KH_2PO_4 and K_2HPO_4 was kept equal to the standard.

- **Confirmation experiments**

A third batch of experiments regarding the change on the culture medium composition was developed in order to confirm the results predicted by the response surface methodology experiment, therefore validating the mathematical model estimated (for further information, check the discussed results in chapter 4).

Table 3.6 exhibits the culture media composition. For all culture media, the composition in KH_2PO_4 and K_2HPO_4 was kept equal to the standard and the iron sulphate heptahydrate (solution C) was not used. In replacement, the same quantity in DI water was added to make up the volume of the (pre-) and main culture media.

Table 3.6 Culture media composition for the confirmation experiment. The * means that solution B was only used on the pre-culture medium.

	Solution A			Solution B
	Glucose (g/L)	Yeast extract (g/L)	Peptone (g/L)	$MgSO_4 * 7 H_2O$ (mM / g in sol. B)
1	15	1	5	406
2	15	1	5	406*
3	15	1	5	0
4	7,5	2	10	406*
5	6.59	0.6	0.8	0
6	6.59	1.84	9.2	0

Culture medium 1, which has the same composition of the standard except the absence of solution C, was used as reference. Culture media 2 and 4 only have solution B on the pre-culture medium. Culture media numbers 3, 5 and 6 do not have solution B on both (pre-) and main culture media, and the same quantity in DI water was added in replacement.

3.5 General chemical synthesis procedures

3.5.1 Synthesis of ethyl-4-hydroxy-3-methylbut-2-enoate 5

Triphenylcarbethoxymethylenephosphorane (4**)** (16 g, 45.8 mmol) was dissolved in 100 mL of toluene at 75 ºC. Hydroxyacetone (3**)** (4 g, 54.0 mmol) was added dropwise and the solution was refluxed for 4 hours at 120 ºC. The solvent was afterwards evaporated under reduced pressure and the yellow solution was purified by column chromatography on silica gel with PE/EtOAc (Ethyl acetate) 1:1 as eluent. The fractions containing the product were collected and evaporated under reduced pressure to give a yellowish oil **5** (4.07 g, 28.2 mmol, 61.6 %) ^{1}H NMR (400 MHz, $CDCl_3$) δ: 1.26 (t, 3H), 2.07 (s, 3H), 4.11 (m, 2H), 4.15 (q, 2H), 5.96 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 14.3, 15.6, 59.7, 67.1, 113.7, 157.1, 166.8. In accordance with literature.[11] Figure 3.1 displays a scheme of the reaction. ^{1}H and ^{13}C NMR spectra are displayed in figures B.1 and B.2, respectively, in Appendix B.

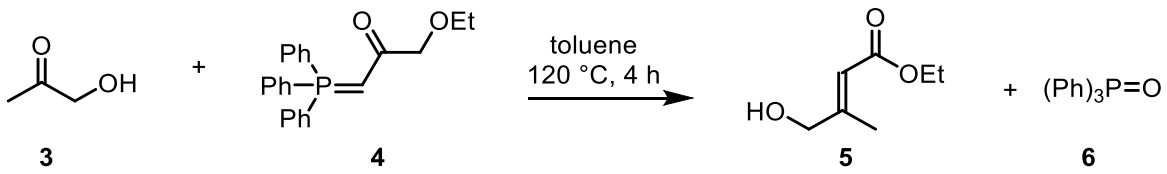


Figure 3.1 Wittig reaction of triphenylcarbethoxymethylenephosphorane (4) with hydroxyacetone (3) to give *(E)*-ethyl-4-hydroxy-3-methylbut-2-enoate (5) and triphenylphosphite (6).

3.5.2 Synthesis of 4-hydroxy-3-methylbut-2-enoic acid 1

(*E*)-ethyl-4-hydroxy-3-methylbut-2-enoate (5) (4,07 g, 28.2 mmol) was dissolved in 28 mL of methanol and 5% aqueous NaOH (35 mL) was added dropwise at 0 °C. The reaction took place overnight at room temperature. Afterwards the solution was diluted with water (56 mL) and washed twice with MTBE (2 x 28 mL). The aqueous layer was then acidified to pH 1 with conc. HCl, saturated with NaCl (1 M) and extracted with ethyl acetate (3 x 50 mL). The collected organic phases were dried over Na_2SO_4, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography on silica gel with PE/EtOAc 1:2 as eluent. The fractions containing the product were collected and evaporated under reduced pressure to give a white powder **1**. (1.33 g, 11.45 mmol, 43.4 %). ^{1}H NMR (400 MHz, DMSO-d_6) δ: 1.95 (s, 3H), 3.92 (s, 2H), 5.84 (s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 15.2, 65.2, 113.0. 158.4, 167.6. In accordance with literature.[11] Figure 3.2 exhibits the hydrolysis reaction scheme. ^{1}H and ^{13}C NMR spectra are displayed in figures B.3 and B.4, respectively, in Appendix B.

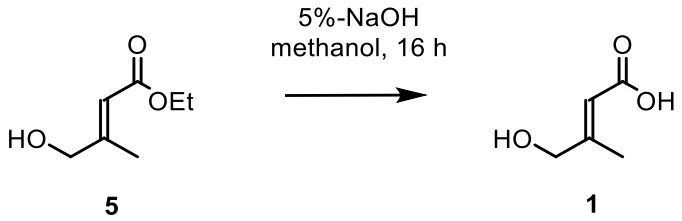


Figure 3.2 Hydrolysis reaction of (*E*)-ethyl-4-hydroxy-3-methylbut-2-enoate (5) to give (*E*)-4-hydroxy-3-methylbut-2-enoic acid (1).

3.6 General biotransformation procedure

The biotransformation procedure was the same for all the experiments developed. After the cultivation time of 72 h, the cells of *R. rhodochrous* ATCC 17895 were harvested by centrifugation (10 000 rpm, 17 696 rcf, 15 min) and washed twice with 100 mM potassium phosphate buffer (pH 6.2). The cells were re-suspended in the same buffer to a final cell content of 100 mg/mL. In 2 mL Eppendorf vials containing 900 μL of the cells re-suspended in buffer, substrate was added to a final concentration of 10 mM. Blank reactions were also performed using buffer instead of whole cells solution. All the reactions were carried out in triplicate, shaking at 26 ºC with 1400 rpm, for 24 hours. Afterwards, the reaction was stopped and the compounds were extracted twice with ethyl acetate (2 x 500 μL) by centrifugation (5 min, 13000 rpm). The combined organic phases were dried over Na_2SO_4 and measured with GC. Figure 3.3 presents the biotransformation reaction scheme, catalysed by whole cells of *R. rhodochrous* ATCC 17895.

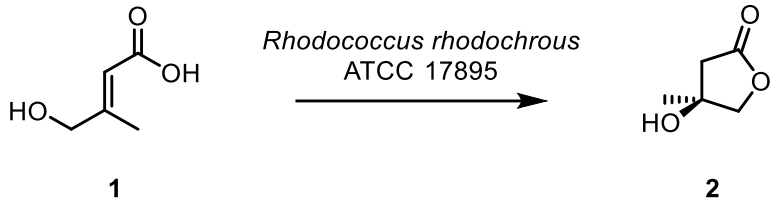


Figure 3.3 Reaction equation of the addition of water catalysed by Michael hydratase present in *Rhodococcus rhodochrous* ATCC 17895 using (*E*)-4-hydroxy-3-methylbut-2-enoic acid (1) as model substrate to give the product 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one (2).

3.7 Preparative scale biotransformation

A large scale biotransformation has been carried out to obtain a sufficient amount of bioconversion product to analyse the extraction efficiency in EtOAc as well as to run a calibration curve. The cells were grown in (literature) standard culture medium. The results are discussed in chapter 4.

Cells of ATCC 17895 were harvested by centrifugation (10 000 rpm, 17 696 rcf, 15 min) and washed twice with 100 mM potassium phosphate buffer (pH 6.2) ; 56.85 g of cells were re-suspended in the same buffer to a final concentration of 100 mg/mL and 15 mM of substrate was added. The biotransformation took place in a 2 L flask shaking at 26 ºC with 180 rpm for 96 hours. Afterwards the solution was centrifuged (10 000 rpm, 17 696 rcf, 15 min) and the collected supernatant was continuously extracted with DCM (Dichloromethane) for 72 hours, using a liquid-liquid extractor (pictures in Appendix C). The collected organic phase was concentrated under reduced pressure and purified by column chromatography on silica gel with heptane/ethyl acetate 1:1 as eluent. The purified compound was again concentrated under vacuum to give a colourless oil (32.3 mg, 0.28 mmol, 4.92 %). GC retention time: 27.41 min. Reference compound NMR data: ^{1}H NMR (400 MHz, $CDCl_3$) δ: 1.46 (s, 3H), 2.51-2.63 (ABq, 2H), 3.28 (s, 1H), 4.09-4.24 (ABq 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 24.6, 43.4, 74.4, 79.7, 176.6. In accordance with literature.[11] ^{1}H and ^{13}C NMR spectra are displayed in figures B.5 and B.6, respectively, in Appendix B.

A calibration line in triplicate was developed analysing samples with different concentrations of the product on the GC, and thus correlate the concentration with the resulting product peak areas from the chromatograms. Table 3.7 displays the dilutions prepared in 1 mL of ethyl acetate, each.

Table 3.7 Concentrations of product used on the dilutions and each quantity of product required. Each dilution was performed in 1 mL of ethyl acetate, using proper GC vials.

Concentration of product (mM)	Quantity of product (mg)
0.5	0.05
1	0.10
2	0.20
4	0.41
6	0.61
8	0.82
10	1.02

The extraction efficiency tests were developed in triplicate using product concentrations of 2 mM, 6 mM and 10 mM, diluted in potassium phosphate buffer (pH 6.2) and extracted twice with ethyl acetate (2 x 500 µL) by centrifugation (5 min, 13000 rpm). The combined organic phases were dried over Na_2SO_4 and measured with GC.

3.8 Fermentor

The growth of the microorganism was tested for the Mhy activity in a 15 L fermentor, with 10 L of culture medium prepared accordingly. Due to the fact that this experiment was carried out after the screening experiments but before the response surface methodology experiments, it was decided to use the best culture medium composition found in the screening experiments, which is the culture medium number 10 described in 4 (for further information check the results and discussion in chapter 4). Flasks (500 mL) of 2 M sulphuric acid, 2 M potassium hydroxide, and antifoam were prepared and autoclaved. The reactor was autoclaved together with the culture medium and all the pieces connected. Pre-culture (according to the standard culture medium but without solution C) was prepared in a mass ratio of 10% of the culture medium used and inoculated. Temperature was set to 26 ºC, pH to 6.70 and mixing speed of 300 rpm; an air flow of 5 slpm was kept. Samples were taken regularly to analyse the OD_{600} (Optical density, 600 nm), and after 64 h the growth was stopped. The grown cells were tested for biotransformation following the general biotransformation procedure described in section 3.6. Results from this experiment are discussed in section 4.6.

CHAPTER 4

Results and discussion

4.1 Calibration curve and extraction efficiency tests

Following the protocol described in section 3.7, the resulting product peak areas from the chromatograms (data not shown) were plotted against the concentrations corresponding to each dilution, as shown in Figure 4.1. Since the calibration was performed in triplicate, each dot is the average of the triplicate with the correspondent standard deviation errors.

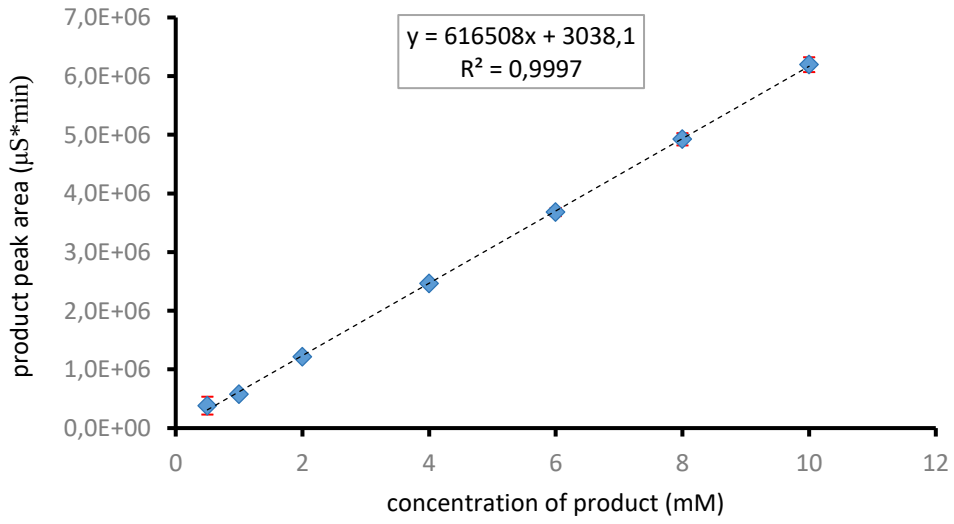


Figure 4.1 Plot of the product peak area in function of the concentration of product.

A virtually perfect linear relation with a high value of R^2 was built. Hereupon, Equation 4.1 specifies the concentration of product in function of the product peak area (inverse function of the observed in figure 4.1).

Equation 4.1 Concentration of product in function of the product peak area

$$c_{prod.} = \frac{prod.\ peak\ area - 3038.1}{616508}$$

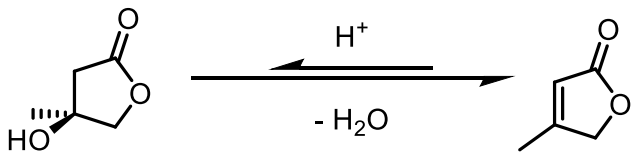
The analysed product peak areas from the extraction efficiency tests (chromatogram data not shown) were compared with the product peak areas from the calibration curve, for the same concentrations. The efficiency of each extraction was therefore calculated dividing the values of the product peak areas after extraction by the values of the product peak areas from the calibration curve and shown in percentage in Table 4.1.

Table 4.1 Concentration of product, the correspondent product peak areas (after extraction and from the calibration curve), and the extraction efficiency of the product for each case.

Concentration of product (mM)	Calibration curve product peak area (µS*min)	After extraction product peak area (µS*min)	Extraction Efficiency (%)
2	1.2E+06	2.4E+05	20.1
6	3.7E+06	7.6E+05	20.7
10	6.2E+06	1.3E+06	21.0

An average of the extraction efficiency was calculated as 20.6 %, having this value been assumed for every extraction that followed the protocol described in section 3.6. This result evidences that most of the product is held in the aqueous phase due to its high hydrophilicity.

The extraction efficiency of a liquid-liquid extraction is determined by the equilibrium constant of the solute's partitioning between the aqueous and organic phases as well as any other reactions involving the solute.[42] A known equilibrium reaction affecting indirectly the extraction of the product 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one **2** in the aqueous phase is the dehydration displayed in Figure 4.2 , yielding 4-methylfuran-2(*5H*)-one **7**. The presence of a smaller peak, which has already been proven to be correspondent to the compound **7**, is always observable in the chromatograms (unpublished data).



2 **7**

Figure 4.2 Equilibrium hydration and dehydration reactions between hydroxy-4-methyldihydrofuran-2(*3H*)-one (2) and 4-methylfuran-2(*5H*)-one (7).

Since the partitioning between the two phases is not the only reaction affecting the extraction efficiency, the calculation of a partitioning coefficient involves a complex algebraic relationship that, along with in-depth studies of the liquid-liquid extraction, were not studied in this project.[42]

In this section, the main purpose of the calculation of the extraction efficiency is to obtain the total concentration of biotransformation product, i.e., the concentration of product present in the aqueous

phase before the liquid-liquid extraction step. Thus, relating the extraction efficiency value with the Equation 4.1, the Equation 4.2 is built.

Equation 4.2 Total concentration of biotransformation product formed in each reaction in function of the analysed product peak area from a chromatogram (mM).

$$c_{prod.\ total} = \frac{(\frac{prod.\ peak\ area}{0.206}) - 3038.1}{616508}$$

Having the concentration of the biotransformation product and the concentration of the limiting reagent, the yield of the reaction can be calculated. Being the water both the solvent and one of the reagents, it is in excess. Thus, the limiting reagent is the substrate (*E*)-4-hydroxy-3-methylbut-2-enoic acid. This reagent has the initial concentration of 10 mM in every biotransformation, since the procedure described in section 3.6 was the same for all the experiments performed. Therefore, Equation 4.3 displays the yield of the reaction.

Equation 4.3 Yield of the biotransformation following the protocol described in 3.6.

$$Yield\ (\%) = \frac{c_{prod.total}}{10} * 100\ \%$$

In the scope of this project it is essential to check either the total concentration of biotransformation product formed or the yield of each reaction in order to, indirectly, evaluate the enzymatic activity of the Michael hydratase present in the whole cells of *Rhodococcus rhodochrous* ATCC 17895. Hence, the influence of the composition of the culture medium on the Mhy activity is evaluated in every experiment by means of the scheme described in Figure 4.3.

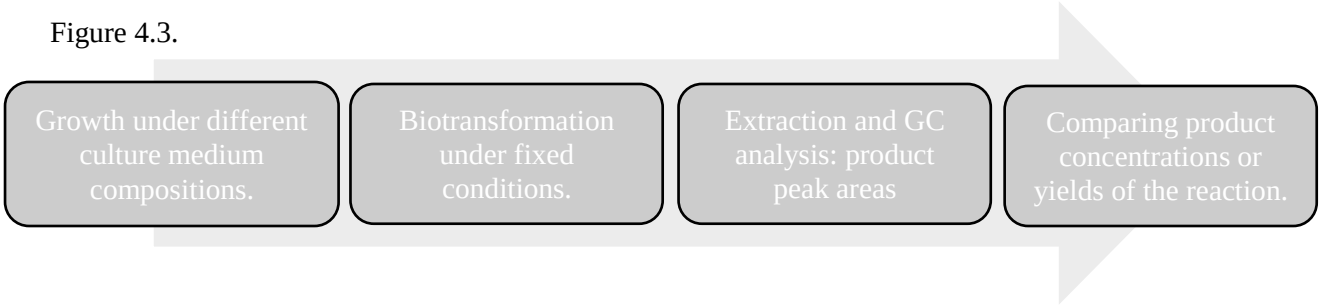


Figure 4.3 Scheme of the steps used to study the influence of the composition of culture medium on the enzymatic activity of the Mhy.

In short, after growing under different culture medium compositions the whole cells of *R. rhodochrous* ATCC 17895 are tested under fixed biotransformation conditions and the product is extracted and analysed with GC (section 3.6). Afterwards, the product peak areas analysed in the chromatograms can be converted to the concentrations of product using Equation 4.2 and the yields of the reaction can be calculated using Equation 4.3.

4.2 OVAT for the study of the influence of $FeSO_4 * 7 H_2O$

4.2.1 Experiments with different concentrations of $FeSO_4 * 7 H_2O$

The first experiment regarding the composition of the culture medium consisted in varying the concentration of iron sulphate heptahydrate while keeping the remaining composition unchanged. A classical One-Variable-At-a-Time (OVAT) strategy was applied to check the importance of this compound for the Mhy enzymatic activity since the literature protocol reports its use in unusual high amounts, besides the OD_{600} analysis issues described in section 1.5.

An average of the product concentration results, given by a duplet of standard culture medium, was used to normalise all other product concentrations. An explanation regarding the normalisation is discussed further in this section. Figure 4.4 displays a graphic representation of the normalised product concentration for each culture medium tested with different concentrations of iron sulphate heptahydrate, as described in Table 3.3 from section 3.5.1.

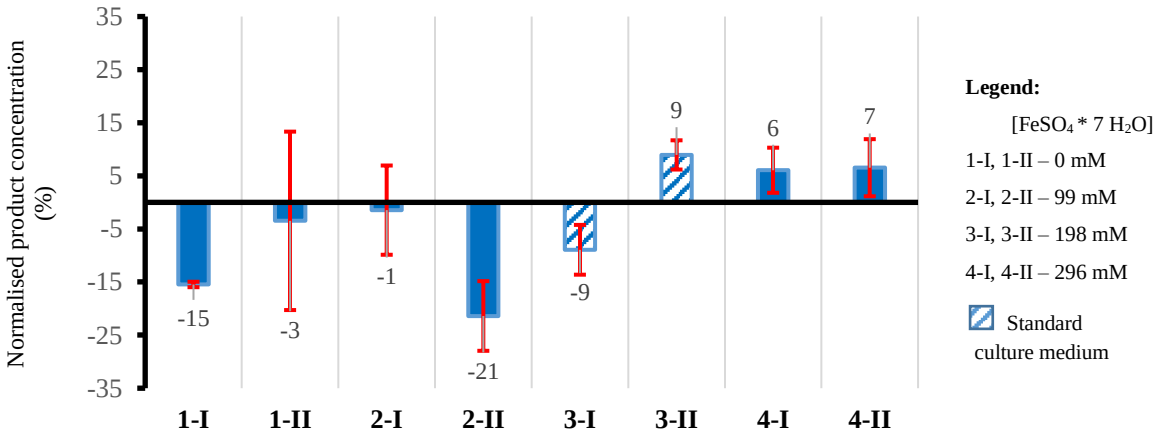


Figure 4.4 Normalised product concentrations (%) obtained under different compositions of $FeSO_4 * 7 H_2O$ in the culture media.

The standard culture media with an iron sulphate concentration of 198 mM are displayed as numbers 3-I and 3-II. Their normalised values have a deviation from the zero of the plot in the same absolute value, 9 %. The remaining culture media with different concentrations of iron sulphate were also studied in duplicate. Each of the eight experiments shows a standard deviation error since it is an average of a triplicate (all biotransformations were carried out in triplicate as described in section 3.6).

Despite an increase on the concentration of product **2** when the iron sulphate concentration is doubled (media numbers 4-I and 4-II), the standard deviation errors are indicators that this increase is insignificant. On the other hand, having half of the iron sulphate concentration (media numbers 2-I and 2-II) as well as in the absence of this compound (media numbers 1-I and 1-II), it is not clear if the effect

on the concentration of product is significantly worse or insignificant, due to the fact that samples of the same duplet demonstrate inconclusiveness.

However, taking these results into account, we showed that iron sulphate has no significant role in terms of the increase of the product concentration and consequently, for the activity of the Mhy.

4.2.2 Experiments without $FeSO_4 * 7 H_2O$

The aim of these experiments is to study the influence of the absence of iron sulphate in both (pre-) and main culture media on the product concentration. It is therefore intended to confirm the previous conclusion.

In this case, no iron sulphate was used in the pre-culture and culture media and the results were normalised to the product concentration given by the standard culture medium and presented in Figure 4.5.

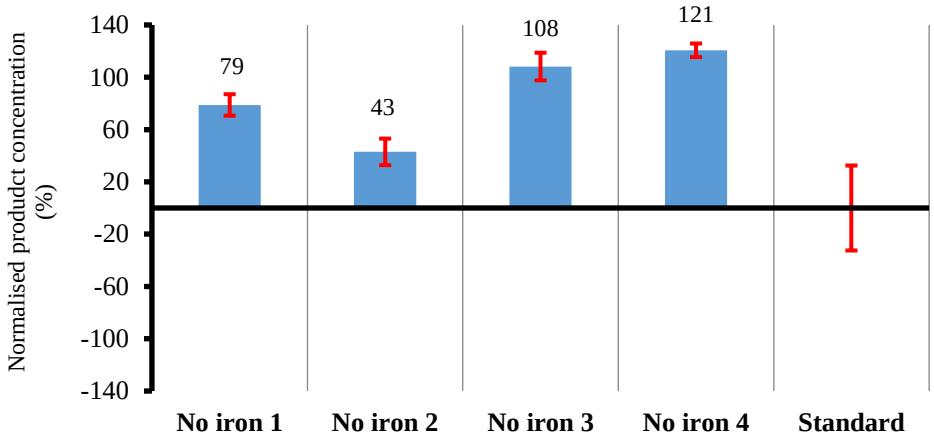


Figure 4.5 Normalised product concentrations (%) obtained without $FeSO_4 * 7 H_2O$ in the (pre)- and main culture media, compared to the literature (pre-) and main culture medium.

Four biotransformations in triplicate and without iron sulphate in both pre-culture and culture medium were carried out in order to confirm the predictions of the insignificance of this compound in the culture medium for the activity of the Mhy (No iron 1, 2, 3 and 4 from Figure 4.5). The standard culture medium does not show any percentage in Figure 4.5 since the normalisations of the product concentrations were done for this culture medium.

From this experiment, one can conclude that the total absence of iron sulphate (in both pre-culture and culture media) yields higher product concentrations. In the case of the experiments “No iron 3” and “No iron 4” more than twice of the product concentration of the standard culture medium was achieved.

4.2.3 General points and main conclusions

It should be noted that, since figures 4.4 and 4.5 are results from different batches of experiments, the results show differences regarding the standard culture medium (data not shown). For this reason, it is important to test a culture medium (in this case, the literature standard culture medium) in each batch of experiments as reference. Uncontrollable factors are permanently associated to the performance of the microorganisms in terms of the enzymatic activity. For instance the beforehand plated microorganisms can be more or less active for the enzyme under study as well as colonies from the same plate, due to different protein-encoding gene expressions. The age of the plate also plays a role in the enzymatic activity since the longer the plates are kept stored, the lesser the enzyme is expressed by the microorganism.

Thus, although the optimisation of the culture medium for the activity of the Mhy is only locally achieved, i.e., optimisation is only valid on the same batch, the results from different batches can be compared as long as all the tested culture experiments are normalised for a reference culture media of the respective batch.

Hence, the combination of the results from Figure 4.5 with the results from Figure 4.4 already discussed lead to the conclusion that there is no contribution of the iron sulphate heptahydrate for the enzymatic activity of the Mhy. Thus, it was decided to discontinue the use of iron sulphate heptahydrate for the further experiments and the protocol described in the literature was modified in this regard.

Without the need of the filter-sterilised solution C (standard culture medium, section 3.5.1), 10 mL of DI water is added to solution A as replacement, thereby saving time and material resources. Also the referred issues with the OD_{600} analysis due to the Fe (III) precipitation once added to water were suppressed and this analysis can now be performed. Attachment E displays the OD_{600} and pH analysis, respectively, performed in order to evaluate the growth of the microorganism in a culture medium with the same composition of the standard, but with solution C replaced by the same quantity in DI water on solution A (without iron sulphate heptahydrate).

4.3 Screening experiments

4.3.1 Evaluation of the product concentration

As described in section 3.3.1, the statistical software Design Expert ver. 7.0.0 was used to test the effect of four different factors on the Mhy activity as well as to study the interactions between these factors in preliminary screening experiments. Twelve culture media were tested of which four are a reference culture media. The composition of each tested culture medium is described in Table 3.4, section 3.5.1 and the results in terms of product concentrations displayed on the following Figure 4.6

are normalised to the average of the four product concentrations given by the reference culture media. Each result is an average of a triplicate shown with the respective standard deviation error.

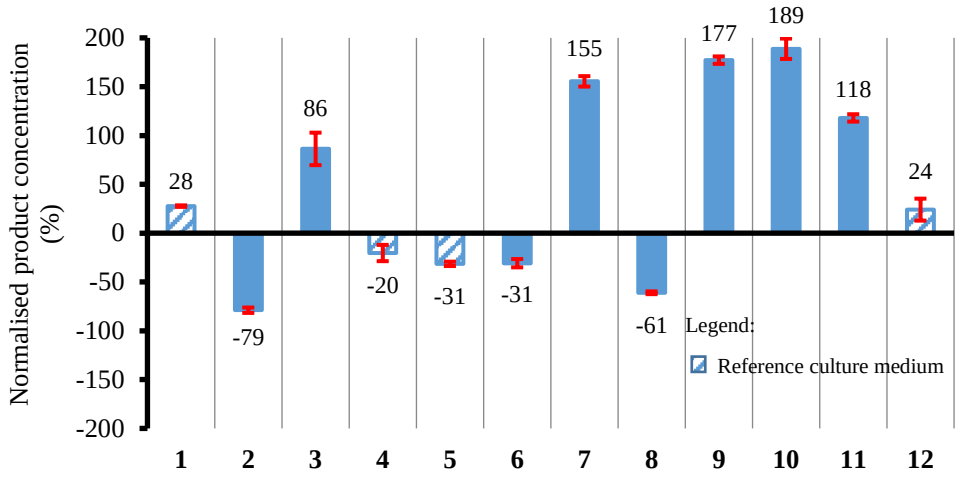


Figure 4.6 Normalised product concentrations (%) of the tested culture medium in the screening experiments.

At the first instance it should be emphasised that a change of the glucose concentration affects the product concentration. In general, high levels of glucose lead to low product concentrations. Analysing the composition of the culture media numbers 2, 6 and 8 described in Table 3.4 from section 3.5.1, glucose is at its high level and these experiments provided the lowest values of product concentration. On the contrary, the low levels of glucose led to higher product concentrations, confirmed by the culture media numbers 9, 10 and 11. Comparing the culture media numbers 9 and 10, the values of the product concentrations show insignificant differences.

Checking the culture medium number 7, this has all the tested variables at its high factorial levels. The concentration of each tested variable is twice the standard concentration. Thus, one could expect a high value of product concentration to be observed due to the upscale in the culture medium composition, however, as it can be observed in Figure 4.6, this is not the culture medium that gave the highest product concentration in this experiment.

Regarding the culture medium number 3, in which all the variable concentrations are half of the standard and no magnesium sulphate heptahydrate was used, the product concentration is even higher than the product concentration given by the reference culture media. However, no insights on this result can be done. Significance and interactions among factors may be the source of inexplicable results, therefore it was necessary to carry out an analysis of variance.

4.3.2 Analysis of variance (ANOVA)

Table 4.2 presents the output results for the analysis of variance and Fisher test. Due to the fact that the ratio of the highest value of the obtained product concentration (culture media number 10) to

the lowest value (culture media number 2) is higher than 10, namely 13.91, a transformation of the response using an inverse square root function was suggested by the software and the analysis of variance was performed for the transformed data. Information concerning the transformation of the data is explained in Appendix F.

Table 4.2 Analysis of variance (ANOVA) and Fisher test.[24]

Source	Sum of square (SS_x)	Degrees of freedom (d.f.)	Mean square (MS_s)	F value	Probability value ($P_x > F$)
Model	17.18	6	2.86	27.86*	0.0011
A- Glucose	8.16	1	8.16	79.41*	0.0003
B – Yeast extract	3.69	1	3.69	35.88*	0.0019
C - Peptone	1.75	1	1.75	17.03*	0.0091
D – $MgSO_4$ * 7 H_2O	0.05	1	0.051	0.50	0.5121
AB	3.22	1	3.22	31.33*	0.0025
AC	0.76	1	0.76	7.44*	0.0414
Residual	0.51	5	0.10	-	-
• Lack of fit	0.27	2	0.09	0.73	0.6222
• Pure error	0.25	3	0.12	-	-
Cor Total	17.69	11	-	-	-

The values of F marked with * are the significant values in the F-test for the confident interval of 5 %. A significant F-value confirms the significance of the source.

As mentioned in chapter 2, in the DoE, a statistical model is fitted to the tested data, and also evaluated using the analysis of variance and F-test. The model F-value of 27.86 implies the model is significant and there is only a 0.11 % chance that a "Model F-Value" this large could occur due to noise. The model error can be attributed to model lack of fit and experimental noise. In this case, the lack of fit F-value of 0.73 implies the lack of fit is insignificant relative to pure error and there is a 62.22 % chance that a "Lack of Fit F-value" this large could occur due to noise. An insignificant lack of fit is desirable, meaning that the model fits well the experimental data. [24]

Despite the statistical model regression attributed to the experimental data (equation and determination coefficients not shown), the purpose of the analysis of variance for the screening experiments was to evaluate the importance of the variables in the response as well as their interactions, i.e., from the four constituents of the culture medium in study, which are the most important for the activity of the Mhy tracked by the biotransformation product concentration.

Thus, with these results, the unveiled significant variables are glucose (A), yeast extract (B) and peptone (C), as well as any significant variable interactions, namely those between glucose and yeast extract (AB) and glucose and peptone (AC). On the other hand, the effect of the magnesium sulphate heptahydrate (D) is insignificant for the Mhy activity.

- **Diagnostics**

To draw conclusions from this empirical experiment, the mathematical model and its respective analysis of variance were assumed as valid, for the defined confident interval of 5 %. However, it is necessary to evaluate the truth of these assumptions, namely if the errors are independent and normally distributed with null average and constant variance. In that way the residuals, which are approximations of the experimental errors, are calculated by the difference between the experimental (observed) values and the correspondent values predicted by the fitted model.[13] The residuals should be unpredictable and not described by the deterministic portion of the model. Besides that, it is desired that no observable predictability of the errors exist. Without randomness and unpredictability as components of any regression model, the model is not a valid approximation of reality.

The analysis of the residual plots (preferably externally studentised residuals) using the Design Expert helps to assess if the errors occur stochastically. A detailed explanation of how to calculate studentised residuals can be found in Appendix G.

Figure 4.7 displays the chart of normal probability distribution of the internally studentised residuals (top left), where it can be observed that the residuals follow the linear tendency of the normal probability, thus they are normally distributed.

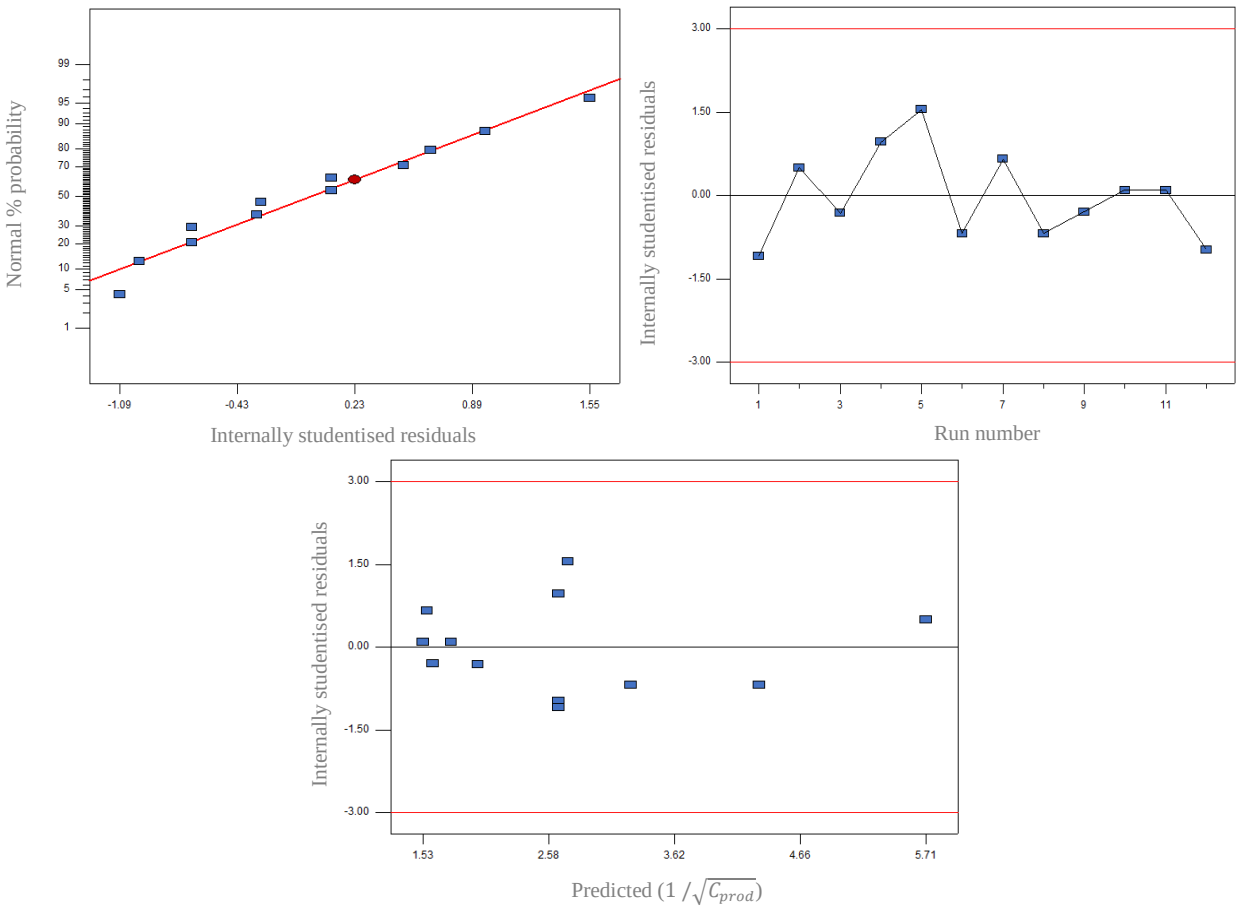


Figure 4.7 Plots of Normal probability distribution of the internally studentised residuals (top left), internally studentised residuals per run number (top right) and internally studentised residuals per predicted concentration of product (bottom).

In the same figure, the independence of the residuals is checked by the chart of the internally studentised residuals per run number (top right), where no tendency of the values of the residuals is observed. They are randomly distributed with positive and negative values, and the closer they are to 0, the closer the predicted values are to the experimental values. Although this is a condition for the validation of the assumptions, these experiments were completed at the same time and the independence was guaranteed by the randomness of the numbered 1-12 culture media compositions.

The chart of internally studentised residuals per predicted values of product concentration (Figure 4.7 bottom) also confirms the homogeneity of the variance. It should be noted that the predicted values of the concentration have the inverse square root transformation applied. In this chart, there is no special tendency of the residuals which also confirms that they are not a function of the predicted values.

If a given value of an internally studentised residual would exceed the limit of ± 3, it would indicate that the residual is a potential outlier, i.e., a residual with a value so distinct from the rest that could affect the analysis of variance.[13] That was not the case of these experiments, were no outliers were detected in the diagnostics, indicating a good and valid ANOVA.

4.3.3 General points and main conclusions

Although a plethora of diagnostic tools can be applied, the ones presented are sufficient to acknowledge that the model is adequate.

Thus, ensuring the truthfulness of the assumptions regarding the ANOVA, the screening experiments confirmed unequivocally that the magnesium sulphate heptahydrate is unnecessary to the culture medium for the activity of the enzyme under study. Therefore, it was decided to discontinue the use of this compound in the culture medium, saving resources. Without magnesium sulphate no filter-sterilised solution B (standard culture media, section 3.5.1) is required and this solution was replaced by the same quantity of DI water on solution A, for both (pre-) and main culture media.

4.4 Response surface methodology (RSM)

4.4.1 Evaluation of the product concentration

As described in section 3.3.2, using a circumscribed central composite design, experiments varying the concentration of glucose, yeast extract and peptone were accomplished with the support of the statistical software. Twenty culture media were tested, of which six are reference culture media, with the same composition of the standard but without the solutions B and C, having DI water used in replacement. The compositions of each culture media are presented in Table 3.5 (section 3.5.1) and the results are displayed in Figure 4.8 in terms of product concentration. The results were normalised to the

average of six product concentrations given by the reference culture media and each result is an average of a triplicate shown with the respective standard deviation error.

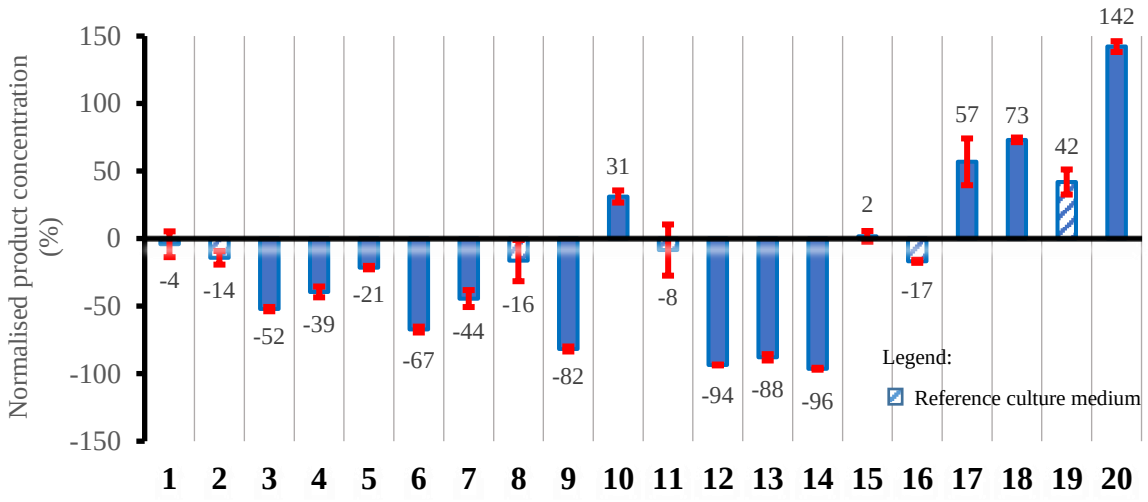


Figure 4.8 Normalised product concentrations (%) of the tested culture media in the response surface methodology experiments.

From these results one can conclude that most of the tested conditions yield worse results in terms of product concentration, compared to the average of the standard culture media product concentrations. All the triplicates also show a small standard deviation error.

Product concentration given by the culture medium number 20 is clearly the best result, yielding more than twice of the product concentration given by the average of the standard culture media.

In accordance with the results from Figure 4.6 and checking Table 3.1 and Table 3.2, the culture media that gives the best results has low levels of glucose and high levels of yeast extract and peptone. This indicates that the best tested culture medium is close to an optimal culture medium, for providing higher biotransformation yields and, consequently, a higher activity of the Michael hydratase. However, in order to reach the best culture media conditions in the range of the central composite design and make predictions for optimal conditions outside of the range of study, a mathematical model (response surface) had to be fitted to the experimental data.

4.4.2 Analysis of variance (ANOVA)

At first instance, an analysis of variance of different models and respective lack of fits was provided by the Design Expert in order to select the highest polynomial order where additional terms are significant, lack of fit is insignificant and the model is not aliased.[24] Explicitly, the software suggests a type of model with polynomial terms which significance and lack of fit are tested using the analysis of variance. Additionally, the models are tested for maximising the adjusted determination coefficient and predicted determination coefficient. The results concerning these analyses are displayed Appendix H.

To further improve the fitting of the chosen model to the experimental data, manual selection of the polynomial terms was performed. Thus, the analysis of variance shown in Table 4.3 contains the selected significant terms for the model that provided the best fit to the experimental data as sources, with less lack of fit and higher determination coefficient (R^2), predicted determination coefficient (Pred R^2) and adjusted determination coefficient (Adj R^2). The importance of these coefficients is discussed further in this section.

The data was transformed using a power function as well as in the previous experiment. In this case a square root transformation was used and the analysis of variance was performed to the transformed data.

Table 4.3 Analysis of variance (ANOVA) of the central composite design, response surface methodology.

Source	**Sum of square (SS_x)**	**Degrees of freedom (df)**	**Mean square (MS_x)**	**F value**	**Probability value $P_x > F$**
Model	0.36	6	0.059	23.60*	<0.0001
A- Glucose	0.040	1	0.040	16.00*	0.0015
B – Yeast extract	0.18	1	0.18	73.34*	<0.0001
C - Peptone	0.050	1	0.050	20.01*	0.0006
BC	0.028	1	0.028	11.34*	0.0050
B^2	0.025	1	0.025	10.08*	0.0073
C^2	0.032	1	0.032	12.71*	0.0035
Residual	0.033	13	2.508E-003	-	-
• Lack of fit	0.025	8	3.063E-003	1.89	0.2500
• Pure error	8.095E-003	5	1.619E-003	-	-
Cor Total	0.39	19	-	-	-

With this analysis of variance, all the tested sources are significant for a confident interval of 5 %, proved by the Fisher's test of which the F-values are marked with a *.

The model F-value of 23.60 implies the model is significant and there is less than 0.01 % chance that a "Model F-Value" this large could occur due to noise. The model error can be attributed to model lack of fit and experimental noise. In this case, the lack of fit F-value of 1.89 implies the lack of fit is insignificant relative to pure error and there is a 25 % chance that a "Lack of Fit F-value" this large could occur due to noise.[24]

In addition to the analysis of variance, a multiple regression analysis was performed. The determination coefficients of the model, adequate precision, standard deviation and coefficient of variation were calculated using the software and displayed in Table 4.4.[24]

Table 4.4 Multiple regression analysis.[24]

Model terms	Values
Std. Dev.	0.050
CV (%)	14.88
R^2	0.916
Adj. R^2	0.877
Pred. R	0.786
Adeq. precision	17.385

The efficiency of fit of the model was checked by the determination coefficient (R^2). In this case, the value of the determination coefficient (R^2) indicates that only 8.4 % of the total variations are not explained by the model. The value of the adjusted determination coefficient (Adj. R^2) is also high, which indicates a high significance of the model.[43-44] The predicted determination coefficient (Pred. R^2) indicates how well a regression model predicts responses for new observations. The value should not be as high as the R^2 in order to prevent an overfitting of the model. When a model is over fitted, there is a higher chance of it to start predicting the random noise, which is not desirable.[45]

Adequate precision measures the signal to noise ratio. A ratio greater than 4 is desirable and in this case the model has an 'adequate precision value' of 17.385. At the same time low values of the standard deviation (Std. dev.) and coefficient of variation (CV) indicate improved precision and reliability of the conducted experiments.

The coefficients of regression of each significant term were also calculated using the Design Expert and the Equation 4.4 was obtained.[24]

Equation 4.4 Regression quadratic model.[24]

$$\sqrt{c_{prod}} = 0.09 - (0.01\ A) + (0.33\ B) + (0.05\ C) + (0.05\ BC) - (0.17\ B^2) - (7.49 * 10^{-3}\ C^2)$$

Where,

c_{prod} – concentration of product (g/L); A – Glucose (g/L); B – Yeast extract (g/L); C – Peptone (g/L);

Interpreting the Equation 4.4 it can be seen that the variable with the largest effect is the yeast extract (B), due to its high values of the coefficients of regression for the linear (B) and quadratic (B^2) terms as well as the interaction (BC). The highest F-value and consequently the lowest probability value (Prob. > F) displayed in Table 4.3 also confirm that B is the most significant variable.

Having Equation 4.4 and Table 3.5 the predicted values can be calculated. The software gives a graphic representation of the predicted values versus the experimental values of the transformed response as output. Figure 4.9 displays this scatter graph where it can be observed 20 dots representing the 20 different culture medium tested in this experiment. The experimental values of the product concentration are under comparison with the values of product concentration predicted by the model.

These values are distributed around a linear relation y=x that represents the equality between the predicted and the experimental responses. The closer the dots are to the line, the equal the predicted responses are to the experimental responses.

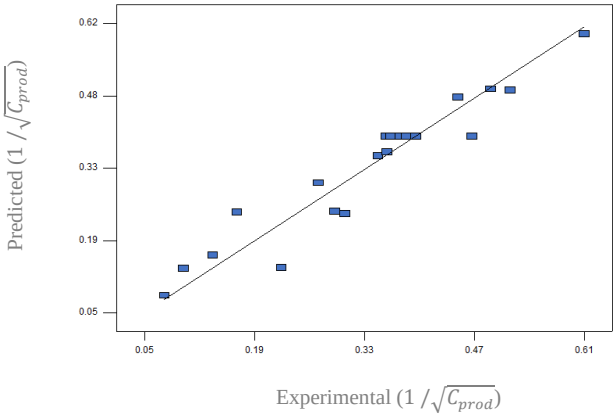


Figure 4.9 Plot of the predicted versus experimental values of the transformed product concentration.

- **Diagnostics**

The same type of diagnostics used in the screening experiments were applied to these experiments. Figure 4.10 displays the chart of normal probability distribution of the internally studentised residuals (top left), where it can be observed that the residuals follow the linear tendency of the normal probability, thus they are normally distributed.

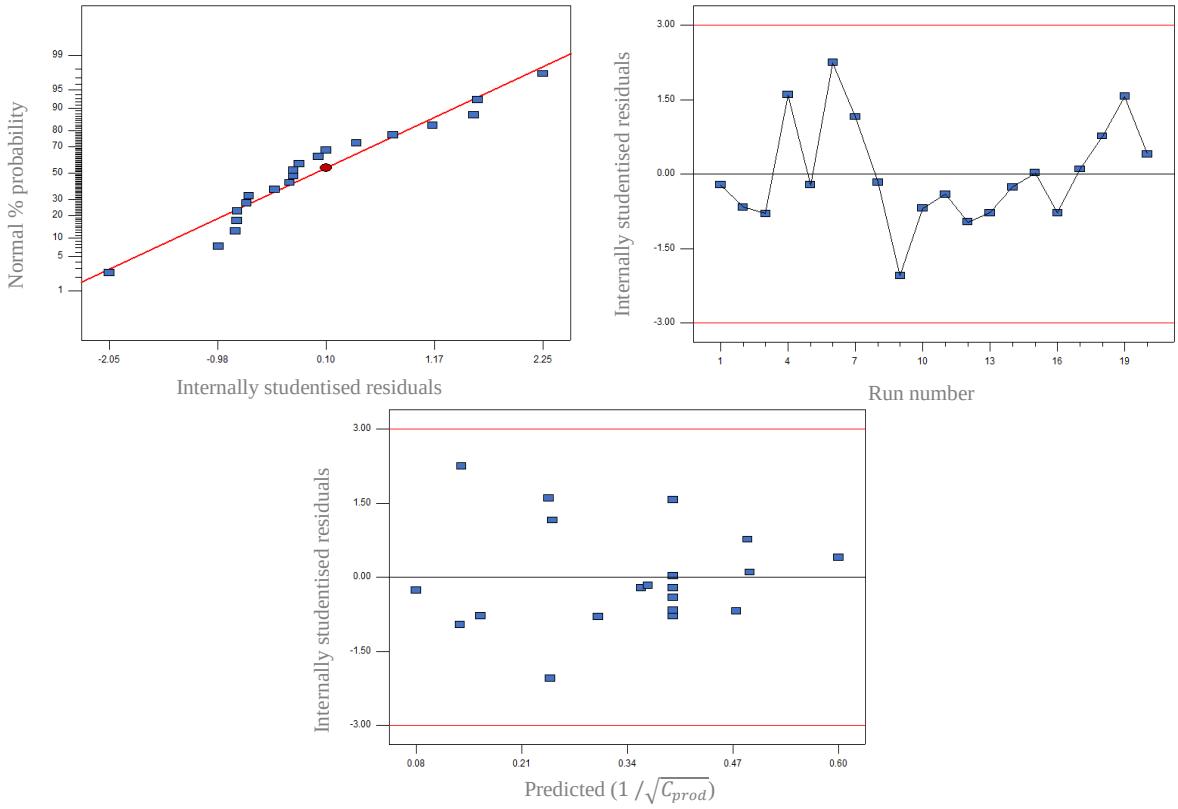


Figure 4.10 Plots of Normal probability distribution of the internally studentised residuals (top left), internally studentised residuals per run number (top right) and internally studentised residuals per predicted concentration of product (bottom).

Once again, the independence of the residuals is checked by the chart of the internally studentised residuals per run number (top right), where no tendency of the values of the residuals is observed (they are randomly distributed with positive and negative values). These experiments were also completed at the same time and the independence was guaranteed by the randomness of the numbered 1-20 culture media compositions.

The chart of internally studentised residuals per predicted values of the transformed product concentration (Figure 4.7 bottom) also confirms the homogeneity of the variance since there is no special tendency of the residuals.

- **Response surface analysis**

The validity of the assumptions concerning the model were confirmed in the diagnostics. Hereupon, the model described by Equation 4.4 can be studied in order to find optimal values of glucose, yeast extract and peptone that optimise the concentration of product and therefore the activity of the enzyme Michael hydratase.

To locally study optimal values of the concentration of product, the software designs a response surface with Equation 4.4 in function of two variables, i.e., a three-dimensional plot. The only possible surface in this experiment is the simultaneous variation of yeast extract (B) and peptone (C) due to their significant interaction (BC). In this case, the variable glucose (A) is fixed to a certain value whose change will also determine the proximity to optimal product concentration values.

In the screening experiments it was found out that low values of glucose contribute to the increase of the product concentration. The same conclusion can be drawn from these experiments.

In Equation 4.4 a negative term is associated to the glucose (A) and testing the response surface for different values of glucose, the proximity to optimal values of product concentration increase when the concentration of glucose decreases. The quadratic terms of yeast extract (B^2) and peptone (C^2) are also negative but the linear terms (B, C) and interaction (BC) are positive. The study intervals of both variables can be defined in the tested response and it was found that the concentration of product increases with an increase of the concentrations of yeast extract and peptone.

However, the mathematical model neither takes the biological effect of a decrease in glucose, nor the biological effect of an increase in yeast extract and peptone into account. Thence, the design space was limited to the lowest value of glucose tested in this experiments, which is the value of 6.59 g/L (Table 3.5) in the axial point $-\alpha$ (Table 3.2). The same was defined for yeast extract and peptone, having the study interval for these variables been limited to the values of each in the axial points ($-\alpha$, $+\alpha$).

Hereupon, Figure 4.11 displays the response surface with the concentration of glucose fixed to 6.59 g/L, where a local maximum of the concentration of product can be found.

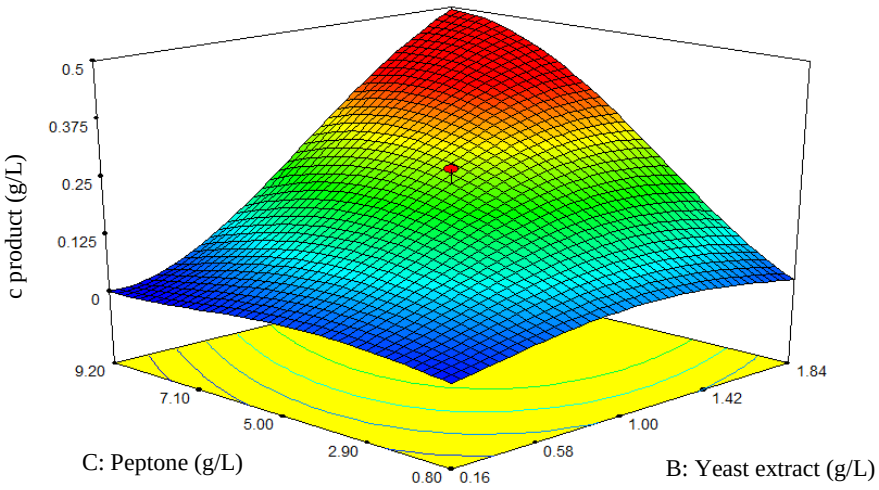


Figure 4.11 Response surface of the concentration of product in function of the concentration of peptone and concentration of yeast extract.

Figure 4.11 shows a colour variation from blue to red with the increase of the concentration of product. Additionally to the response surface, Figure 4.12 displays the contour plot of the response surface that helps in the observation of the maximum product concentration in this range of study.

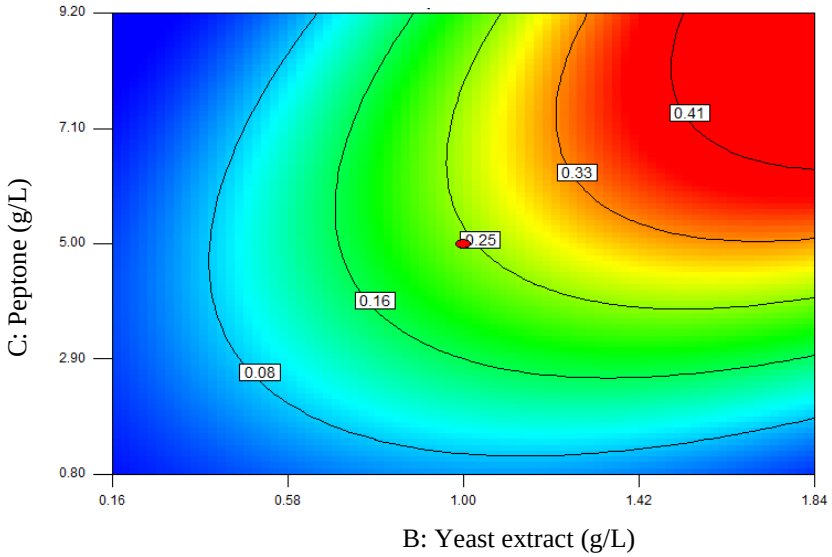


Figure 4.12 Contour plot of the response surface.

The optimal values were therefore predicted to be 9.20 g/L for peptone, 1.84 g/L for yeast extract and the fixed 6.59 g/L for glucose. With this values, the model estimates the best value of product concentration of 0.49 g/L (4.8 mM) in the case of this experiments, however this has to be confirmed in further confirmation experiments.

4.5 Confirmation experiments

4.5.1 Evaluation of the product concentration

It should be noted that there are variations of the results from batch to batch due to uncontrollable factors (noise), as it was explained in section 4.2.3. Therefore higher or lower values of product concentration can be obtained using the best values predicted by this model. Still, if the model makes a good prediction, the best combination described before will give the best yield in biotransformation product, when compared to different culture media compositions of the studied compounds.

The confirmation experiments are therefore important to prove the best combination of factors given by the model in a new batch of experiments. New culture media were tested with the same compositions of some that were tested in previous experiments (check section 5.3.1 – confirmation experiments) and compared in terms of normalised product concentration. Figure 4.13 displays a scheme that relates the different culture media tested in these experiments with the culture media tested in previous experiments and shows the evolution from experiment to experiment in terms of the presence of magnesium sulphate on both (pre-) and main culture media.

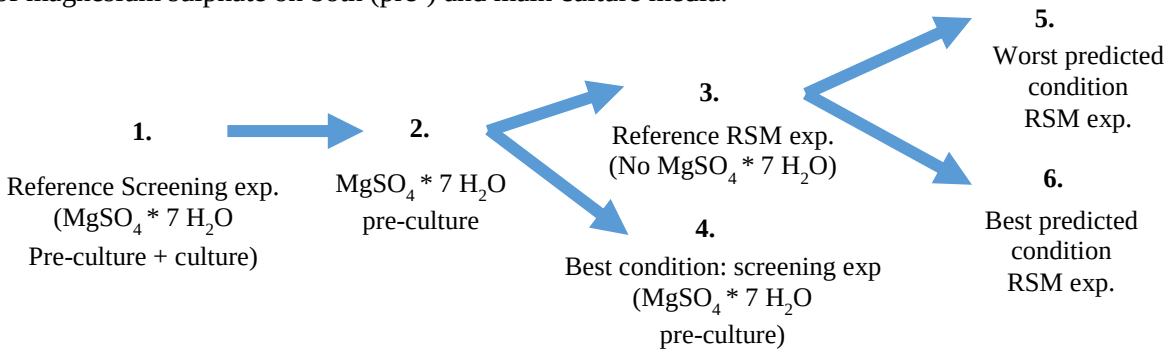


Figure 4.13 Scheme of the relation between the different culture medium tested in these experiments.

Figure 4.14 displays the results where the product concentrations are normalised to the culture medium number 1, established as the standard in these experiments, and each experiment is an average of a triplicate shown with the respective standard deviation error.

Culture medium number 6 has the best composition predicted by Equation 4.4 that indeed led to the best result in terms of product concentration. More than triple of the product concentration given by the standard culture medium was obtained with this culture medium. On the other hand, the composition of the culture medium number 5 gave a low prediction of the concentration of product by the same equation, confirmed by the practical results. A decrease of 11 % of the product concentration given by the standard culture medium is observed in the results given by the culture medium number 5.

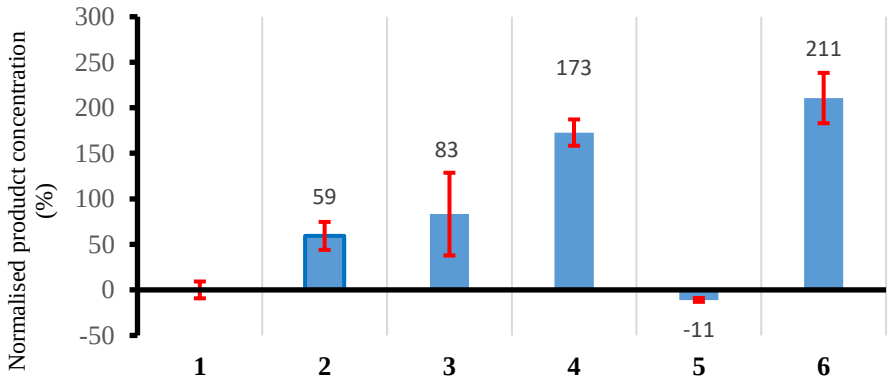


Figure 4.14 Normalised product concentrations (%) of the tested culture media in the confirmation experiments.

In these experiments, the standard culture medium (1) has the same composition of the reference culture medium of the screening experiments. This culture medium can be compared to the culture medium 3, which has the same composition of the reference culture medium of the response surface methodology experiments. It can be observed in Figure 4.14 a significant increase of 83 % of the product concentration comparing these culture media. They only differ in the addition of magnesium sulphate. Culture medium 3 does not have any magnesium sulphate while culture medium 1 possesses this compound. Therefore these results confirm the conclusions drawn from the screening experiments, namely that the magnesium sulphate has no added value to the culture medium.

The culture medium number 2 only possesses magnesium sulphate on the pre-culture medium, thus only 16 µM of $MgSO_4$ is present in the culture medium. This can explain the insignificant difference in the result of product concentration between the culture media numbers 2 and 3.

The culture medium number 4 has the same composition of the culture medium number 10 of the screening experiments, which gave the best product concentration result in that batch. In this case, the product concentration is higher than the given by the culture medium number 3, corroborating once again the results of the screening experiments.

4.5.2 General points and main conclusions

More than a confirmation that the predictions of the valid mathematical model gave the best result in terms of product concentration, these experiments served to conclude that an improvement of this product concentration has been achieved throughout all the experiments. Comparing the standard culture media defined for each batch of experiments was important, since this "standard" has also been improved from batch to batch due to the absence of magnesium sulphate.

4.6 Fermentor

The fermentation proceeded using the culture medium number 10 from the screening experiments, as described in section 3.8. During the fermentation time, a continuous analysis of the pH and pressure of oxygen was performed. Figure 4.155 displays the graphical analysis of the referred three parameters.

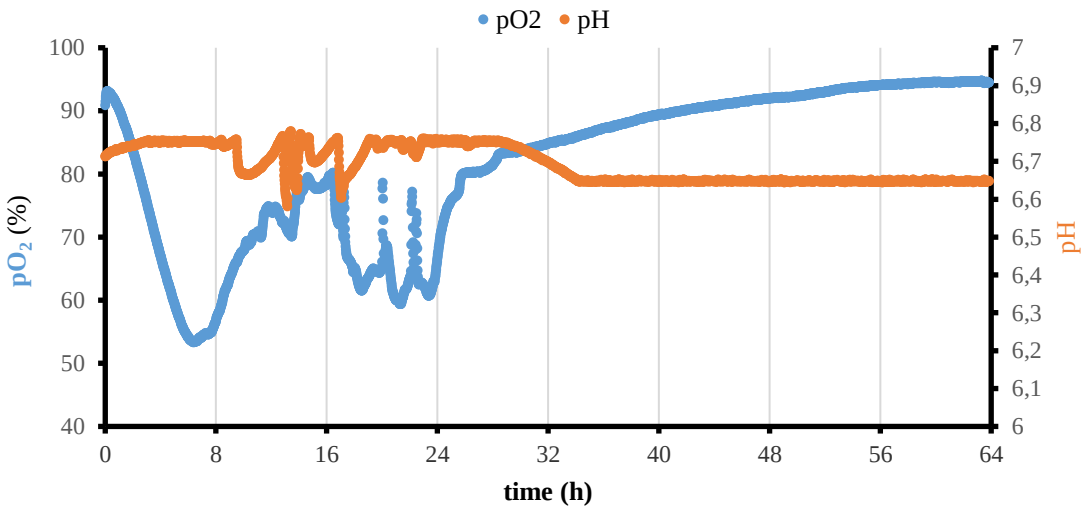


Figure 4.15 Analysis of pO_2 and pH performed during the fermentation.

It can be observed that the pH shows a small variation during the time of growth. The culture medium becomes more acidic, which was also confirmed by the pH analysis performed after the experiments with iron sulphate (check Appendix E). In this case the addition of acid and base are used to control the pH to the set value of 6.7, explaining the slopes between 8 and 24 h. A small decrease to 6.65 and further stabilisation is observed after 32 h, suggesting the end of the exponential phase of growth.

Analysing the pO_2 the slope observed at 8 h suggests that the lag phase ended and an exponential phase of growth began. However, an increase of the pO_2 can be observed and subsequent slopes between 16 h and 24 h. This result suggest that a change in the carbon source might be affecting the growth of the microorganism. Thus, the glucose might not be used as carbon source after a certain time of growth and may even work as biological inhibitor.

This assumption is also in agreement with the product concentration results from both screening and response surface experiments, were low values of glucose provided higher yields of biotransformation product and therefore higher enzyme activity. However no biological explanation can be drawn from this result since the isolation of the enzyme is still in progress and the type of metabolism associated to its synthesis is yet to be discovered.

In terms of the tested product concentration following the procedure described in section 3.6, the results of the GC analysis (data not shown) showed high product peak areas, confirming a high activity

of the grown cells. Besides that, the 15 L fermentor gave a substantially higher amount of biomass (data not shown), compared to what is generated in incubators.

Thus, from this experiment one can conclude that transferring the growth from the incubators to a fermentor with optimised medium compositions led to large amount of biomass with high Michael hydratase activity thereby saving a substantial amount of growth time.

4.7 Further discussion

In this study, the effect of the culture medium composition on the activity of the enzyme Michael hydratase from *Rhodococcus rhodochrous* ATCC 17895 was investigated. Starting from a culture medium composition described in the literature, the values of the parameters were changed in order to increase the yield in biotransformation product **2**.

The activity of the enzyme was tracked by the product concentration. All the biotransformation conditions were kept constant, namely the cell content in each tested vial, quantity of substrate used, temperature, buffer and reaction time and speed. Also the growth time, temperature and speed in the incubators were maintained as well as the buffer of the culture media. Thus, it was ensured that the effects on the product yield were only due to the change of the composition of the culture medium.

It was concluded that two compounds, namely iron sulphate heptahydrate and magnesium sulphate heptahydrate, are not essential to the culture medium in order to achieve higher amounts of product concentration. Contrariwise, it was proven that the presence of these compounds can be prejudicial to the yield of biotransformation product. Also, issues regarding the OD_{600} analysis due to the Fe (III) precipitation were supressed with the absence of iron sulphate. The protocol was therefore changed, saving resources and time since there is no need to filter-sterilise the solutions that contained these compounds.

An optimisation of the composition of the culture medium in order to increase the yield in biotransformation product was achieved using the software Design Expert, and a valid mathematical model was created. Optimal variable values' were found using the model response surface in the interval defined by the central composite design. The values of 6.59 g/L for glucose, 1.84 g/L for yeast extract and 9.20 g/L for peptone provide the best result in terms of product concentration, within the defined interval, as it was proved in the confirmation experiments.

However, these optimal values might not be the best obtained by the mathematical model due to the limitation of the interval of study of the variables. Increasing the interval of study of peptone and yeast extract in the response surface and maintaining the glucose at its lowest tested value of 6.59 g/L, it is possible to check an increase in the biotransformation product. Figure 4.16 displays the response surface to a higher interval of concentration of yeast extract and peptone and the shading presented is the studied interval displayed in Figure 4.11.

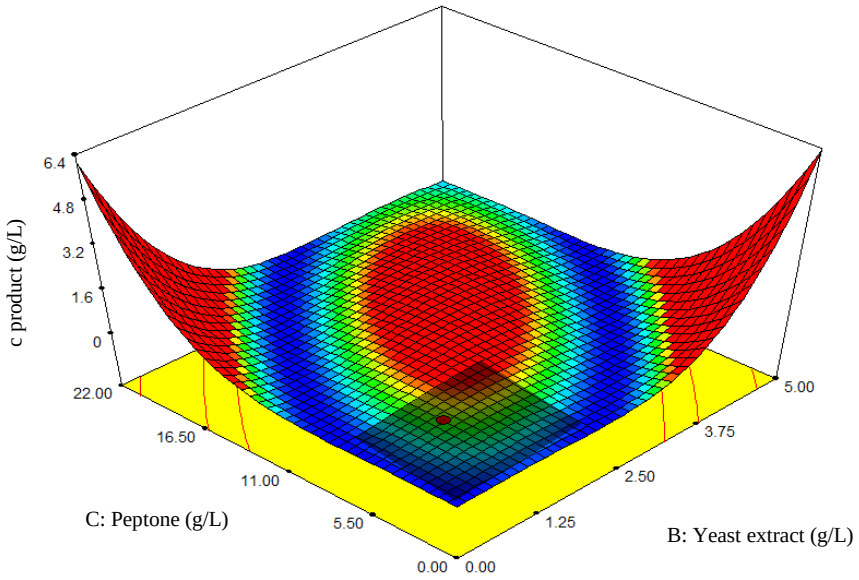


Figure 4.16 Response surface of the product concentration dependent on the concentration of peptone and concentration of yeast extract.

Figure 4.17 helps in the analysis of the predictions of a higher concentration of product. As it can be observed, the shade area is the interval of study in Figure 4.12.

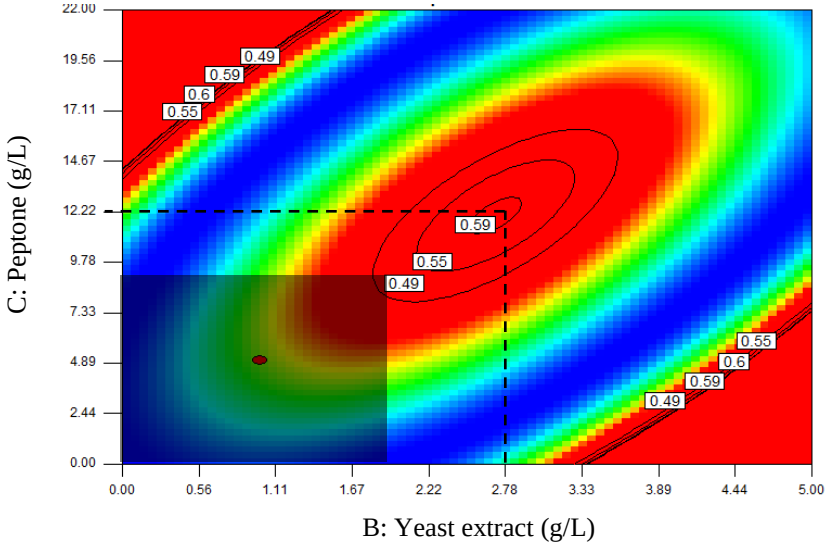


Figure 4.17 Contour plot of the response surface.

A higher amount of concentration of product (0.59 g/L, 5.8 mM) is estimated by Equation 4.4 when increasing peptone and yeast extract to values of 12.22 g/L and 2.78 g/L, respectively. However, this was not experimentally tested.

The values on the boarders of the Figure 4.17 predicted by the model as optimal were neglected. It is believed that there would be biological consequences of having either higher amounts of peptone and lower amounts of yeast extract or lower amounts of peptone and higher amounts of yeast extract. Therefore, it should be emphasised that Equation 4.4, as an empirical model, is incapable to predict the

truthfulness of a higher yield in biotransformation product (activity of the Mhy) in intervals of variables that go far beyond the defined ones.

CHAPTER 5

Conclusion and outlook

5.1 Conclusion

More than an optimisation of the culture medium composition in order to improve the activity of an enzyme, this study should serve as a proof that studies concerning the culture medium are important for the expression of protein-encoding genes from microorganisms. Enzymes are extremely sensitive molecules, and culture media formulations should always be a target of study for the enhancement of the production of these biological molecules.

With this study the Michael hydratase from *R. Rhodochrous* ATCC 17895 showed high dependence on the composition of the culture medium. It was possible to triple the activity of this enzyme reformulating the old culture medium protocol, as it was proved by the confirmation experiments.

Furthermore, using a fermentor for the growth of highly Mhy active cells revealed to be advantageous when it is desired to obtain a higher amount of biomass in a shorter amount of time.

The implementation of a completely empiric approach using DoE was useful since the biological interactions and the type of metabolism assigned to the expression of this enzyme are still unknown. Moreover, with a complex culture medium, it is difficult to assign a role to each medium constituent since peptone and yeast extract are compose of several chemical compounds. The monitoring of the pO_2 in the fermentor experiment suggested a change in the carbon source and glucose revealed to have a negative effect in the Mhy activity proved in both screening and response surface methodology experiments. It would not be possible to take such conclusions if an OVAT approach was used.

The present results will aid in the ongoing process of isolation of the Michael hydratase from this microorganism.

5.2 Outlook

At first instance, taking into account Figure 4.17 and the subsequent discussion, the values of 6.59 g/L for glucose, 12.22 g/L for peptone and 2.78 g/L for yeast extract should be tested in a new batch in order to confirm that the predictions of the model outside the interval of study are reliable.

The effects of the *R. rhodochrous* ATCC 17895 growth conditions studied in the scope of this project concentrated on the composition of the culture medium while the remaining growth conditions like temperature, pH, speed and time, were maintained.

The growth temperature in use (26 ºC) is the recommended by the ATCC for the propagation of the strain and also reported in the literature as the optimal growth temperature.[46]. It was also proven that the same temperature is the optimal in the biotransformation step, however, this might not be true for the growth step.[11] The Mhy enzymatic activity from *Rhodococcus rhodochrous* ATCC 17895 is likely to be influenced by the growth temperature, therefore studies regarding this would be a useful investigation.

An optimal pH of 6.2 is also reported for the biotransformation step using a KPi-buffer (Potassium phosphate buffer), however no optimal growth pH was reported for the enhancement of the Mhy activity from this microorganism.[11] The same KPi-buffer of the biotransformation step is used in the literature culture medium to give an initial growth pH of 7.2.[12] When the iron sulphate and magnesium sulphate are removed and the quantities of the some constituents were changed, the initial pH decreased (data not shown). The quantities of K_2HPO_4 and KH_2PO_4 used for the buffer were however maintained. Thus, in addition to the temperature investigation, it would be interesting the performance of experiments in order to check the role of the culture medium pH in the Mhy activity.

The literature reports that *Rhodococcus rhodochrous* ATCC 17895 is able to use other carbon sources as for instance, fructose, mannose or sucrose.[46] Thus, substituting the currently used carbon source (glucose) in order to check the effect on the Mhy activity would be an investigation worth to try.

However, since the glucose showed negative influence for the Mhy activity and the carbon source is likely to change during the growth time of the microorganism (fermentor results), a use of a defined culture medium, i.e. a type of medium that does not have complex ingredients (peptone, yeast extract) and all the chemical species are known, is suggested. The lesser complex the medium is, the easier the understanding of the influence of each ingredient on the Mhy activity would be.

Bibliography

[1] K. Faber, *Biotransformations in Organic Chemistry*, 6th ed., Springer, **2011**.

[2] R. Singh, M. Kumar, A. Mittal, P. K. Mehta, *3 Biotech* **2016**, *6*, 174.

[3] M. Wink, *An Introduction to Molecular Biotechnology*, 2nd ed., WILEY-BLACKWELL, Weinheim, **2011**.

[4] H. Gröger, *Angewandte Chemie International Edition* **2014**, *53*, 3067-3069.

[5] A. J. J. Straathof, P. Adlercreutz, *Applied Biocatalysis*, 2nd ed., harwood academic pulishers, Taylor & Francis e-Library, **2005**.

[6] S. Ojigbo. *Applications of enzymes in industry*. **2016** 31/12 [cited 2017 09/08]; Available from: http://www.pharmanewsonline.com/applications-of-enzymes-in-industry/.

[7] Novozymes. *What are enzymes?* [cited 2017 26/06]; Available from: http://www.novozymes.com/en/about-us/our-business/what-are-enzymes.

[8] V. Resch, U. Hanefeld, *Catalysis Science & Technology* **2015**, *5*, 1385-1399.

[9] B.-S. Chen, L. G. Otten, U. Hanefeld, *Biotechnology Advances* **2015**, *33*, 526-546.

[10] J. Jin, U. Hanefeld, *Chemical Communications* **2011**, *47*, 2502-2510.

[11] B.-S. Chen, V. Resch, L. G. Otten, U. Hanefeld, *Chemistry – A European Journal* **2015**, *21*, 3020-3030.

[12] H. L. Holland, J.-X. Gu, *Biotechnology Letters* **1998**, *20*, 1125-1126.

[13] Z. L. Pereira, J. G. Requeijo, *Qualidade: Planeamento e Controlo Estatístico de Processos*, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, **2008**.

[14] J. Antony, *Design of Experiments for Engineers and Scientists*, 1st ed., Elsevier, **2003**.

[15] A. S. Rathore, H. Winkle, *Nat Biotech* **2009**, *27*, 26-34.

[16] T. P. S. University. *1.1. - A Quick History of the Design of Experiments (DOE)*. **2017** [cited 2017 29/08]; Available from: https://onlinecourses.science.psu.edu/stat503/node/6.

[17] *Handbook - Design of Experiments in Protein Production and Purification* - GE Healthcare Life Sciences, **2014**.

[18] *NIST/SEMATECH e-Handbook of Statistical Methods*. **2003** 30/10/2013 [cited 2017 August]; Available from: http://www.itl.nist.gov/div898/handbook/.

[19] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, L. A. Escaleira, *Talanta* **2008**, *76*, 965-977.

[20] D. Baş, İ. H. Boyacı, *Journal of Food Engineering* **2007**, *78*, 836-845.

[21] S. Vieira, R. Hoffmann, *Estatística experimental*, Atlas, **1989**.

[22] R. F. Teófilo, M. Ferreira, *Quimiometria II: planilhas eletrônicas para cálculos de planejamentos experimentais, um tutorial*, **2006**.

[23] R. E. Bruns, I. S. Scarminio, B. de Barros Neto, *Statistical design-chemometrics, Vol. 25*, Elsevier, **2006**.

[24] Design Expert® v 7.0.0 ed., Stat-Ease, Inc., Minneapolis, MN 55413, **2005**.

[25] G. S. N. Naidu, T. Panda, *Enzyme and Microbial Technology* **1999**, *25*, 116-124.

[26] T. Garcia, N. Sanchez, M. Martinez, J. Aracil, *Enzyme and Microbial Technology* **1999**, *25*, 591-597.

[27] S. P. J. Namal Senanayake, F. Shahidi, *Food Chemistry* **2002**, *77*, 115-123.

[28] Q. K. Beg, V. Sahai, R. Gupta, *Process Biochemistry* **2003**, *39*, 203-209.

[29] A. Ismail, M. Linder, M. Ghoul, *Enzyme and Microbial Technology* **1999**, *25*, 208-213.

[30] D. Çelik, E. Bayraktar, Ü. Mehmetoğlu, *Biochemical Engineering Journal* **2004**, *17*, 5-13.

[31] B. Manohar, S. Divakar, *Process Biochemistry* **2004**, *39*, 847-853.

[32] P. Rai, G. C. Majumdar, S. Dasgupta, C. Chopra, *Journal of Food Engineering* **2004**, *64*, 397-403.

[33] A. Vohra, T. Satyanarayana, *Process Biochemistry* **2002**, *37*, 999-1004.

[34] X. Q. Mu, Y. Xu, Y. Nie, J. Ouyang, Z. H. Sun, *Process Biochemistry* **2005**, *40*, 2345-2350.

[35] G. Ghoshal, A. Kamble, U. S. Shivhare, U. C. Banerjee, *International Journal of Research and Reviews in Applied Sciences* **2011**, 132-137.

[36] L.-L. Yuan, Y.-Q. Li, Y. Wang, X.-H. Zhang, Y.-Q. Xu, *Journal of Bioscience and Bioengineering* **2008**, *105*, 232-237.

[37] M. S. Tanyildizi, D. Özer, M. Elibol, *Process Biochemistry* **2005**, *40*, 2291-2296.

[38] M. P. Pal, B. K. Vaidya, K. M. Desai, R. M. Joshi, S. N. Nene, B. D. Kulkarni, *Journal of Industrial Microbiology & Biotechnology* **2009**, *36*, 747-756.

[39] M.-T. Lee, W.-C. Chen, C.-C. Chou, *Biotechnology and Applied Biochemistry* **1998**, *28*, 229-233.

[40] M. P. C. Rodrigues, *High-Throughtput Resin Screening for the Purification of Aspartic Proteases by Expanded Bed Adsorption*, MSc Thesis from Instituto Superior Técnico, **2016**.

[41] G. E. P. Box, J. S. Hunter, W. G. Hunter, *Satistics for experimenters: An introduction to design, data analysis and model building*, Wiley, **1978**.

[42] *7.7: Liquid-Liquid Extractions*. **2017** 16/02/2017 [cited 2017 08/09]; Available from: https://chem.libretexts.org/Textbook_Maps/Analytical_Chemistry_Textbook_Maps/Map%3A_Analytical_Chemistry_2.0_(Harvey)/07%3A_Collecting_and_Preparing_Samples/7.7%3A_Liquid%E2%80%93Liquid_Extractions.

[43] A. I. Khuri, J. A. Cornell, *Response Surfaces: Designs and Analyses, Vol. 152*, 2nd ed., Marcel Dekker, Inc, University of Florida, Gainesville, Florida, **1996**.

[44] R. M. Joyce, *Journal of Polymer Science Part C: Polymer Letters* **1984**, *22*, 372.

[45] J. Frost. *Multiple Regression Analysis: Use Adjusted R-Squared and Predicted R-Squared to Include the Correct Number of Variables*. **2013** 13/06/2013 [cited 2017 09/15]; Available from: http://blog.minitab.com/blog/adventures-in-statistics-2/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables.

[46] B.-S. Chen, L. G. Otten, V. Resch, G. Muyzer, U. Hanefeld, *Standards in Genomic Sciences* **2013**, *9*, 175-184.

Appendix

A. GC data: retention time of the product

Figure A.1 displays the gas chromatogram of the purified 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one **2** (biotransformation product) with the retention time of 27.1 min and the retention time of the compound resulting from the natural equilibrium of the product in water, 4-methylfuran-2(*5H*)-one **7**, with the retention time of 18.1 min. The black line is corresponding to a GC vial only containing ethyl acetate. The retention time of these molecules was tested in previous unpublished studies.

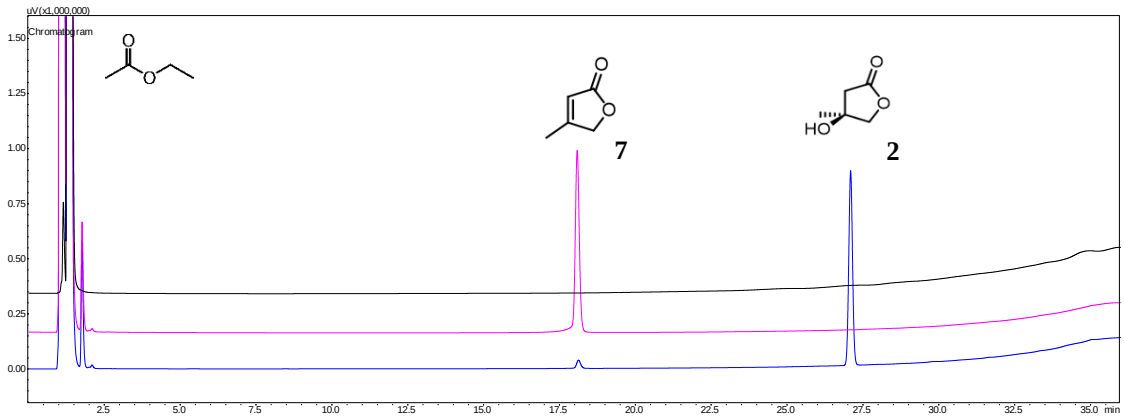


Figure A.1 Gas chromatogram of 4-hydroxy-4-methyldihydrofuran-2(*3H*)-one (2) and 4-methylfuran-2(*5H*)-one (7).

As it can be observed, the compound 7 is always presented in the chromatogram of compound 2, due to the explained natural equilibrium.

B. NMR Data

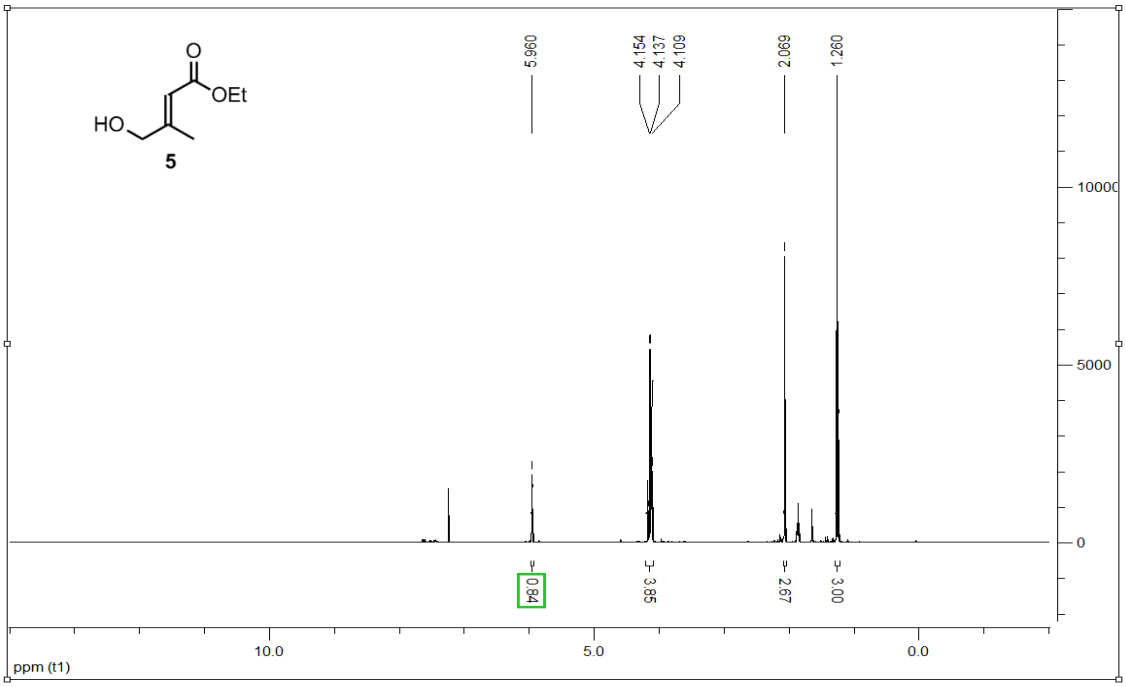


Figure B.1 ^{1}H NMR of ethyl-4-hydroxy-3-methylbut-2-enoate (5) (400 MHz, $CDCl_3$) δ: 1.26 (t, 3H), 2.07 (s, 3H), 4.11 (m, 2H), 4.15 (q, 2H), 5.96 (m, 1H). In accordance with literature.[11]

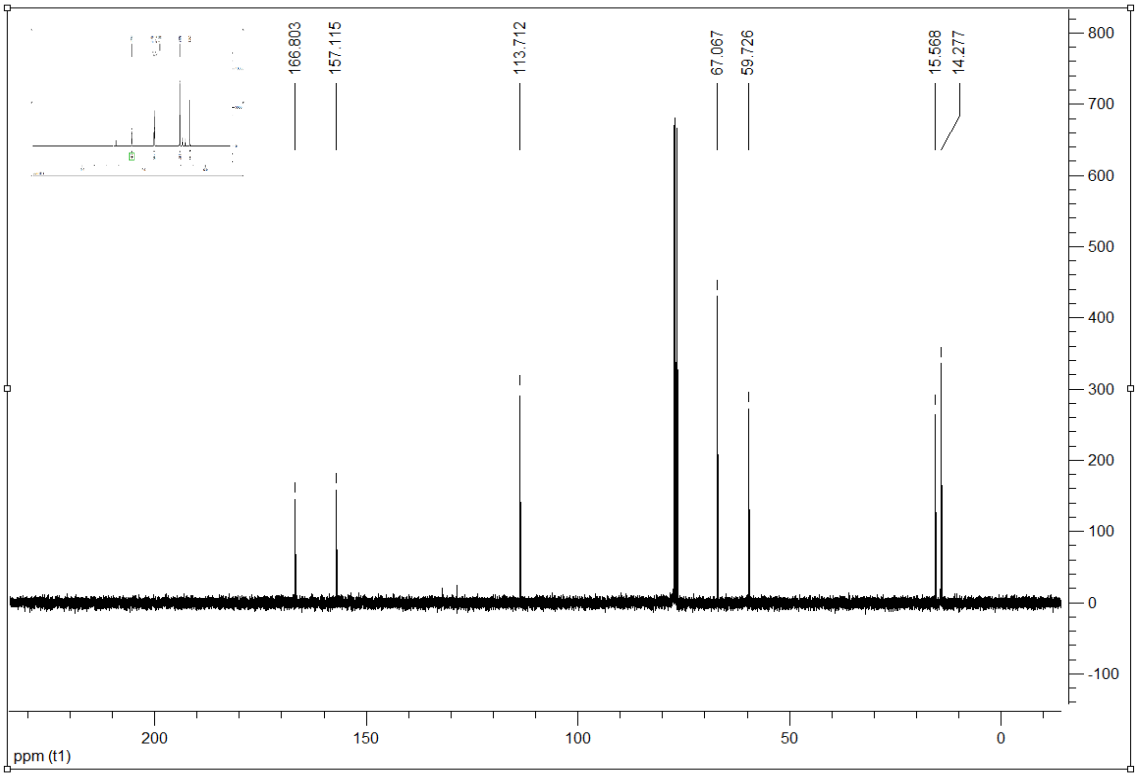


Figure B.2 ^{13}C NMR of ethyl-4-hydroxy-3-methylbut-2-enoate (5) (100 MHz, $CDCl_3$) δ: 14.3, 15.6, 59.7, 67.1, 113.7, 157.1, 166.8. In accordance with literature.[11]

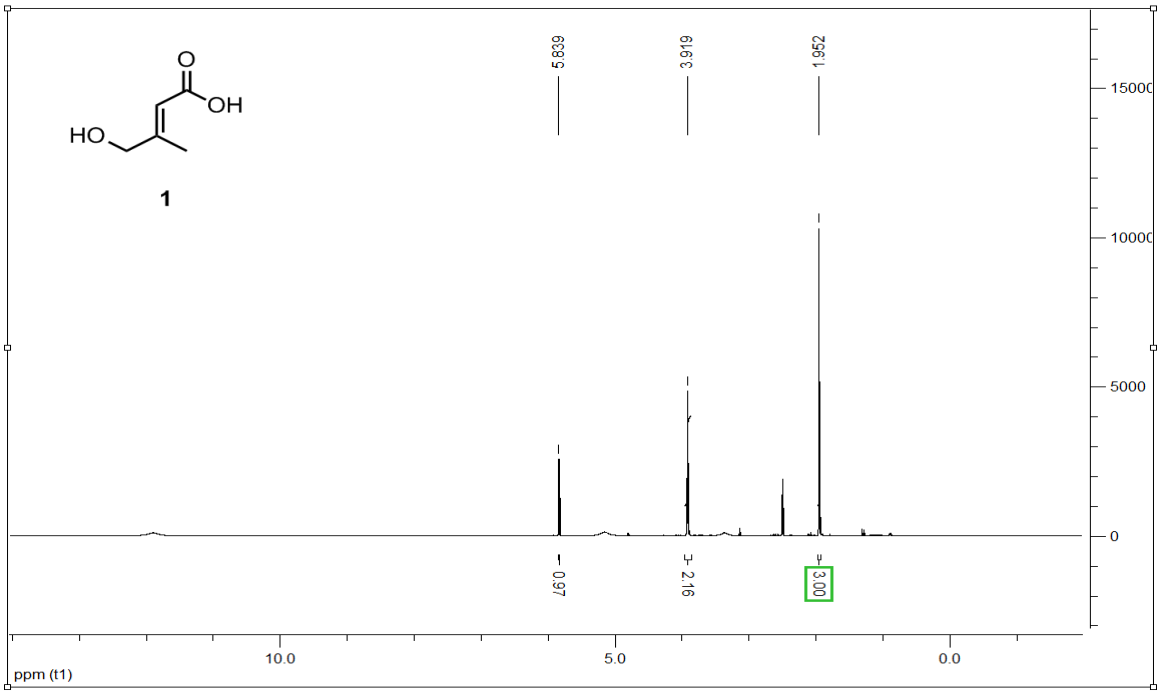


Figure B.3 ^{1}H NMR of (*E*)-4-hydroxy-3-methylbut-2-enoic acid (1) (400 MHz, DMSO-d_6) δ: 1.95 (s, 3H), 3.92 (s, 2H), 5.84 (s, 1H). In accordance with literature.[11]

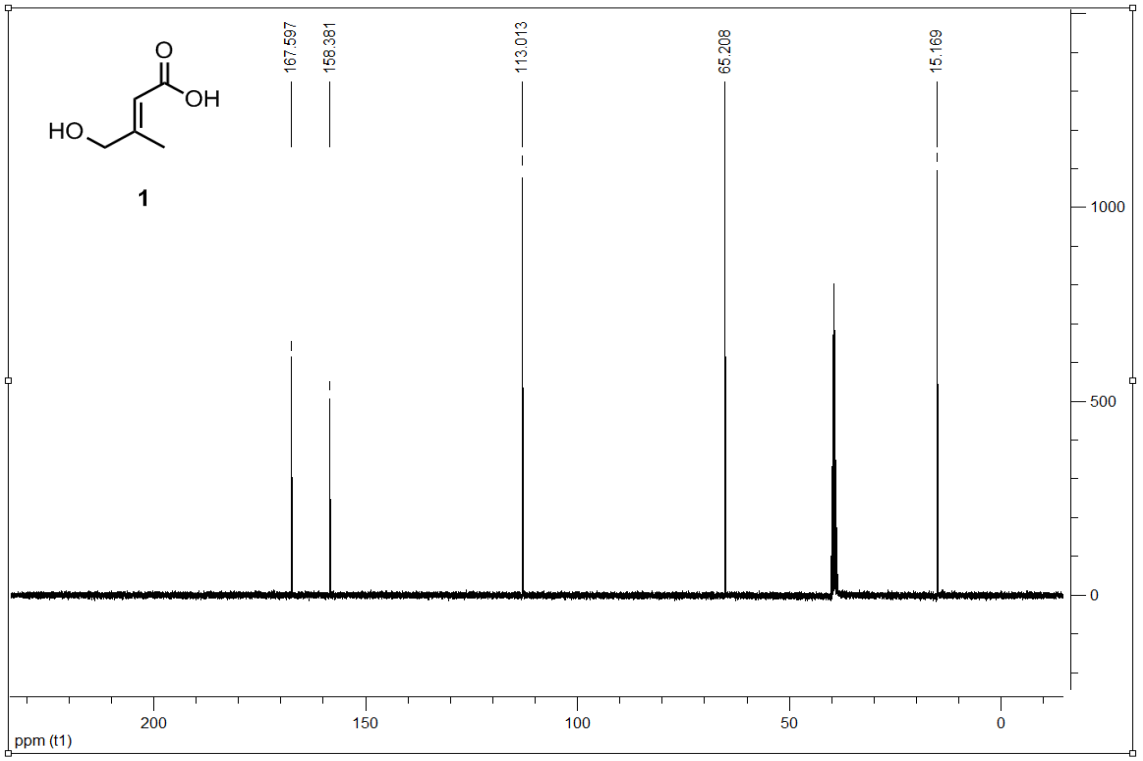


Figure B.4 ^{13}C NMR of (*E*)-4-hydroxy-3-methylbut-2-enoic acid (1) (100 MHz, DMSO-d_6) δ: 15.2, 65.2, 113.0, 158.4, 167.6. In accordance with literature.[11]

The following NMR data from 4-hydroxy-4-methyldihydrofuran-2(3H)-one **2** is of a standard obtained in earlier experiments (unpublished data). It was not possible to run a NMR due to the low yield of the product **2** on the preparative scale biotransformation. However, comparing the GC retention times with the stated in Appendix A, it can be confirmed that this is the right compound and the present NMR data are displayed additionally to the reader.

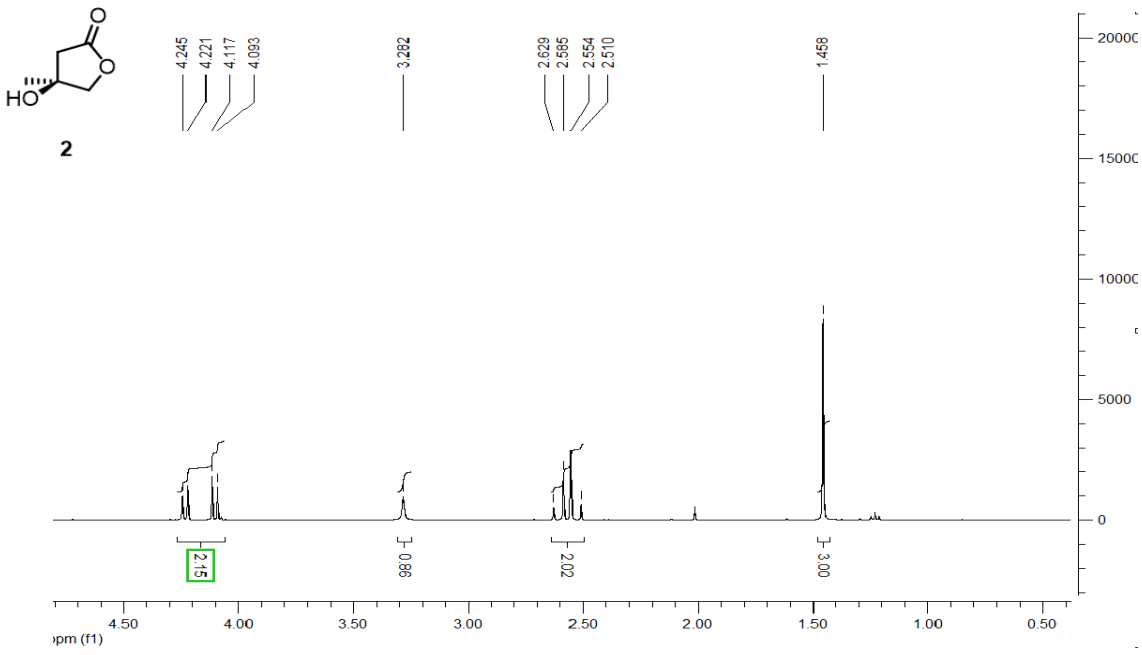


Figure B.5 ^{1}H NMR of 4-hydroxy-4-methyldihydrofuran-2(3H)-one (2) (400 MHz, DMSO-d_6) δ: 1.46 (3H,s), 2.51-2.63 (2H,ABq), 3.28 (1H,s), 4.09-4.24 (2H,ABq). In accordance with literature.[11]

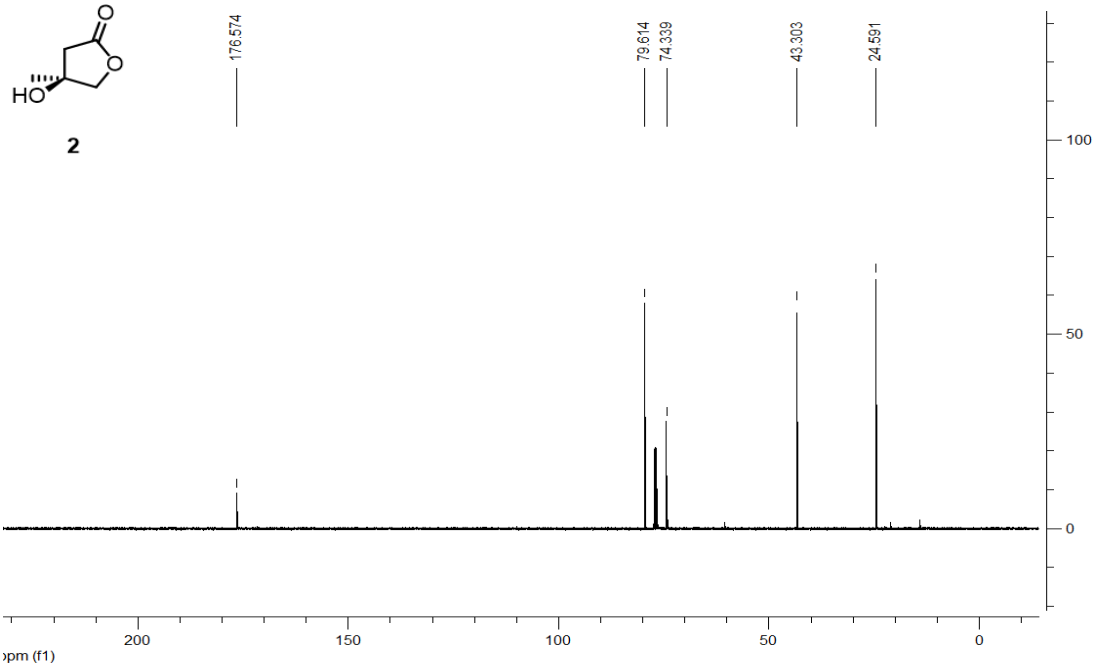


Figure B.6 ^{13}C NMR of 4-hydroxy-4-methyldihydrofuran-2(3H)-one (2) (100 MHz, DMSO-d_6) δ: 24.6, 43.4, 74.4, 79.7, 176.6. In accordance with literature.[11]

C. Preparative scale biotransformation

Figure C.1 Preparative scale continuous liquid-liquid extraction using the supernatant from biotransformation and DCM as extracting solvent.

Figure C.2 Zoom-in of the preparative scale extraction. Left: flask with supernatant from the biotransformation on the top and DCM on the bottom. Right: flask with DCM being heated in oil bath.

D. Precipitation of $FeSO_4 * 7 H_2O$

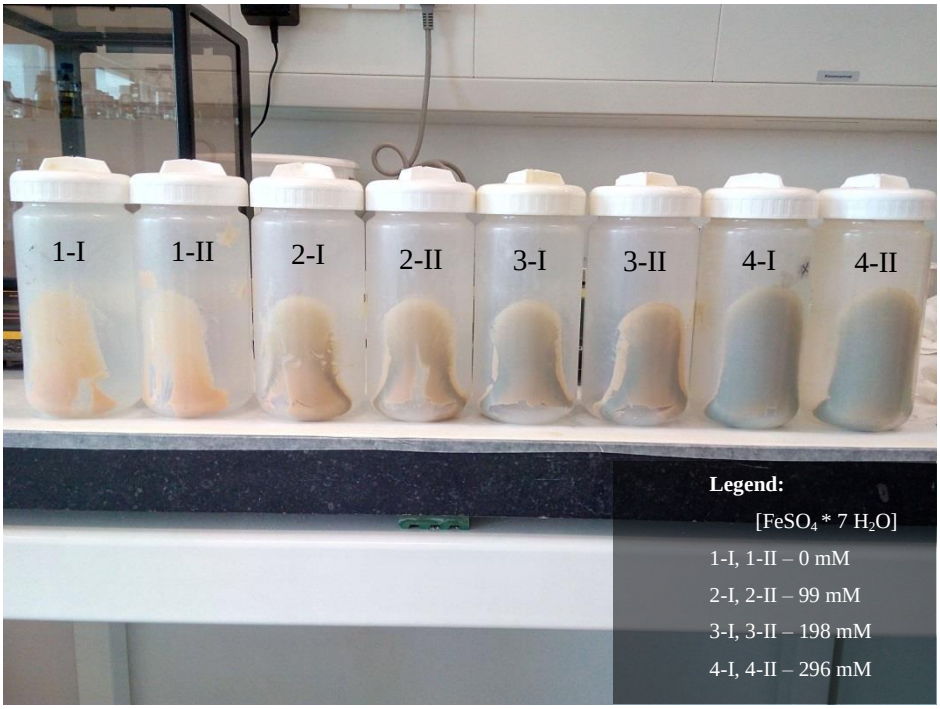


Figure D.1 Precipitation of iron sulphate heptahydrate in the centrifuge tubes after centrifugation.

E. OD_{600} and pH

Figure E.1 displays the OD_{600} analysis to a culture growing in a standard medium with no iron sulphate added. In figure E.2, a pH decrease can be observed at the same time the OD_{600} from figure E.1 increases, indicating that during the exponential phase of the growth, the culture medium becomes slightly more acidic. When the culture reaches the stationary phase, there is an insignificant increase of the pH, which remains within a smaller range until the end of the experiment.

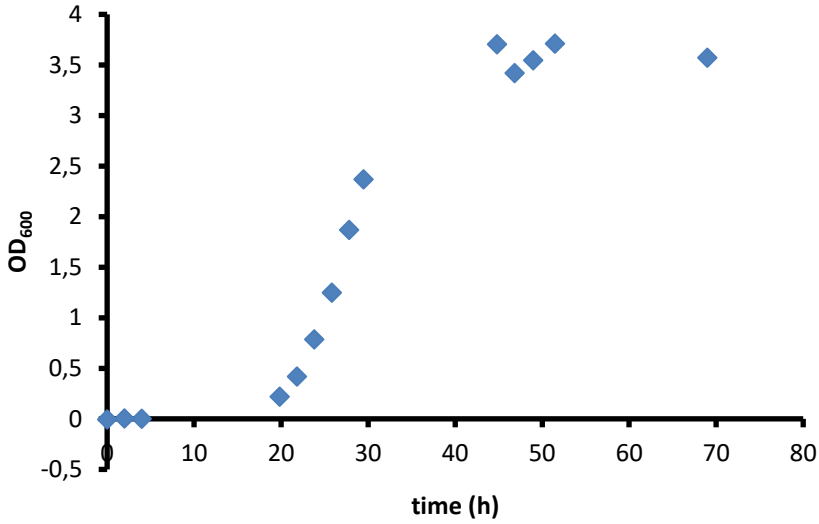


Figure E.1 OD_{600} measurements displaying the growth of *Rhodococcus rhodochrous* ATCC 17895 over a time course of 72 hours. Culture growing in a standard medium.

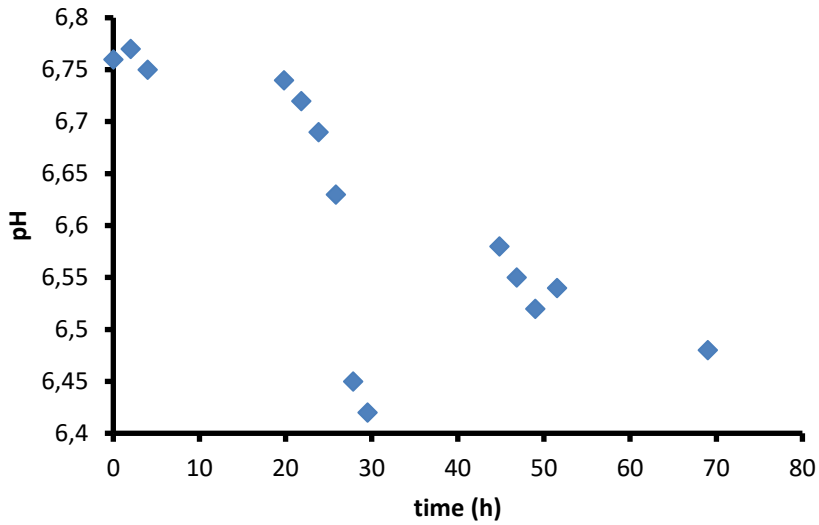


Figure E.2 Analysis of the pH of the standard medium in which the microorganism grew**.**

F. Transformation of data

The transformation of data is an important component of any data analysis. It applies a mathematical function to all the response data. Transformations may be needed to meet the assumptions that make the ANOVA valid, for instance the diagnosis that the residuals must be normally distributed with a constant variance. Also, transformation is needed if the error (residuals) is a function of the magnitude of the response (predicted values).[24]

Most data transformations can be described by the power function $\sigma = fn(\mu^{\alpha})$, where sigma (σ) is the standard deviation and mu (μ) is the mean. The power alpha (α), is given in the list above. Lambda (λ) is 1- α in all cases. If the standard deviation associated with an observation is proportional to the mean raised to the α power, then transforming the observation by the 1- α power gives a scale satisfying the equal variance requirement of the statistical model.[24]

The appropriate choice of a response transformation relies on subject matter knowledge and/or statistical considerations.[24] The transformations available from the Design Expert are shown in table F.1.

Table F.5.1 List of transformations available from the Design Expert.[24]

Transformation	Transformation	Requirements
Square root (λ= 0.5)	$y' = \sqrt{y + k}$	$(y + k) > 0$
Natural log (λ= 0)	$y' = \ln(y + k)$	$(y + k) > 0$
Base 10 log (λ= 0)	$y' = \log_{10}(y + k)$	$(y + k) > 0$
Inverse square root (λ= -0.5)	$y' = \frac{1}{\sqrt{y + k}}$	$(y + k) > 0$
Inverse (λ= -1)	$y' = \frac{1}{(y + k)}$	Response range, (low+k) to (high+k), must not include 0.
Power	$y' = (y + k)^{\lambda}$	$-3 < \lambda < 3$
Logit	$y' = \ln(\frac{y - lower}{upper - y})$	-
ArcSin square root	$y' = \sin^{-1}\sqrt{y}$	$0 < y < 1$

G. Studentised residuals and normal distribution

As displayed in equation G.1, the residual (e_i) is a difference between actual (y_i) and predicted ($\hat{y}_i$) values for each point, and can assume negative or positive values.[13]

Equation G.1 Definition of residual.[13]

$$e_i = y_i - \hat{y}_i$$

The studentised residual is the residual divided by the standard deviation (Std Dev) of that residual. The standard deviation of each residual can be estimated by the square root of the mean square of the residuals ($\sqrt{MS_{res}}$), and the mean square of the residuals is given as output of the statistical software and displayed in the ANOVA tables.[13] Therefore, equation G.2 displays the calculation of a studentised residual, r_i.

Equation G.2 Calculation of a studentised residual.[13]

$$r_i = \frac{e_i}{\sqrt{MS_{residual}}}$$

H. Fit summary

H.1 Sequential Model Sum of Squares

The Sequential Model Sum of Squares table shows the accumulating improvement in the model fit as terms are added. For example, the linear line (A, B, C) from table H.1. shows the significance of the terms after accounting for the mean and block terms (AB, AC, BC). The Quadratic line indicates the significance of adding the quadratic terms (A^2, B^2, C^2) to the linear, block and mean terms. Each line from table H.1. is not a complete model, but only the statistics for those additional terms.[24]

Table H.1 ANOVA of the sequential model sum of squares.[24]

Source	Sum of squares (SS_x)	Degrees of freedom (df)	Mean square (MS_x)	F value	Probability value $P_x > F$
Mean (vs Total)	2.27	1	2.27	-	-
Linear (vs Mean)	0.27	3	0.091	12.89	0.0002
2FI (vs Linear)	0.033	3	0.011	1.81	0.1943
Quadratic (vs 2FI)	**0.053**	**3**	**0.018**	**6.51**	**0.0102**
Cubic (vs Quadratic)	0.017	4	4.373E-003	2.73	0.1306
Residual	9.595E-003	6	1.599E-003	-	-
Total	2.65	20	0.13	-	-

The criteria of choice is the highest order model where the additional terms are significant. In Table H.1., it is the quadratic, bold and green highlighted.

H.2 Lack of fit test

Additionally, when the design has extra design points beyond what is needed for the model where points are replicated to provide an estimate of pure error (normally, the central points are replicated in this regard), a Lack of fit test of each model can be compared. As explained in section 2.3., the lack of fit compares the MS_{lof} with the MS_{pe} and is not desirable.[24] Thus, best choice is that where Lack of fit is insignificantly higher. In this case, is the cubic line, is it can be observed in table H.2. However, since the cub terms are aliased, the choice is the insigficantly higher lack of fit for a non-aliased model, which is the quadratic model.

Table H.2 ANOVA of the Lack of Fit test applied to each model.[24]

Source	Sum of squares (SS_x)	Degrees of freedom (df)	Mean square (MS_x)	F value	Probability value $P_x > F$
Linear	0.11	11	9.582E-003	5.92	0.0312
2FI	0.072	8	8.989E-003	5.55	0.0376
Quadratic	**0.019**	**5**	**3.798E-003**	**2.35**	**0.1855**
Cubic	1.500E-003	1	1.500E-003	0.93	0.3799
Pure Error	8.095E-003	5	1.619E-003	-	-

H.3 Model summary statistics

The Model Summary Statistics of models fit shows the standard deviation, the R-squared and adjusted R-squared, predicted R-squared and the PRESS statistic for each complete model. Low standard deviation, R-squared near 1 and relatively low PRESS are the best combination.[24]

Focus on the model maximizing the R-Squared, adjusted R-Squared and the Predicted R-Squared, and minimising the PRESS and standard deviation. As shown in table H.3, the quadratic model offers once again the best combination.[24]

Table H.3 Calculations of the different coefficients and PRESS for each model.[24]

Source	Std. Dev.	R-Squared	Adj. R-Squared	Pred. R-Squared	PRESS
Linear	0.084	0.7073	0.6524	0.5107	0.19
2FI	0.078	0.7936	0.6984	0.3995	0.23
Quadratic	**0.052**	**0.9301**	**0.8673**	**0.5433**	**0.18**
Cubic	0.040	0.9753	0.9216	0.1169	0.34

- **Additional information (PRESS)**

In statistics, the **predicted residual error sum of squares (PRESS) statistic** is a form of cross-validation used in regression analysis to provide a summary measure of the fit of a model to a sample of observations that were not themselves used to estimate the model. It is calculated as the sums of squares of the prediction residuals for those observations. The PRESS for the chosen model should be small relative to the other models under consideration.[24]