Thermochemical conversion of waste biomass (cherry rejects) to renewable platform chemicals: FUR and HMF

Master of Chemistry

Renewable Chemistry

Author: Heidi M. Hammer Mayhew

Supervisor: Tanja Barth

Co-supervisor: Camilla Løhre

University of Bergen

Department of Chemistry

Acknowledgements

First and foremost, I would like to thank my supervisor Tanja Barth for excellent guidance and encouragement throughout my past years as a student at the University of Bergen. Secondly, thank you to my co-supervisor Camilla Løhre for all the help and proof-reading of my thesis.

Thank you to Bjarte Holmelid for all the training and help with the HPLC-MS/UV instrument and subsequent analyses. I would also like to thank everybody on my research group for your help and second opinions when I've encountered problems during my thesis and laboratory work.

Lastly, I would like to thank my family and friends for all the encouragement and support throughout my education.

Abstract

Furfural (FUR) and 5-hydroxymethylfurfural (HMF) are furanic compounds that can be derived from carbohydrates in biomass. FUR and HMF can potentially serve as essential biobased platform chemicals due to their diverse applications as reaction intermediates. As platform chemicals they can be upgraded into a plethora of different fuels and chemicals currently produced from fossil-derived resources. In this thesis, the thermochemical conversion reaction of carbohydrates in waste biomass (cherry rejects) to FUR and HMF was optimized. This was done with the use of a biphasic reaction system developed according to the principles of green chemistry. The influence of temperature, residence time, and volume of aqueous/organic solvents on FUR and HMF yield was investigated by utilizing experimental designs. The greatest FUR yield obtained in this thesis was 1.16m%, the experiment which generated this result was conducted with a temperature of 210°C, residence time of 30 minutes, and an organic-to-aqueous ratio of 1:4. The greatest HMF yield obtained in this thesis was 20.04m%, the experiment which generated this result was conducted with a temperature of 230°, residence time of 30 minutes, and an organic-to-aqueous ratio of 1:1. The yields were quantified with a developed qNMR method. Additionally, a HPLC-MS/UV method was developed for separation and quantification of FUR, HMF, and three other selected compounds (levulinic acid, 5-methylfurfural, and 2-acetylfuran). This analytical method was developed to investigate the presence of other compounds in addition to FUR and HMF in the product solutions of the conversion reaction. The results in this thesis have demonstrated that waste biomass (specifically cherry rejects) could serve as a promising biobased feedstock for the production of fuels and chemicals via renewable platform chemicals, FUR and HMF, that have traditionally been produced from fossil-derived resources.

List of contents

Abbreviations

Abbreviation	Meaning
ADC	Analog-to-digital converter
CAGR	Compound annual growth rate
ESI	Electrospray Ionization
FDCA	2,5-Furandicarboxylic acid
FD	Factorial Design
FFD	Fractional Factorial Design
FT	Fourier Transformation
FUR	Furfural
HMF	5-Hydroxymethylfurfural
HPLC	High-Performance Liquid chromatography
HPLC-MS/UV	High-Performance Liquid chromatography with both mass- and UV-Vis detector
MS	Mass spectrometry
PC	Principal component
PCA	Principal component analysis
PEF	Polyethylene furanoate
PET	Polyethylene terephthalate
PLS	Partial least square
qNMR	Quantitative Nuclear Magnetic Resonance
RF	Radio Frequency
TIC	Total Ion Chromatogram
UV-Vis	Ultraviolet/Visible light

1. Introduction

In the past 50 years, both the world population and the primary energy consumption has increased substantially. The world primary energy consumption grew by 31 exajoules during 2021, this is the largest increase in history (BP, 2022). Fossil-derived energy resources have been shown to cause emissions of GHG gases that are harmful to the environment, and the fossil fuel deposits are finite and decreasing. Despite this, it is still the main global energy source with an estimated market share of 81% in 2020 (Sundaram, 2015, IEA, 2020). Chemical products are also strongly dependent on fossil resources. Consequently, the world is now facing two major challenges: energy crisis and environmental pollution (Gupta and Tuohy, 2013).

Society's acknowledgement of the issues posed by fossil energy has led to financing of research and development for clean and sustainable energy resources. This in turn has led to many publications, policies and agreements regarding global energy and climate. This includes the development of the 12 principles of green chemistry (Anastas and Warner, 1998), the 17 sustainable development goals (SDGs) by UN and the subsequent signing of the Paris Agreement on Climate Change in 2015.

There is an increasing global recognition that biomass feedstocks have the potential to replace a large fraction of fossil feedstocks for industrial productions. This will address both the energy and non-energy (chemicals and materials) sectors (Cherubini, 2010). The recognition of biomass as a renewable resource has been a major driving force for the advancement of biorefineries. A biorefinery with the aim of producing bulk chemicals from biomass will need a selection of platform molecules from which all desired chemicals can be derived (Cherubini, 2010). Furanic compounds like furfural (FUR) and 5-hydroxymethylfurfural (HMF) produced from sustainable biomass sources can potentially serve as platform chemicals due to their diverse applications as reaction intermediates (Huber et al., 2006, Danon et al., 2014, Millan et al., 2019, Liu et al., 2020). FUR and HMF may therefore play a major role in providing the future with biobased, sustainable energy.

2. Background

2.1 Green chemistry

Green chemistry is sustainable chemistry with the goal of developing chemical processes that are safe, efficient, economical, and renewable (Mohrig et al., 2014). The twelve principles of green chemistry were developed in 1998 by Paul Anastas and John Warner (Anastas and Warner, 1998). These principles govern the implementation of greener chemical practices today and are listed in table 2-1.

Table 2-1. List of the 12 principles of green chemistry. Taken from Mohrig et al. (2014).

1. Prevention:	Avoid generating waste
2. Atom Economy:	Incorporate most atoms from the reagents into the product
3. Less Hazardous Chemical Syntheses:	Use and generate the least toxic materials
4. Designing Safer Chemicals:	Ensure that final products are nontoxic
5. Safer Solvents and Auxiliaries:	Use minimal and innocuous supporting materials
6. Design for Energy Efficiency:	Minimize energy requirements
7. Use of Renewable Feedstocks:	Whenever possible, use renewable raw materials
8. Reduce Derivatives:	Avoid introducing atoms that have to be removed later
9. Catalysis:	Use catalysts for efficient and less wasteful processes and re-use them when possible
10. Design for Degradation:	Plan for products to break down naturally into benign substances
11. Real-Time Analysis for Pollution Prevention:	Monitor the process to avoid accidental exposure to hazards
12. Inherently Safer Chemistry for Accident Reduction:	Avoid using chemicals that are highly reactive

Green chemistry has changed the mindset of today's chemists. When creating an innovative product, they are required to think beyond the production of the product and evaluate its entire life span, "cradle to grave". Not only the raw materials need to be evaluated, but also the process for making it, its lifetime, and disposal.

Principle number 7 of green chemistry states that renewable raw materials should be used whenever possible. This has motivated both the scientific and industrial industries to put an effort into finding potential renewable bio-based feedstocks that can replace petroleum-based feedstock.

2.2 The Biorefinery Concept

In an effort to minimize the use of non-renewable pollution-causing energy sources, there are several clean and renewable resources to consider. Wind, solar and hydropower are examples of renewable energy sources that can, to some extent, replace fossil-based resources. However, the production of chemicals is also strongly dependent on the petrochemical industry (Kumar and Verma, 2021, Cherubini, 2010, Hülsey, 2018). Consequently, fossil-based resources need to be replaced by clean, renewable sources for both energy and chemical products.

A biorefinery has been defined by IEA as "the sustainable processing of biomass into a spectrum of marketable products and energy" (Cherubini, 2010). For this reason, biorefineries based on renewable biomass sources have the potential to fully replace the fossil-based petrochemical industry. The structure of biorefineries share some similarities with conventional oil refineries (figure 2-1). The most important difference between them is that conventional oil refineries has been in development for more than 150 years, whereas biorefineries are much younger and less developed. It is clear that further technological advancements are necessary to develop the full potential of biomass resources (Hülsey, 2018).

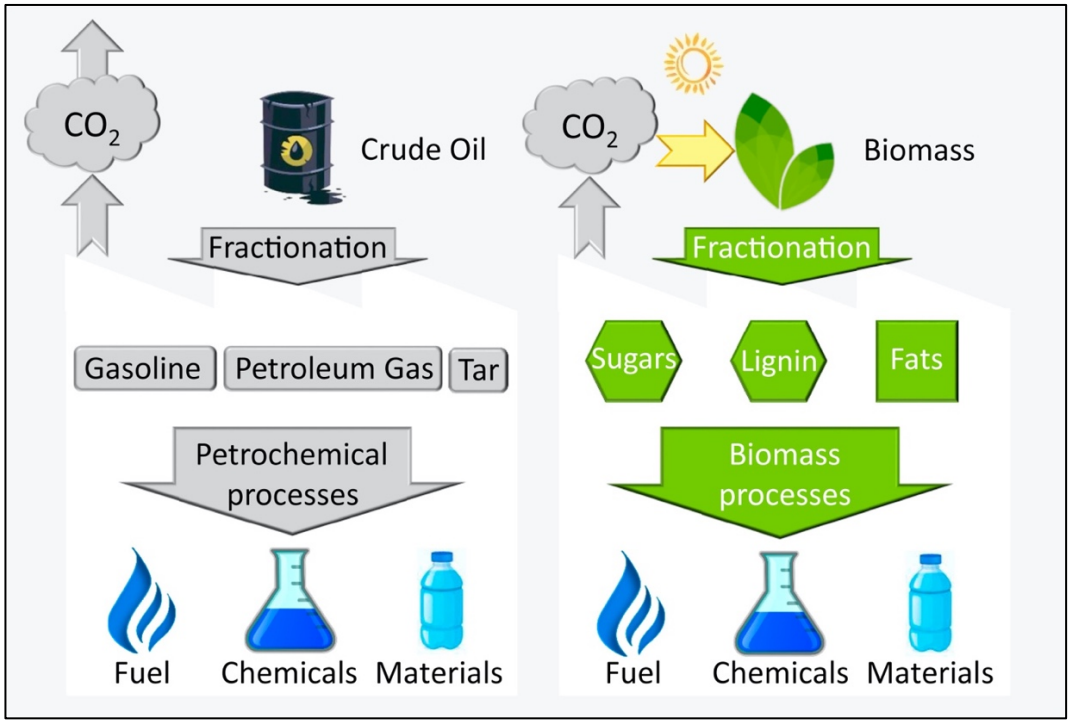


Figure 2-1: Demonstrating the similarities between a biorefinery and a conventional oil refinery. Taken from Hülsey (2018).

Although biorefineries can use sustainable feedstock sources, this is not enough to ensure that the final products are green and sustainable. When developing methods and techniques for future biorefineries it is important that the environment impact is minimized and that greener methodologies are used (Cherubini, 2010). Suitable methods and techniques will depend on which type and generation of biofeedstock used.

2.3 Biofeedstocks

2.3.1 First generation feedstock

First generation feedstocks include sugar, starch, vegetable oil and animal fats. This feedstock is usually taken from raw materials like wheat, corn, sugarcane, and sugar beet. The most common first generation biofuels produced from this type of feedstock are bioethanol, biomethanol, biodiesel, and starch-derived biogas (Cherubini, 2010).

First generation biofuels are one of the most commercially available types of biofuels (Acheampong et al., 2017). World production of bioethanol reached 112.7 billion litres, and world biodiesel production reached 43.4 billion litres in 2019 (IEA, 2021). The high sugar or oil content in first generation biofeedstock is advantageous because of their easy conversion into biofuel.

There is an ethical dilemma regarding production of first generation biofuels, because they are in competition with food and feed for biomass and land use. However, biorefineries can utilize waste biomass as feedstock, thereby helping in waste management along with simultaneously meeting energy demands (Kumar and Verma, 2021). This is the case for the feedstock used in this thesis.

A short introduction of second- and higher generation feedstocks are included for the sake of their relevancy to the biorefinery concept and its future development.

2.3.2 Second generation feedstock

Second generation biofuels are produced from lignocellulosic feedstock such as residues from agriculture, forestry, and industry (Cherubini, 2010). Contrary to first generation biofuels, this type of biofuel utilizes or contributes to the utilization of the entire plant, not just the grains, sugars, or fats. This means that the energy yield per hectare of land can be much higher (Acheampong et al., 2017).

2.3.3 Third and fourth generation feedstock

Third generation biofuels are produced from algae feedstock specially engineered through advanced biotechnology (Acheampong et al., 2017). The algae are a high-energy and entirely renewable feedstock and are predicted to produce more energy per acre of land than possible by conventional crops like sugarcane and corn (Chisti, 2007). They are however limited in commercial availability as they are still under broad research.

Fourth generation biofuels are produced from algae undergone metabolic engineering and are expected to bring essential advances in the biofuels field (Acheampong et al., 2017). This technology is an emerging field, and it is dependent on revolutionary advancements in synthetic and system biology for further development.

2.4 Furfural (FUR)

Furfural (FUR) or furan-2-carbaldehyd is an aldehyde that consists of a furan with the hydrogen at position 2 substituted by a formyl group. The chemical structure of both furan and FUR is displayed in figure 2-2. It appears as a colourless liquid with a reddish-brown tint. It is a naturally occurring compound found in agricultural residues like corn cobs, oat hulls and sugar cane bagasse (Eseyin and Steele, 2015). Table A-1 in Appendix A summarizes the physical and chemical properties of FUR.

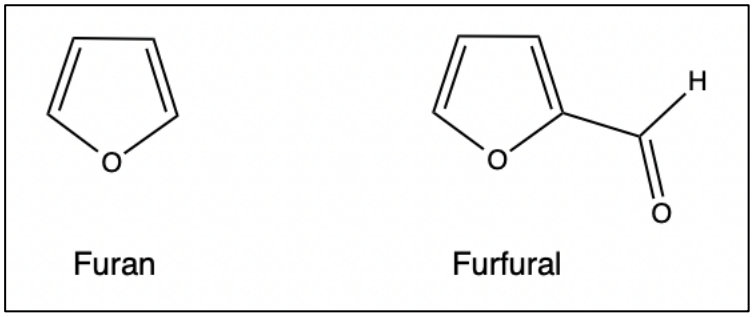


Figure 2-2. The chemical structure of furan and furfural (FUR).

2.4.1 Mechanistic pathways to FUR

When processing lignocellulosic biomass into biofuels, and other potential biorefinery products, the biomass is pre-treated. This is necessary because of its complex structure. Lignocellulosic materials consist mainly of three polymers: cellulose, hemicellulose, and lignin. “These polymers are associated with each other in a hetero-matrix to different degrees and varying relative composition depending on the type, species, and even source of the biomass” (Bajpai, 2016, p. 7) (figure 2-3).

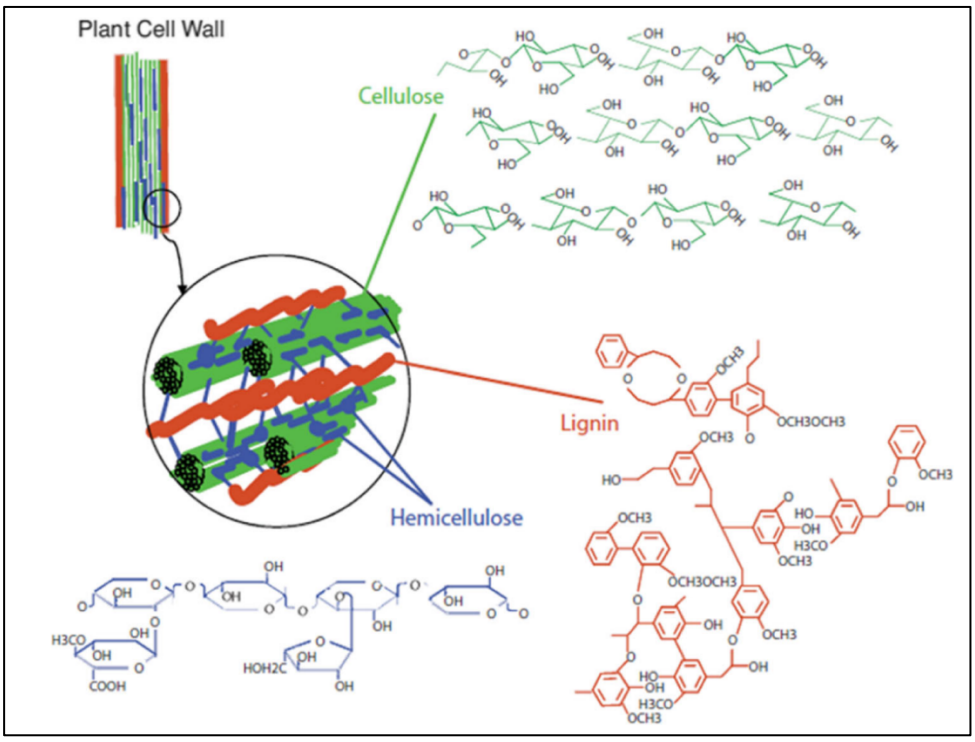


Figure 2-3. The structure of the three polymers comprising lignocellulosic biomass: Cellulose, hemicellulose, and lignin. Taken from Wang et al. (2019).

The pre-treatments are carried out with the aim to break down the complex structure of lignocellulose, reduce the cellulose crystallinity, and preserve pentoses from hemicellulose, improving the sugar availability (Xu et al., 2020). The biomass pre-treatment, especially when involving acid and/or temperatures above 160-180°C, induces the formation of several degradation products. FUR and HMF are considered the main degradation products from the carbohydrate fraction of the biomass (Rasmussen et al., 2014).

Hydrothermal treatment is often chosen because it is environmentally friendly compared to chemical pre-treatment, as it uses only water for reaction medium without additional chemicals (Jeong and Lee, 2015). During hydrothermal treatment of biomass, pentoses are released from hemicellulose. FUR is formed from pentoses, mainly D-xylose and L-arabinose (figure 2-4).

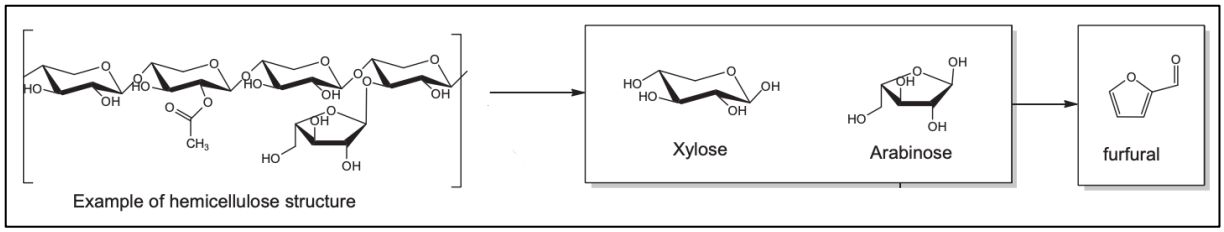


Figure 2-4: Suggested summary reaction route for the degradation of hemicellulose to FUR during hydrothermal treatment of biomass. Adapted from Rasmussen et al. (2014).

There are three different suggested mechanisms for the degradation of xylose to FUR. One of the mechanisms include a ring-opening of xylose, an acyclic mechanism, and the two others are direct cyclic mechanisms (figure 2-5). These mechanisms are supported by experimental results (Ahmad et al., 1995, Antal Jr. et al., 1991), this indicates that the degradation of xylose is a complex process and may occur according to several mechanisms.

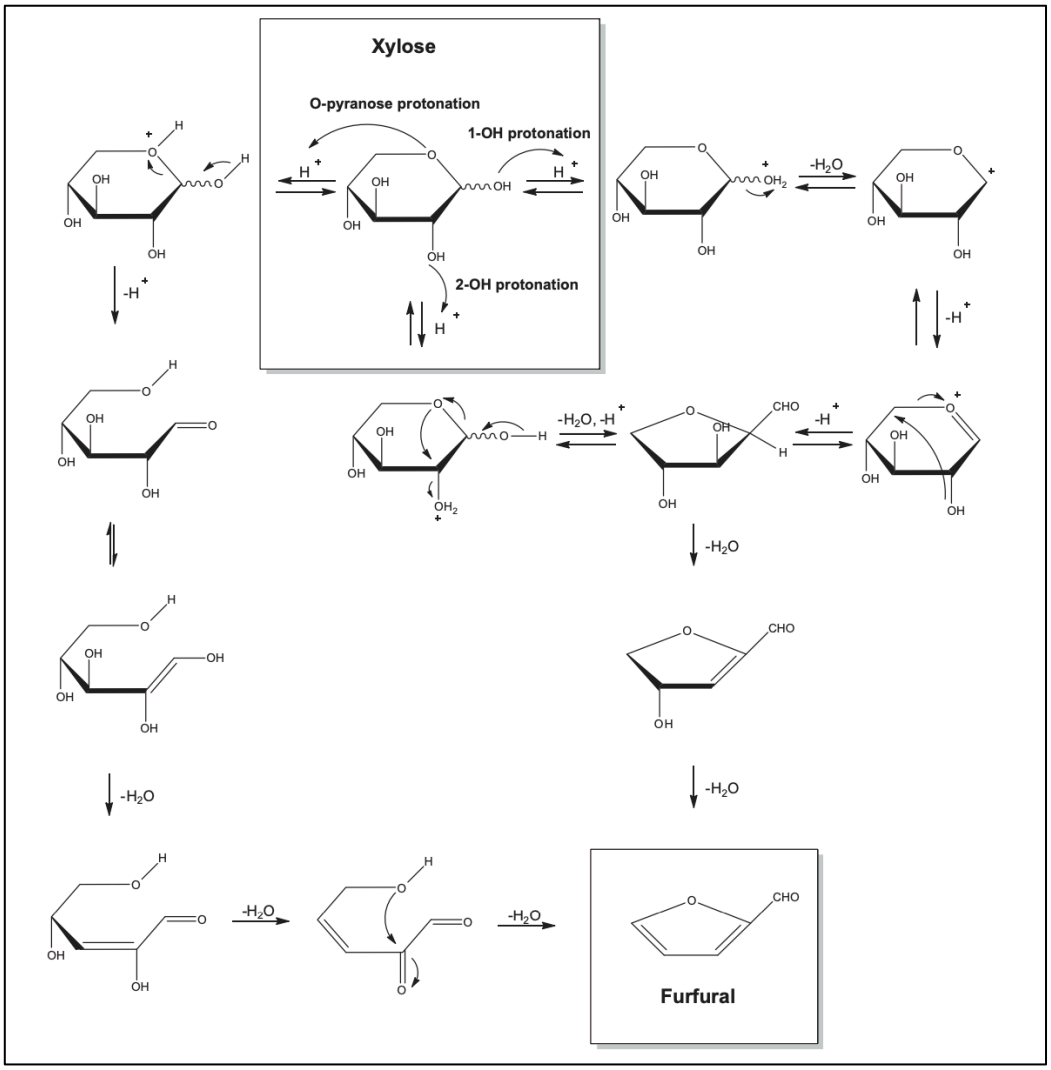


Figure 2-5. Suggested mechanisms for FUR formation from xylose, resulting from different protonation sites. Taken from Rasmussen et al. (2014).

With the use of ab initio molecular dynamic simulations, the rate-limiting step in the degradation reaction of xylose has been found. This is the protonation of a hydroxyl group on the xylose molecule or direct protonation of the pyranose oxygen (the oxygen atom in the six-membered ring). Ab initio molecular dynamic simulations and quantum mechanics modelling combined with NMR with no solvent effects (in vacuum) results in the reaction scheme outlined in figure 2-6 (Rasmussen et al., 2014). Hence, the degradation product formed depends on the protonation site.

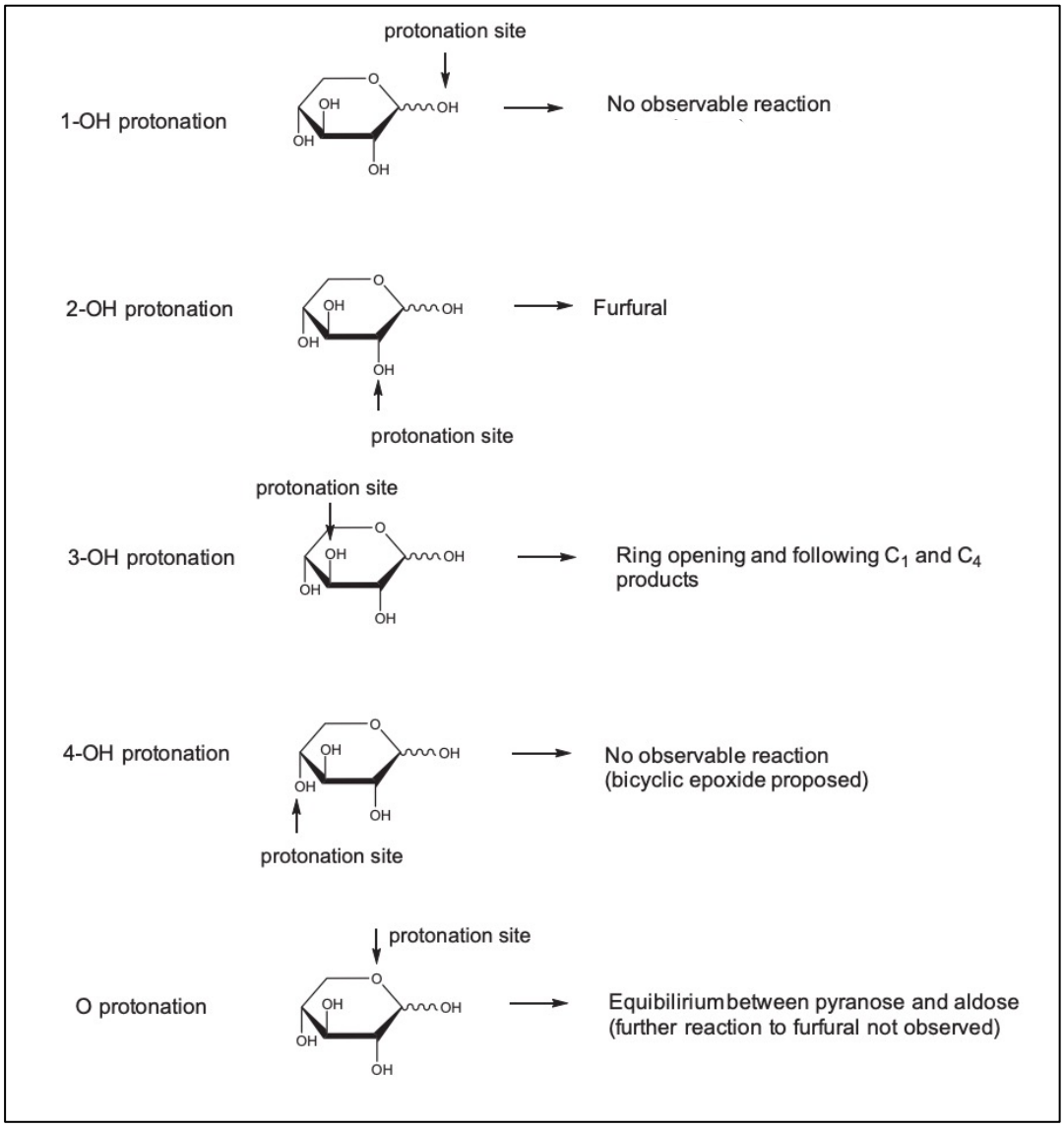


Figure 2-6. The different degradation products resulting from different protonation sites on xylose. Determined from ab initio molecular dynamic simulations and quantum mechanics modelling with no solvent water. Adapted from Rasmussen et al. (2014).

According to the models and simulations (summarised in figure 2-6) the furfural formation from the acyclic mechanism followed by protonation of the pyranose oxygen (shown in figure 2-5), does not occur. Protonation of the pyranose oxygen only results in an equilibrium between pyranose and aldose. Hence, the models contradict the experimental data that support furfural formation via this acyclic mechanism (Ahmad et al., 1995, Antal Jr. et al., 1991). Via this modelling, furfural formation will only arise from the 2-OH protonation of xylose (figure 2-5). Since water molecules compete for protons and hydrogen bond to the hydroxyl groups, the solvent water structure is crucial for the protonation site. In addition, reaction conditions (pH, solvent, salts etc.) can alter the water molecule surroundings and

therefore the protonation site (Rasmussen et al., 2014). This suggests that the protonation site, and resulting degradation product, can be chosen by controlling the reaction conditions.

It has also been hypothesised that xylose could degrade via a furan isomerisation intermediate (xylulose) that further dehydrates to FUR (figure 2-7).

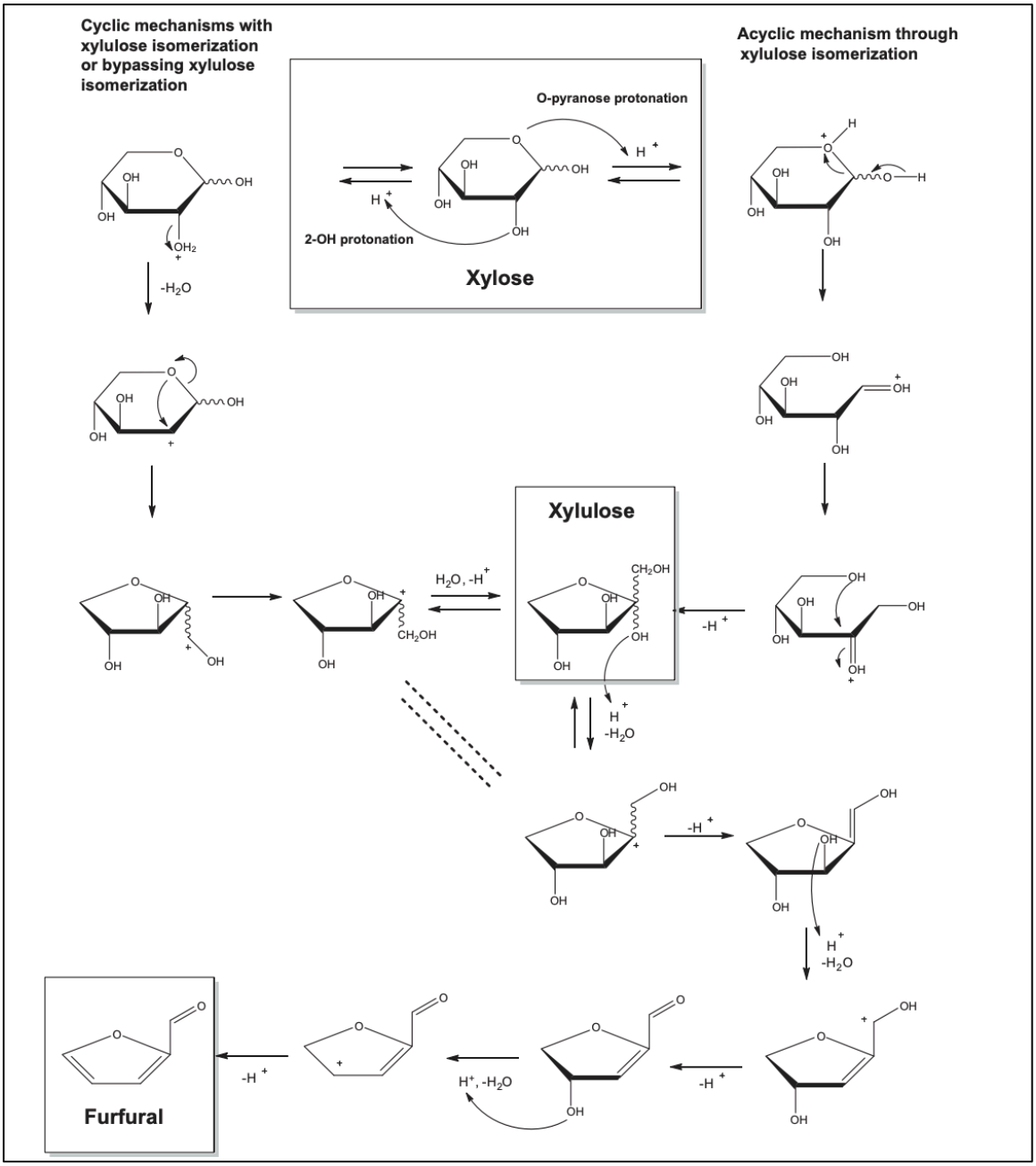


Figure 2-7. A hypothesised new mechanism for degradation of xylose to FUR.
Taken from Rasmussen et al. (2014).

2.5 5-Hydroxymethylfurfural (HMF)

5-Hydroxymethylfurfural (HMF) or 5-hydroxymethyl-2-furaldehyde consists of a furan substituted at positions 2 and 5 by a formyl and a hydroxymethyl group, respectively. The chemical structure of HMF is displayed in figure 2-8. HMF is solid at room temperature. It is a member of furans, an arenecarbaldehyde, and a primary alcohol. HMF is present in many different foods that are subjected to thermal treatment. Examples are honey, coffee, and tomato paste (Sevenich et al., 2020). Table A-2 in Appendix A summarizes the physical and chemical properties of HMF.

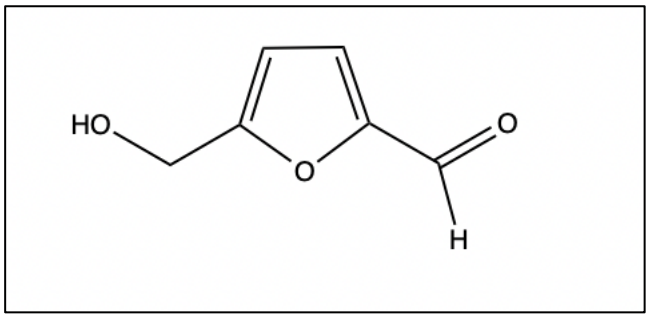


Figure 2-8: The chemical structure of HMF.

2.5.1 Mechanistic pathways to HMF

As mentioned in section 2.4.1, when biomass undergoes pre-treatment, HMF and FUR are considered the main degradation products from the carbohydrate fraction of the biomass (Rasmussen et al., 2014). The chosen pre-treatment method is often hydrothermal treatment because it is a more environmentally friendly approach compared to other methods (Jeong and Lee, 2015).

During hydrothermal treatment of biomass, D-glucose releases from cellulose. D-glucose will then dehydrate to HMF, which can further degrade to levulinic acid and formic acid (Yang et al., 2012) (figure 2-9). The monomeric substituents in hemicellulose include both hexoses and pentoses, hemicellulose may therefore also degrade to HMF. Further degradation of HMF into levulinic acid and formic acid can then take place (Rasmussen et al., 2014) (figure 2-9).

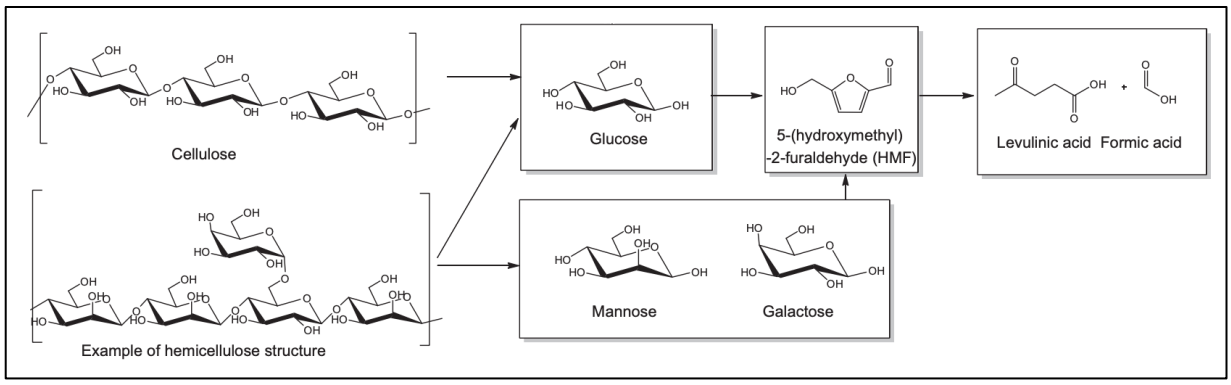


Figure 2-9. Suggested summary routes for the degradation of cellulose and hemicellulose to HMF, levulinic acid and formic acid during hydrothermal treatment of biomass.
Adapted from Rasmussen et al. (2014).

The formation of HMF from glucose is complex and several mechanisms and reaction routes have been suggested. The mechanisms suggested are both cyclic and acyclic, and they can both go via fructose isomerisation or without (figure 2-10). When glucose is protonated at the pyranose oxygen, ring opening will occur. This is the start of the acyclic mechanism. After the ring has opened, it's followed by either the aldose acyclic mechanism (Antal Jr. et al., 1991) or rearrangement and ring closure to fructose (Ahmad et al., 1995) (figure 2-10). Regardless of which mechanism takes place, the compounds will further dehydrate in a series of steps to form HMF (figure 2-10).

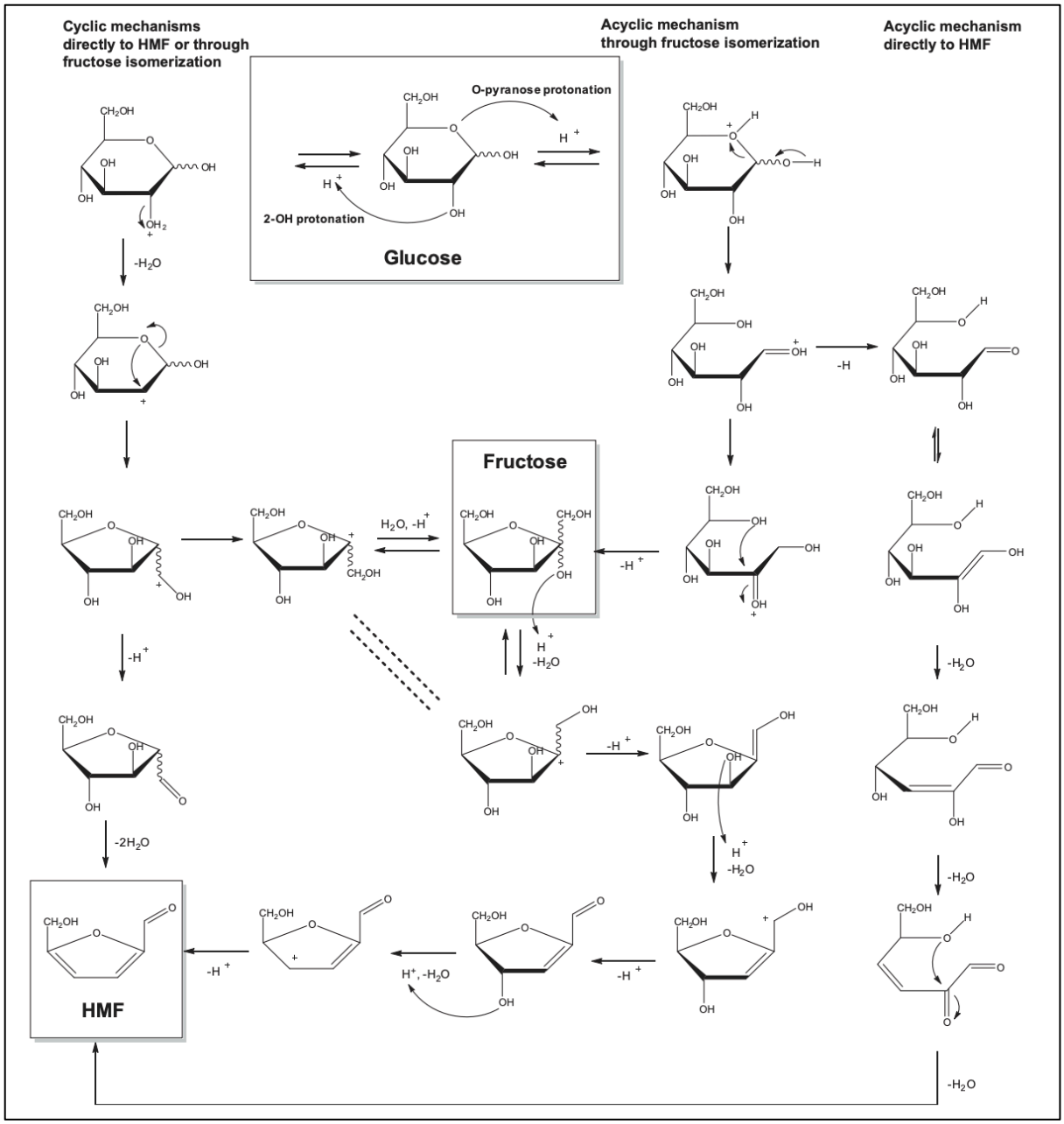


Figure 2-10. Formation of HMF from glucose through both cyclic and acyclic mechanisms with and without fructose isomerisation. Protonation of the pyranose oxygen leads to acyclic mechanism and protonation of 2-OH leads to the cyclic mechanism. Taken from Rasmussen et al. (2014).

The 2-OH protonation of glucose initiates the cyclic mechanism. This is followed by formation of fructose via a furanose tertiary carbocation or by formation of HMF via a furanose secondary carbocation without isomerisation to fructose (figure 2-10). Isomerisation to fructose has been given in situ NMR support for hydrothermal conditions, but no revealing of an acyclic or cyclic mechanism prior to isomerisation has been determined (Rasmussen et al., 2014). The

isomerisation of glucose to fructose with the closed ring structure is supported by Car-Parrinello molecular dynamics metadynamic simulations (CPMD-MTD) in the gas phase (Rasmussen et al., 2014). The simulation may however change drastically when water solvent molecules are incorporated in the model (Qian et al., 2005). The acyclic isomerisation to fructose is supported by quantum chemical calculations with density functional theory in water (Yang et al., 2012). The direct cyclic mechanism for HMF formation with no fructose intermediate is supported by CPMD-MTD simulations in water, but there is still a lack of experimental evidence (Rasmussen et al., 2014).

The dotted lines in figure 2-10 indicates another hypothesised cyclic mechanism without a fructose intermediate. This mechanism is a possibility because of the formation of the same tertiary fructose carbocation both before and after fructose formation (Rasmussen et al., 2014). The protonation site on glucose determines the subsequent degradation products (figure 2-11), and it is also dependent on reaction conditions (Qian et al., 2005).

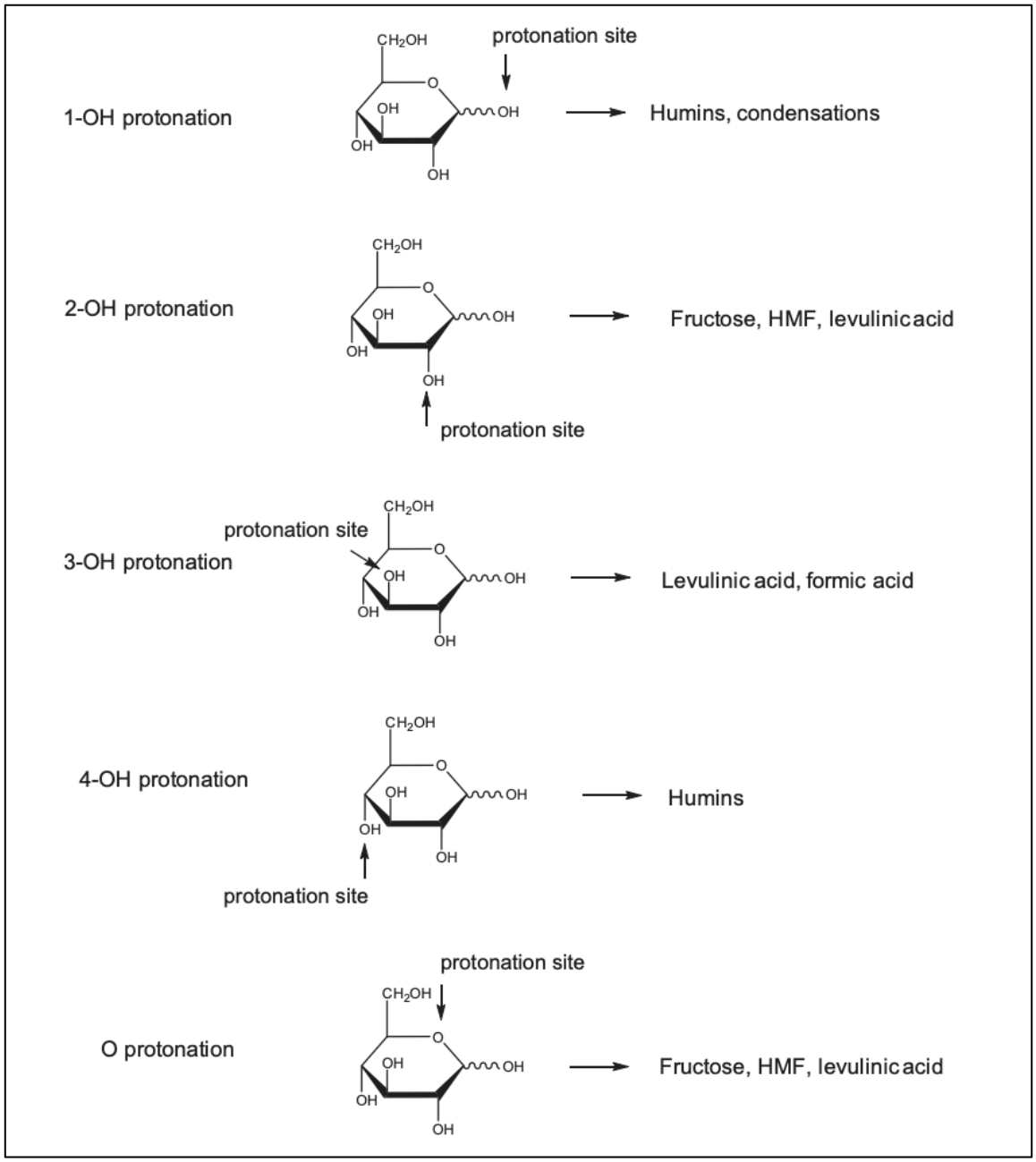


Figure 2-11. Different protonation sites on glucose and subsequent degradation products determined by ab initio molecular dynamics in water and quantum mechanics modelling with solvent water. Taken from Rasmussen et al. (2014).

There are some contradictions in the literature regarding the degradation products after protonation of glucose. For instance, Yang et al. only observed formation of levulinic acid as a result of 2-OH protonation with no formation of HMF (Yang et al., 2012). This suggests that levulinic acid has an alternative path of formation than only degradation of HMF. However, Qian suggested HMF as the only product formed after 2-OH protonation of glucose (Qian, 2012).

A new mechanism for HMF formation via 1-OH protonation of glucose has also been hypothesised (figure 2-12).

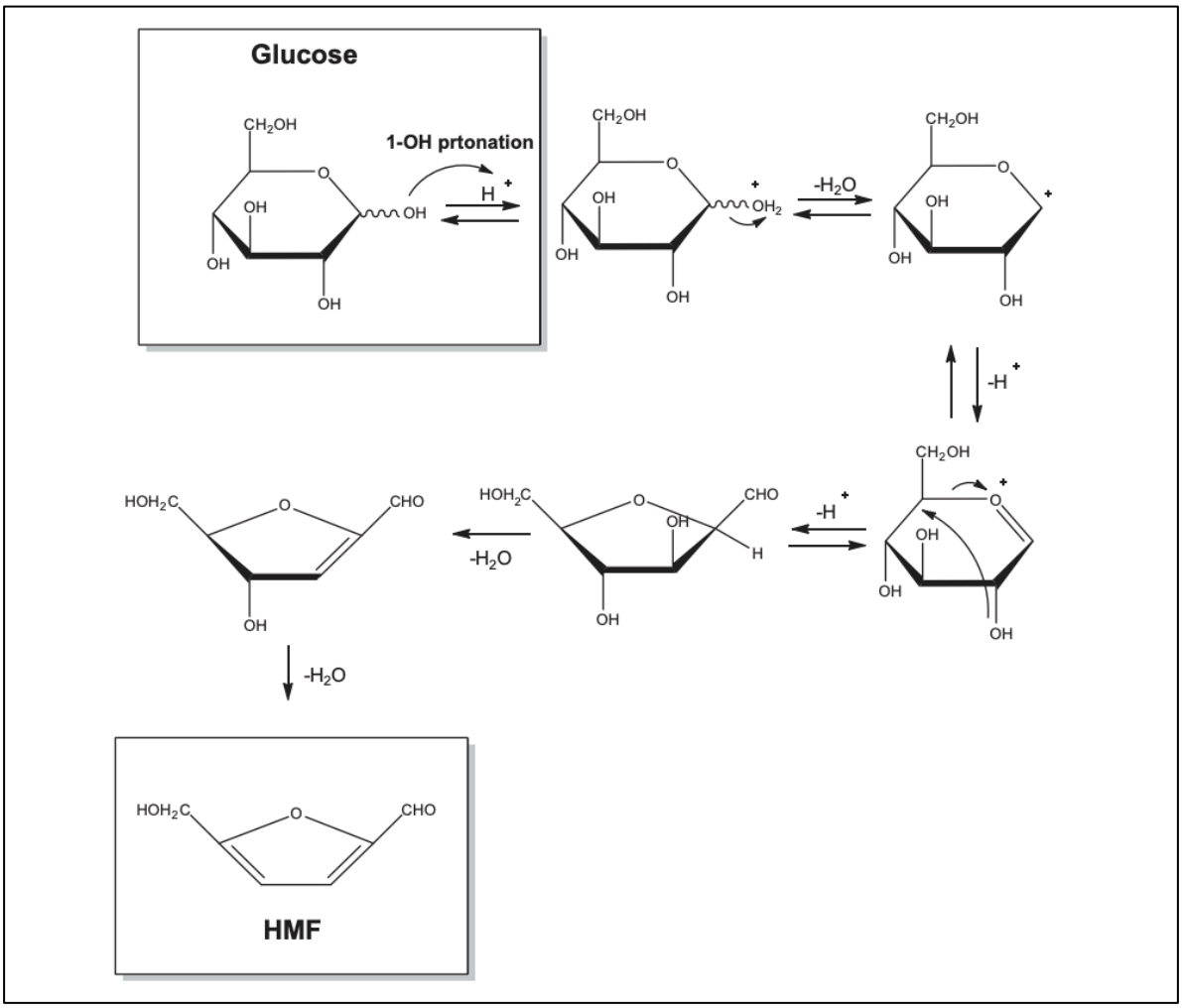


Figure 2-12. A new hypothesised mechanism for glucose degradation to HMF. Taken from Rasmussen et al. (2014).

2.6 Use of FUR and HMF as renewable platform chemicals

FUR and HMF are among the most used furan-based platform chemicals, this is because their chemical structure allows the preparation of various high-value-added chemicals and products. These bio-furan derivatives are generally obtained from the cellulose and hemicellulose fractions of biomass via acid-catalysed dehydration of their C6-C5 sugars (Xu et al., 2020). In 2004, the U.S. Department of Energy released a list of the most promising bio-derived platform molecules. Among them are FUR and HMF, which are receiving increased consideration for their versatile use as biobased platform chemicals (Werpy et al., 2004).

FUR and HMF can undergo many reaction processes in the presence of molecular hydrogen. These include C=O and C=C hydrogenation, decarbonylation and C-O and C-C hydrogenolysis (Xu et al., 2020). Figure 2-13 and 2-14 shows a selection of simplified oxidative and reductive catalytic transformations of FUR and HMF, respectively.

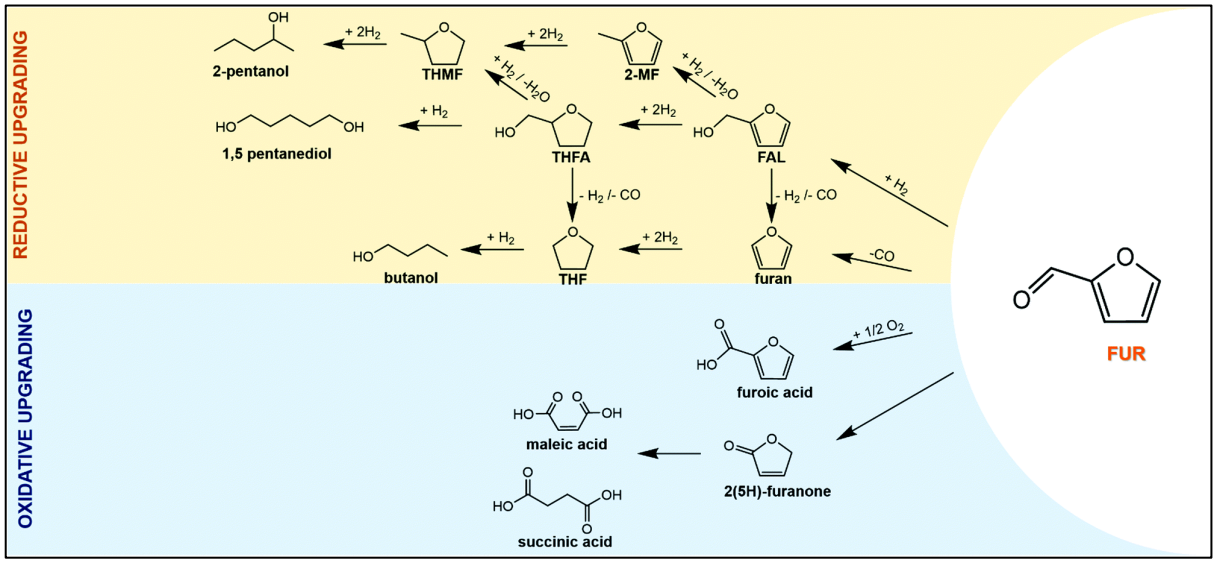


Figure 2-13. A selection of simplified reductive and oxidative catalytic transformations of FUR. Taken from Xu et al. (2020).

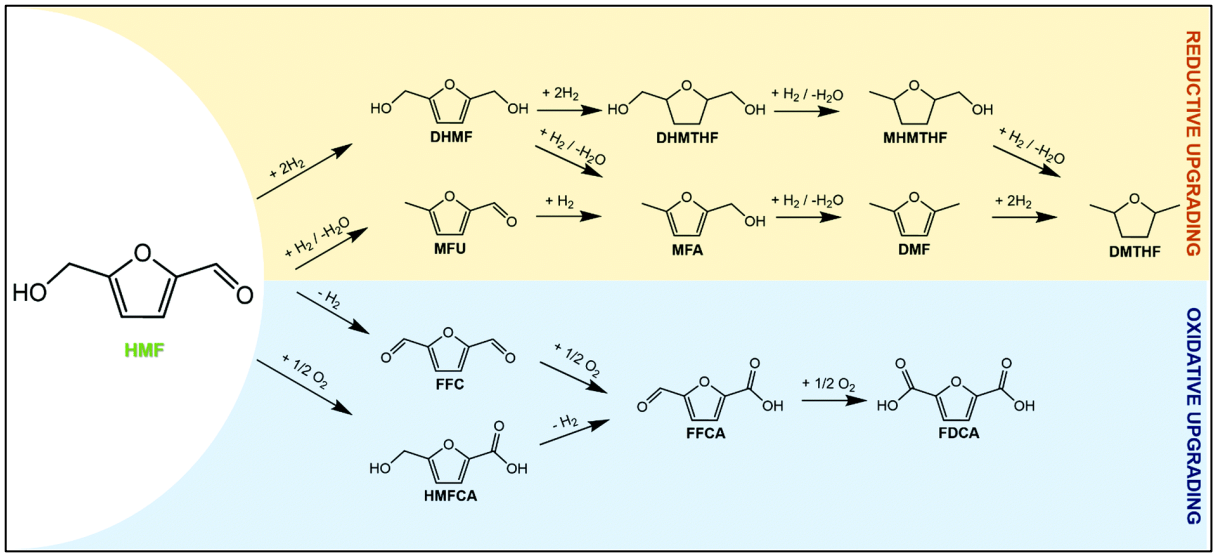


Figure 2-14. A selection of simplified reductive and oxidative catalytic transformations of HMF. Taken from Xu et al. (2020).

As HMF consists of both an aldehydic and a hydroxymethyl group it can be converted into a plethora of high value products. Through catalytic oxidation of HMF, 2,5-furandicarboxylic acid (FDCA) can be produced. FDCA is one of the two key building blocks for the production of polyethylene furanoate (PEF), which is a 100% bio-based polymer that can replace polyethylene terephthalate (PET). PET is a synthetic polyester material that is widely used in plastic products of daily life and has caused serious pollution to the global environment (Liu et al., 2018). PEF has a higher tensile strength compared to PET, and consequently shows excellent barrier properties towards water, oxygen, and carbon dioxide. It is projected that PEF will reach a market value of EUR 770 million in 2025 with a compound annual growth rate (CAGR) of 11% (Xu et al., 2020). At present, the high production costs of HMF limits its competitiveness at an industrial scale (Mika et al., 2018).

Almost 100 chemicals can be directly or indirectly produced from FUR. When furfural is hydrogenated, tetrahydrofuran (THF) can be formed (figure 2-13). THF is a precursor for a wide range of chemical synthesis. It is an important commercial solvent used as the starting material in the production of nylon and polytetramethylene ether glycols (spandex) (Eseyin and Steele, 2015). Furfuryl alcohol (FAL) and 2-methylfuran (2-MF) are formed during reductive upgrading of FUR (figure 2-13). These two chemicals are employed in making antimicrobial agents (Eseyin and Steele, 2015). There is a wide range of possible applications

of FUR and its derivatives, and they have a high demand for application in the plastic, food, pharmaceutical and agricultural industries (Eseyin and Steele, 2015).

The market size of FUR is projected to grow from EUR 500 million in 2019 to EUR 630 million by 2024 at a compound annual growth rate (CAGR) of 4.9%. The market size of HMF is also projected to grow, albeit slower, from EUR 50 million in 2019 to EUR 55 million in 2024 with a CAGR of 1.4% (Xu et al., 2020).

3 Objectives

The primary objective of this thesis is to optimize a thermochemical conversion reaction of carbohydrates in waste biomass (cherry rejects) into renewable platform chemicals like HMF and FUR using a biphasic reaction system. The cherries used were waste biomass because they were considered overripe (too close to expiration) and therefore not good enough for the modern consumer. This makes them a second generation biofeedstock. The biphasic reaction system used (developed by Molnes, 2021) was developed according to the principles of green chemistry. For instance, the solvents and catalyst chosen for the system, sulfamic acid and methyl-isobutyl-ketone (MIBK), are considered green. The optimization of the conversion reaction will be done to evaluate whether cherry rejects can be used as a feedstock in biorefineries for the large-scale production of HMF and FUR. In order to accomplish this, the thesis has been divided into four main tasks:

1. In the first task a HPLC-MS/UV method for separation and quantification of FUR, HMF, and three other selected compounds (5-methylfurfural, 2-acetylfuran and levulinic acid) is developed. The three compounds have been selected based on the available literature of possible valuable compounds HMF and FUR degrade to. These compounds cannot be quantified with qNMR due to a big area of overlapping signals in the NMR spectra making quantification difficult. The HPLC-conditions for separation were found by use of a standard solution mixture containing all five compounds. The quantification was achieved by preparing calibration curves for each compound.

2. In the second task a fractional factorial design (FFD) is performed as a screening for the biphasic reaction system used with cherry rejects as substrate. Four main factors will be included in the screening FFD, namely reaction temperature, residence time, amount of aqueous solvent, and amount of organic solvent. Multivariate analysis by principal component analysis (PCA) and partial least squares regression analysis (PLS) was used to evaluate the effects of the main factors on FUR and HMF yield. qNMR was used to quantify FUR and HMF in the product solutions.

3. In the third task the results from the multivariate analysis of the screening FFD are used to uncover how the main factors can be adjusted to optimize the reaction conditions with regard to FUR and HMF yield. With these results in mind a second optimizing (full) factorial design (FD) is performed using the same biphasic reaction system and substrate. Multivariate analysis (PCA and PLS) was used to evaluate the effects of the main factors on FUR and HMF yield. qNMR was used to quantify FUR and HMF in the product solutions.

4. The fourth task is to investigate the presence of the five selected compounds in the aqueous product solutions of the optimizing FD, using the developed HPLC-MS/UV method. The developed calibration curves were used to quantify the compounds present.

4. Analytical Methods

4.1 Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) is a very versatile spectroscopic technique that enables both structural elucidation and accurate quantification of organic molecules. Any atomic nucleus with either odd mass, odd atomic number, or both has a quantized spin angular momentum and a magnetic moment. Each atomic nucleus with the spin property has a different ratio of magnetic moment to angular moment since each has different charge and mass. This ratio is called the magnetogyric ratio, it is denoted by γ and it is a constant for each nucleus. NMR is dependent on this characteristic (Pavia et al., 2014).

The most common nuclei that possess spin include ${}^{1}_{1}H, {}^{2}_{1}H, {}^{13}_{6}C, {}^{14}_{7}N, {}^{17}_{8}O$ and ${}^{19}_{9}F$. The nuclei of the most abundant isotopes of carbon and oxygen (${}^{12}_{6}C$ and ${}^{16}_{8}O$) are not included among those with the spin property and can therefore not be analysed by NMR. This thesis will only focus on ^{1}H-NMR.

4.1.1 ^{1}H-NMR

Each nucleus with spin has a quantized number of allowed spin states it may adopt, this is determined by its nuclear spin quantum number I. The number I is a physical constant and there are 2I+1 allowed spin states. A proton (hydrogen nucleus) has the spin quantum number I = 1/2, and therefore has two allowed spin states: -1/2 and +1/2. Any moving charge will generate a magnet field of its own. Therefore, when an external magnetic field (B_0) is applied, a magnetic moment, μ, will be generated by the charge and spin of the nucleus. A hydrogen nucleus has a clockwise (+1/2) and a counterclockwise (-1/2) spin, the nuclear magnetic moments generated by each of these spins will point in opposite directions. All the protons will have their magnetic moments either aligned with the field or opposed to it (Pavia et al., 2014). This is illustrated in figure 4-1.

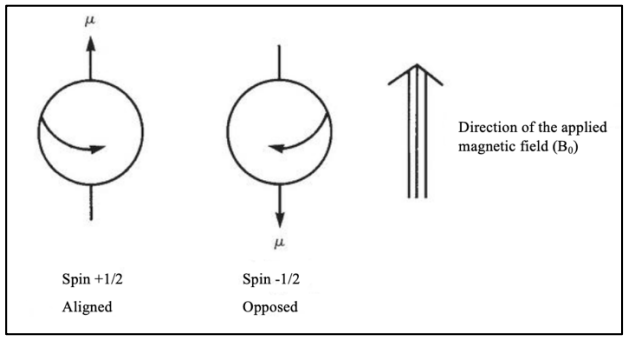


Figure 4-1. The two allowed spin states for a proton and the resulting magnetic moments when an external magnet field is applied. Adapted from Pavia et al. (2014).

The spin state +1/2 is of lower energy because it is aligned with the field. Hence, the spin states split into two states of unequal energy when an external magnetic field is applied. A nuclear magnetic resonance phenomenon occurs when nuclei aligned with the field absorb energy and change their spin orientation with respect to the applied field. The energy absorbed must equal the energy difference between the two states. The energy difference between the two states is a function of the strength of the applied magnetic field B_0. The stronger the field, the greater the energy difference between the possible spin states. The energy difference between the states is also dependent on the magnetogyric ratio γ, which is constant for each atomic nucleus (Pavia et al., 2014). This means that the energy needed to induce a resonance phenomenon is characteristic for each atomic nucleus.

In the presence of an external magnetic field, the nucleus will precess about its own axis of spin with an angular frequency ω, called the Larmor frequency. If radiofrequency waves of the same frequency are supplied to the precessing proton, and only if it's the same, the energy can be absorbed. The nucleus' spin orientation will then change, this is called resonance. To obtain a signal from this resonance phenomenon, the populations of the two different spin states need to be unequal. If the upper and lower states become equal, no net signal is observed. This is called saturation (Pavia et al., 2014).

Not all protons in a molecule will have resonance at the exact same frequency. This is because protons on different sites on a molecule will experience slightly different electronic (magnetic) environments. An applied magnetic field will cause the valence electrons of the protons to circulate, this is called a local diamagnetic current. This current will generate a counter magnetic field opposing the applied magnet field. As a result, each proton is shielded from the

applied field to an extent that depends on the electron density surrounding it. The difference in resonance frequency between the protons are very small, so small that it is difficult to measure the exact frequencies to the precision needed. For this reason, a reference compound is placed in the sample so that the resonance frequency of each proton is measured relative to the resonance frequency of the protons of the reference compound. The universal standard is tetramethylsilane (TMS) (Pavia et al., 2014).

The shift from the reference compound depends on the strength of the applied magnetic field, so comparing different spectra obtained with different field strengths can be confusing. A parameter independent of field strength was therefore defined, chemical shift δ. The chemical shift in δ units express the amount a proton resonance is shifted from the resonance of the reference compound, in ppm of the spectrometer's basic operating frequency (Pavia et al., 2014). This means that the chemical shift value of a given proton is the same regardless of the field strength when the resonance signal is measured.

To use NMR as an effective analytical tool and interpret the resulting spectra, there are a few additional phenomena that need to be understood. These are spin-spin splitting, coupling constants, and relative peak intensities. The spin-spin splitting phenomena arises because hydrogens on adjacent carbon atoms can "sense" one another. The hydrogen on carbon A can sense the spin direction of the hydrogen on carbon B (figure 4-2). In a sample, there will be a mixture of molecules where hydrogen B will have +1/2 spin in some molecules and -1/2 spin in others. Proton A is said to be coupled to proton B. In scenario X (figure 4-2) the spin of proton B is aligned with the applied field, this means that it will slightly deshield proton A. In scenario Y proton A is more shielded from the coupling to proton B.

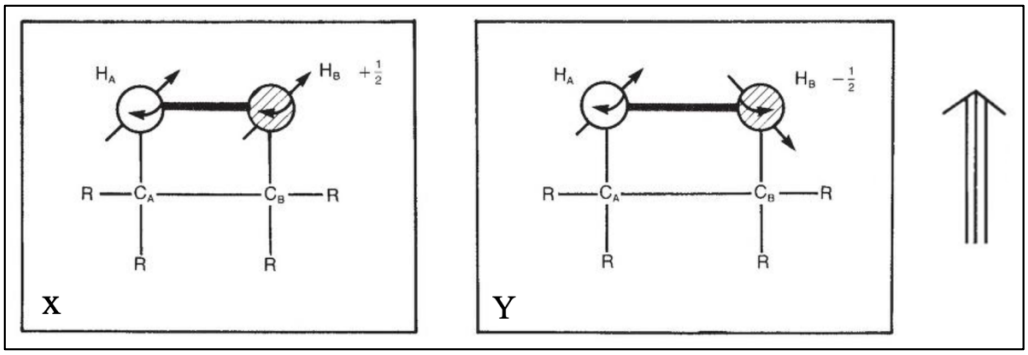


Figure 4-2. The origin of a doublet signal for proton A, owing to a mixture of molecules in scenario X and Y. The right-side arrow denotes the direction of the applied magnetic field.

Adapted from Pavia et al. (2014).

In a sample solution, there will be approximately equal numbers of X and Y molecules at any given time. The result will therefore be two absorptions of nearly equal intensity for proton A (figure 4-3). The signal of proton A is split into what is called a multiplet, more specifically a doublet. The splitting pattern of a signal follows the n+1 rule, where n is the number of protons sensed on neighbouring carbon atoms. The distance between the peaks in a multiplet is called the coupling constant J (figure 4-3). The coupling constant is a measure of how strongly the proton is affected by its neighbours' spin states. It is independent of magnetic field strength and measured in hertz. In this scenario we have a $^{3}J_{HH}$ coupling because the protons are coupled via a three-bond vicinal coupling.

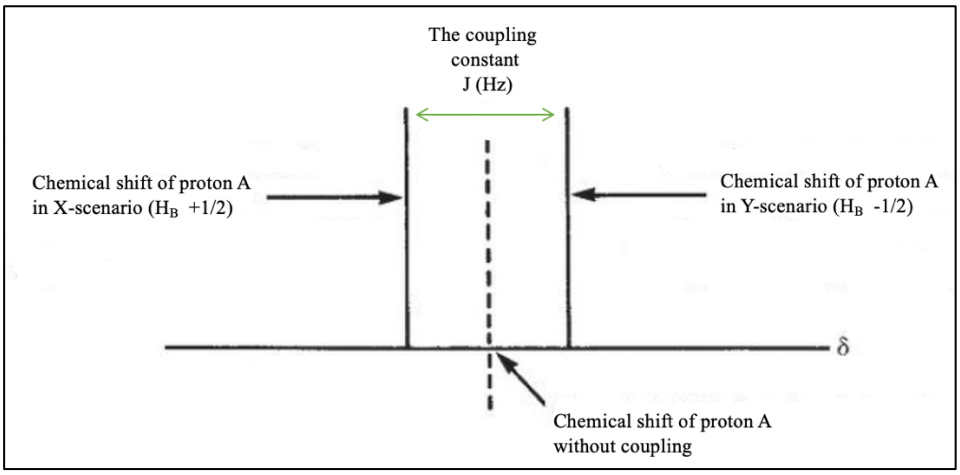


Figure 4-3. The resulting doublet signal for proton A when coupled to proton B, including the coupling constant denoting the strength of the coupling.

Adapted from Pavia et al. (2014).

The area under each peak in the NMR spectra is proportional to the number of hydrogens generating the peak. Hence, the number of peaks tells us how many sets of protons are in chemically equivalent environments (Pavia et al., 2014). This characteristic makes NMR an effective tool for accurate quantification of compounds.

Modern pulsed Fourier transformation (FT) NMR spectrometers use powerful superconducting magnets. They are cooled by liquid helium and nitrogen (figure 4-4). When a sample is prepared for analysis, it is transferred to a NMR tube. The NMR tube is lowered into the magnet and then spun around its own axis to ensure a uniform magnetic field throughout the sample. The probe contains a coil attached to a synthesizer that generates the radiofrequency (RF) pulse needed for the protons to reach resonance. During a RF pulse, the pulse gate closes and allows the radiofrequency to reach an amplifier (figure 4-4). The duplexer directs the high-energy pulse to the NMR-probe, which will excite all the protons in the sample simultaneously.

The pulse ends by opening the pulse gate, the excited nuclei will then return to their original spin states and emit electromagnetic radiation while doing so. Since one sample contains several protons in different chemical environments, the electromagnet radiation emitted during relaxation will contain several different overlapping frequencies. This is called a free-induction decay (FID) signal. The FID signal will be sent to the RF receiver, which converts the signal to audio frequency signals. The audio frequency signals will then be forwarded to an analog-to-digital converter (ADC) that digitalizes the signal and sends it to a computer. Fourier transformation (FT) is applied to the FID signal, which mathematically converts the waves from time domain to frequency domain, and the signal is displayed as separate peaks in the NMR spectrum (figure 4-4).

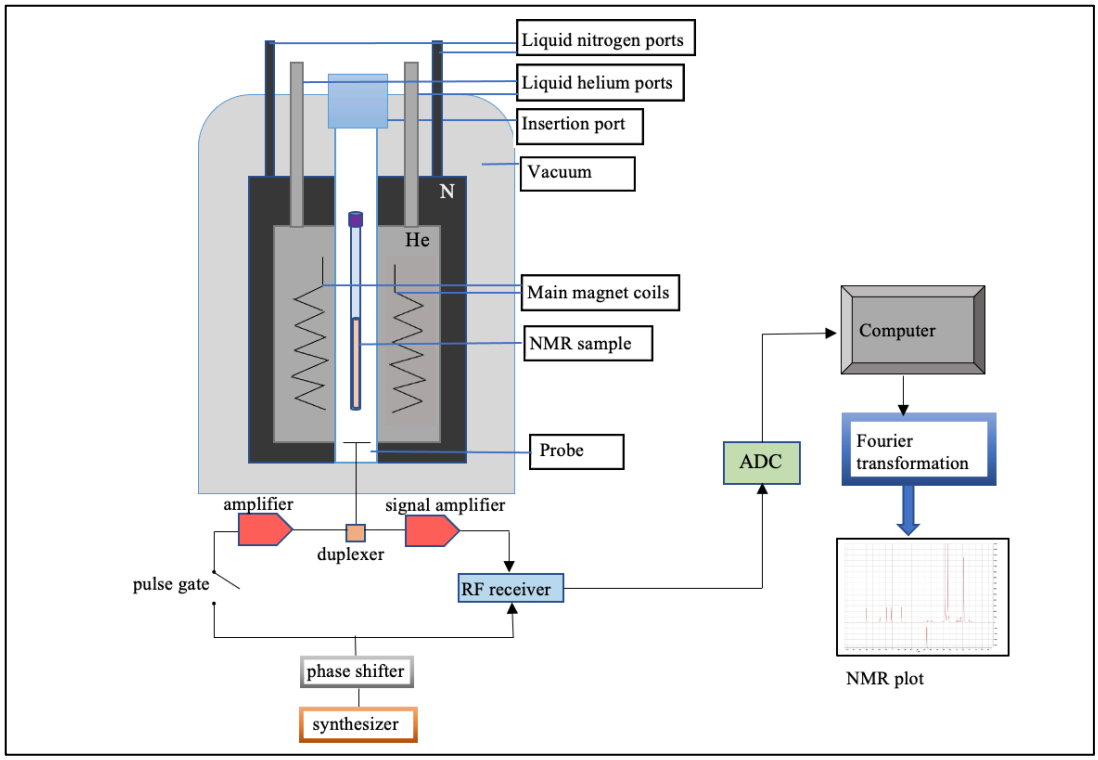


Figure 4-4. An overview of the NMR spectrometer and its parts.

Figure adapted and redrawn from Levitt (2008) and Molnes (2021).

An advantage with pulsed FT-NMR is that several pulses can be sent to the sample consecutively, and the signal-to-noise ratio improves as a function of the square root of the number of scans. This is especially advantageous when analysing nuclei of limited natural occurrence, or very dilute samples (Pavia et al., 2014). However, in order to use NMR quantitatively (qNMR) it is important that the relaxation delay is sufficient. The relaxation delay is the delay time between each pulse. The loss of signal intensity parameter (T1) reports the time each nucleus in a sample needs to lose resonance intensity following pulse excitation. The value of T1 for the nuclei can be estimated with inversion recovery experiments. To ensure that the relaxation delay is sufficient, it needs to be five times the T1 time for the slowest relaxing nuclei in the sample. With correct relaxation delay, quantitatively precise spectra can be recorded and analysed.

4.2 High-Performance Liquid Chromatography (HPLC)

High-performance liquid chromatography is an important chromatographic method used for separations of organic compounds. Figure 4-5 depicts a typical HPLC system.

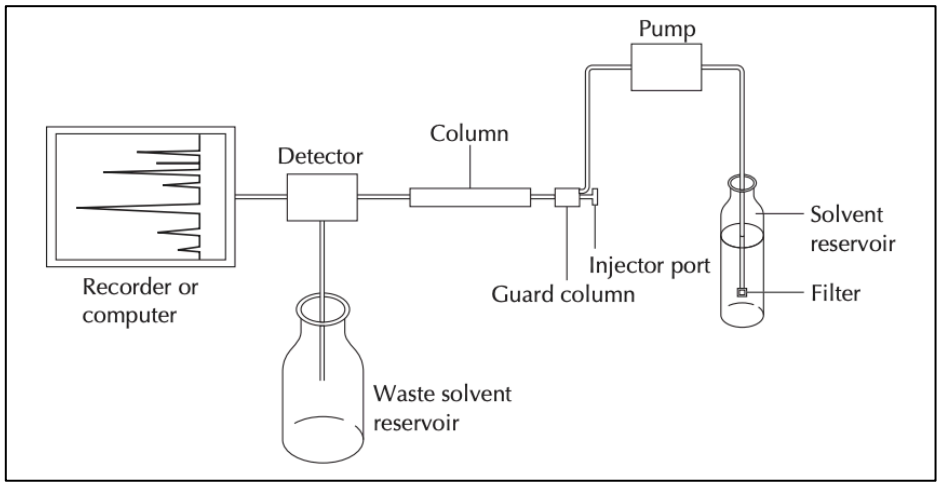


Figure 4-5. The layout of a typical high-performance liquid chromatograph. Taken from Mohrig et al. (2014).

In HPLC, the column is packed with particles that act as the stationary phase. The mobile phase is a liquid that is forced through the column by high pressure (typically 50-200 atm). HPLC utilizes small samples that are injected by an autosampler. There is generally a guard column placed before the more expensive main column. This guard column retains fine particles and strongly adsorbed compounds that would otherwise degrade the main column, it must therefore be replaced periodically. The length of the main column ranges from 5 to 30 cm, with an inner diameter of 1-5 mm. The columns usually have a liquid stationary phase that is covalently bonded to microporous spherical silica (SiO_2) particles. The attached detector can be a number of different instruments (Mohrig et al., 2014).

There are generally two factors that contribute to how well compounds are separated by chromatography. The first is difference in elution times between peaks: The bigger the difference, the better their separation. The second factor is how broad the peaks are: The wider the peaks, the poorer their separation. A band of solute will broaden as it moves through a chromatography column (figure 4-6). One main cause of band spreading is diffusion. Diffusion

is the movement of a solute from a region of high concentration to a region of low concentration, caused by the random movement of molecules (Harris, 2016).

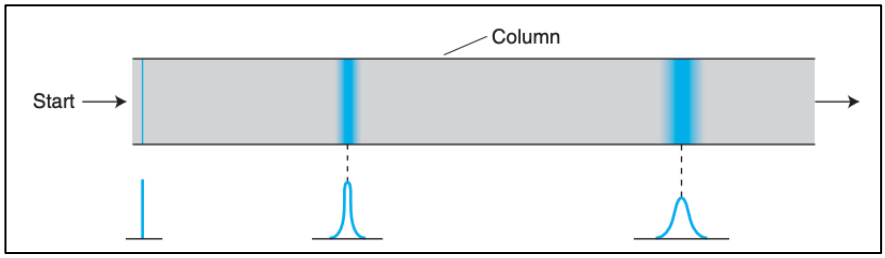


Figure 4-6. Illustration of how an initially sharp band of solute broadens as it moves through a chromatography column. Taken from Harris (2016).

The standard deviation for band spreading is $\sigma = \sqrt{2Dt}$, where D is the diffusion coefficient of the solute and t is time. Plate height, H, is the constant of proportionality between the variance of the band, σ^2, and the distance it has travelled, x ($H = \sigma^2/x$). The plate height is a measure of column efficiency. The ability of a column to separate components of a mixture is improved by decreasing plate height. An efficient column has more theoretical plates, N, than a less efficient column. For a column of length L, the number of theoretical plates is N = L/H. The van Deemter equation tells us how the column and flow rate affect plate height:

$$H \approx A + \frac{B}{u_x} + C_{u_x} \qquad \text{(Equation 4-1)}$$

Where u_x is the linear velocity and A, B, and C are constants for a given column and stationary phase. A is the multiple paths term, B is the longitudinal diffusion term and C is the equilibration time (mass transfer) term. The term A arises from multiple effects for which the theory is unclear. Figure 4-7 shows a pictorial explanation for one of the effects. Molecules entering the column at the same time are eluted at different times because of different flow paths. For simplicity, many different effects are approximated by the constant A (Harris, 2016).

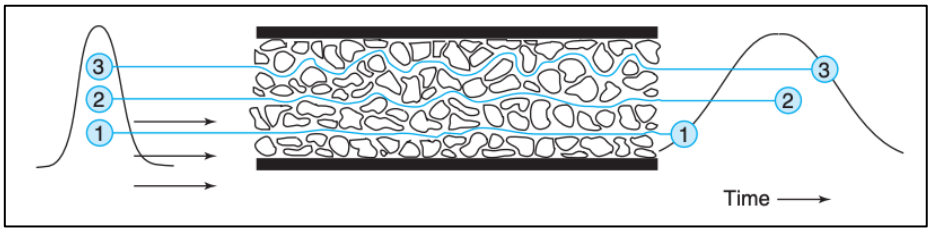


Figure 4-7. Demonstrates the effect of multiple flow paths on band spreading. Taken from Harris (2016).

The term B arises from longitudinal diffusion and is illustrated in figure 4-6. It takes place along the axis of the column and occurs while the band is transported along the column by solvent flow. The faster the linear flow, the less time is spent in the column and less diffusional broadening occurs. The term C arises from the finite time required for solute to equilibrate between the mobile and stationary phases. The solute must diffuse from the mobile phase to the stationary phase's surface for this equilibration to occur. The time required depends on the distance the solute must diffuse to get to the stationary phase and on how fast the solute diffuses. The faster the mobile phase velocity, the less time is available for this transfer to occur (Harris, 2016).

The efficiency of a packed column increases as the particle size of the stationary phase decreases (figure 4-8). Smaller particles give better resolution because they provide more uniform flow through the column, which reduces the multiple path term A. The distance solute must diffuse to equilibrate between the two phases is also smaller, thereby reducing the C term for finite equilibration time. The consequence of smaller particle size is that a higher pressure is needed in order to push the mobile phase through the column. Heating the column will decrease the viscosity of the solvent which will reduce the required pressure for faster flow. Increased temperature decreases retention times and affects relative retention of compounds (Harris, 2016). However, a consequence is that increased temperature can degrade the stationary phase and decrease column lifetime.

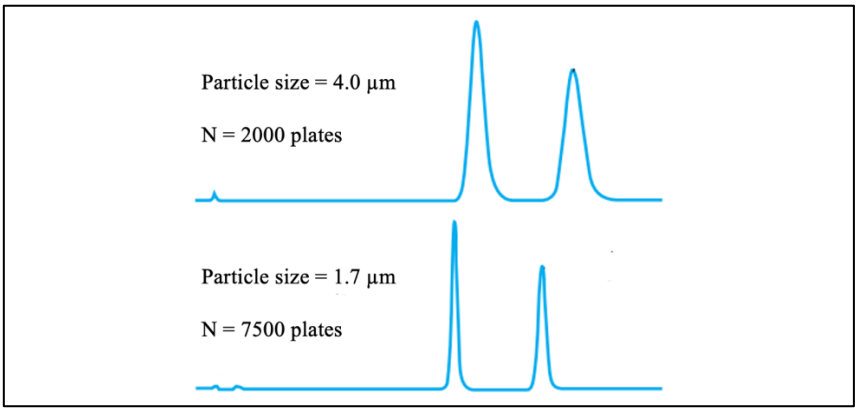


Figure 4-8. Difference in resolution between two chromatograms of the same sample with equal conditions except particle size. Adapted from Harris (2016).

The most common mode of HPLC is reversed-phase chromatography, this is also the mode used in this thesis. The stationary phase of the column is nonpolar or weakly polar, while the mobile phase is more polar. This means that the more polar compounds will elute first from the column and have lower retention times than the less polar compounds. With the type of stationary phase used in this thesis, the solvent molecules compete with solute molecules for retention sites on the stationary phase. This is called adsorption chromatography. Elution occurs when solvent displaces solute from the stationary phase.

Isocratic elution is when one mobile phase of constant composition is used. If one solvent does not elute all the components of the sample fast enough, then gradient elution can be used. Retention is decreased by making the mobile phase more like the stationary phase. In reversed-phase mode this means that a less polar solvent will have a higher eluent strength, it is said to be a stronger mobile phase (Harris, 2016). A gradient elution where the mobile phase is gradually made less polar would therefore make the elution of all components faster.

4.3 Mass Spectrometry (MS)

Mass spectrometry (MS) is a technique used to determine the molecular weights and structures of compounds. In MS, molecules are ionized using a variety of different techniques (ionization methods) to generate either cationic or anionic compounds. Electric fields are then used to manipulate the movement of these ions in a way that enables us to determine their mass-to-charge (m/z) ratios. The resulting mass spectra from these measurements are plots of the number of ions observed at each m/z value (Harris, 2016). The basic components of a mass spectrometer are illustrated in figure 4-9.

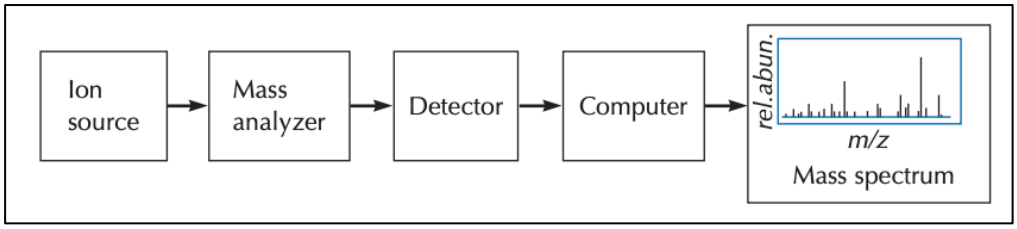


Figure 4-9. The basic components of a mass spectrometer. Taken from Mohrig et al. (2014).

In this thesis, a chromatography-mass spectrometry interface will be used. The ionization method used is electrospray ionization (ESI), which is suitable for a liquid sample arriving straight from the chromatography column. Figure 4-10 displays a mass spectrometer with ESI. Liquid from the chromatography column enters the steel nebulizer capillary at the upper left, along with a flow of N_2 gas. ESI can be used in both positive and negative mode. This is done by changing the voltage in the spray chamber. Positive ESI was used in this thesis and will therefore be the primary focus going forward.

The strong electric field at the nebulizer outlet, combined with the flow of N_2 gas, creates a fine aerosol of charged droplets. Positive ions from the aerosol are attracted toward the glass capillary by a negative potential of about -4500 V, leading the ions into the mass spectrometer. Initially, the droplets will have multiple charges and will shrink by evaporation of the solvent. When the droplets become smaller, the electrostatic repulsion from the charges will be bigger than the surface tension of the solvent. The droplets will then split into smaller droplets. The process continues until single ions are formed (Harris, 2016).

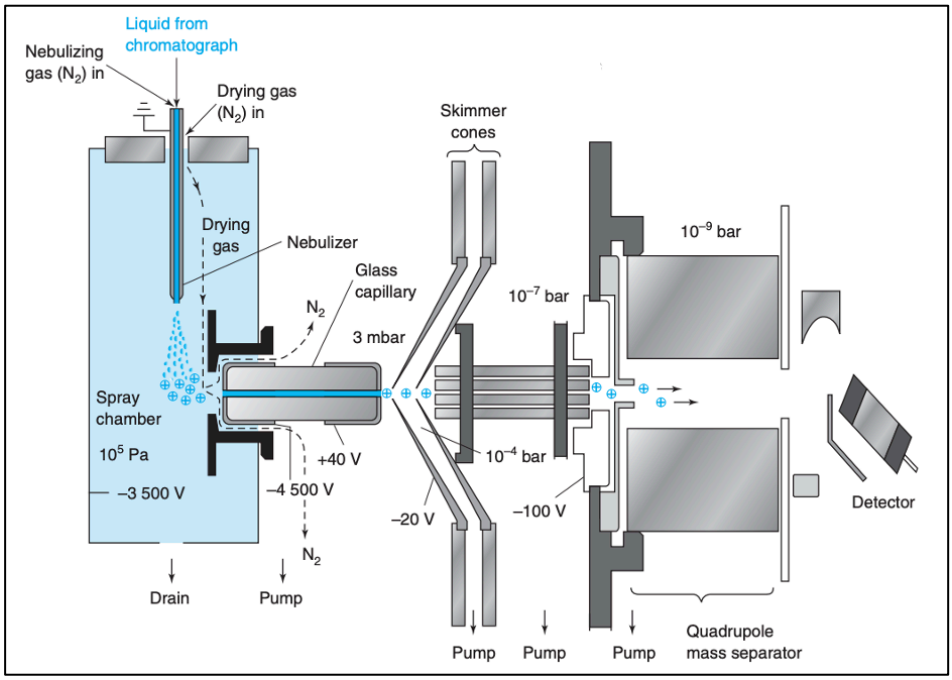


Figure 4-10. A mass spectrometer with electrospray ionization (ESI) as ionization method. Adapted from Harris (2016).

Positive ESI will mainly give $[M+H]^+$ ions, that is the molecule analysed (M) plus a hydrogen atom. The compound will give a signal in the mass spectra at a mass equal M+1. Other ions will also arise from the complexation between the analyte (M) and stable ions from the solution. Common examples are listed in table 4-1.

Table 4-1. List of common ions arising from complexation between the analyte (M) and stable ions in the solution with positive ESI as ionization method.

Ion:	**Mass:**
$[M+H]^+$	M + 1
$[M+Na]^+$	M + 23
$[M+K]^+$	M + 39
$[M+NH_4]^+$	M + 18
$[M+HCO_2]^+$	M + 45
$[2M+H]^+$	2M + 1
$[2M + Na]^+$	2M + 23

There are several components that need to be adjusted and optimized to obtain a satisfactory MS spectrum of a given compound. The temperature needs to be high enough to evaporate the solvent, but not too high as it can cause decomposition of the analytes or unwanted fragmentation. The magnitude of the potential difference (capillary voltage) can to a certain degree control which compounds are ionized, and which adducts are formed. It also affects the degree of ionization. The fragmentor at the end of the glass capillary also affects fragmentation and its voltage needs to be adjusted. The nebulizer pressure needs to be optimized. The flow rate of N_2 gas is also important. At too low flow rates, the evaporation of the solvent will not be sufficient. Too high flow rates can give insufficient transfer of charge to the droplets, and the sensitivity will be reduced (Harris, 2016).

ESI is considered a "soft" ionization method, meaning that it causes little or no fragmentation. The most intense peak in a mass spectrum is called the base peak. The intensities of other peaks in the same spectrum are expressed as a percentage of the base peak intensity. In mass spectra obtained with positive ESI, the $[M+H]^+$ ion is usually the peak of highest intensity and therefore the base peak. When recording a mass spectrum, it results in a total ion chromatogram (TIC) which is a plot of the total ion current (ions reaching the MS detector) vs. retention time. The TIC monitors a large range of mass-to-charge units. The mass chromatogram is extracted from the TIC. Some compounds are hard to ionize with ESI. For a compound to be analysed with ESI, it needs to be able to stabilize the charge given during ionization.

While recording a mass spectrum it's important that the acquisition rate/scan frequency is adequate, especially when performing quantitative analyses. A too high acquisition rate can give loss of signal strength, and too low will give a poor description of the chromatographic peaks and inaccurate integration. Selected ion monitoring (SIM) is a mass spectrometry technique that increases the selectivity of mass spectrometry for a few chosen analytes and decreases the response to everything else. During sample recording with this SIM scan type, the mass spectrometer is set to monitor just a few values of m/z, giving a greater signal-to-noise ratio than with the typical MS scan (Harris, 2016).

4.4 Ultraviolet and Visible Spectroscopy (UV-Vis)

When a compound absorbs ultraviolet (UV) or visible (Vis) light, an electron is excited from a bonding or nonbonding molecular orbital to an antibonding molecular orbital. Absorption of UV radiation can for example promote an electron from a π-bonding orbital into a higher energy π^*-antibonding orbital (figure 4-11). The energy required for an excitation process like this is comparable to the strength of a chemical bond.

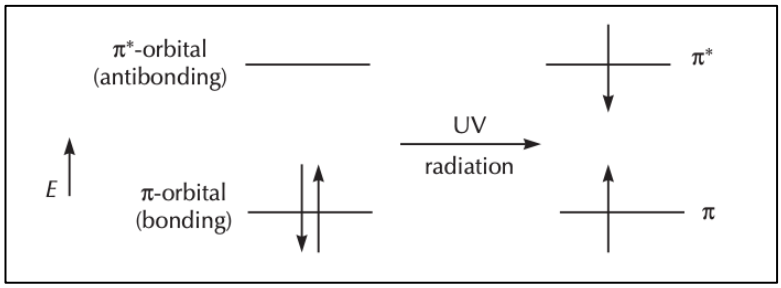


Figure 4-11. Demonstration of an electron's excitation from a bonding to an antibonding π-orbital by cause of UV radiation absorption. Taken from Mohrig et al. (2014).

There are two major classes of UV-Vis spectrometers: dispersive and multiplex diode-array spectrometers. Figure 4-12 shows the instrumental design for a dispersive UV spectrometer with a double-beam light pathway. In this type of UV spectrometer, the light passes through a monochromator, which scans through narrow bands of separate light frequencies. After passing through the monochromator, the radiation is split into two beams which are directed by mirrors through sample and reference cells. The two beams are then recombined later in the optical path. Detectors for dispersive UV-Vis spectrometers are either photocells or photomultipliers. The photocell is the simplest detector. It has a metal surface, and when radiation hits it, electrons are ejected and then converted into a signal (Mohrig et al., 2014).

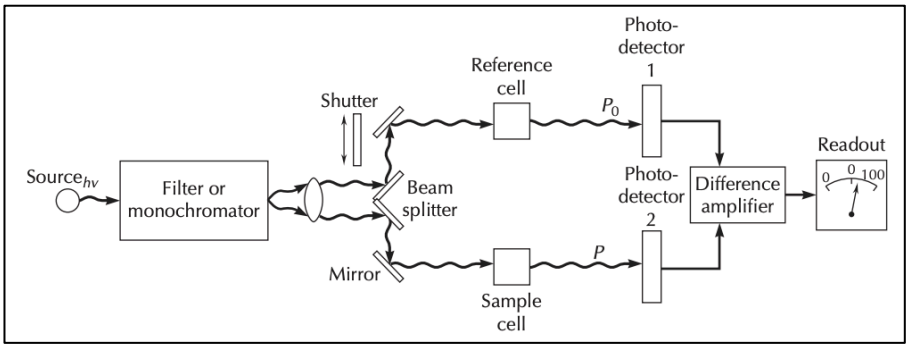


Figure 4-12. An illustration of the instrumental design for a dispersive UV spectrometer with a double-beam light pathway. Taken from Mohrig et al. (2014).

All organic compounds absorb UV light. However, there are only a few commercial spectrometers that can effectively scan the wavelengths where C-H, C-C and nonconjugated C=C bonds absorb, this is due to interference from strong UV-absorption by O_2 and CO_2 in the atmosphere (Mohrig et al., 2014). In UV spectroscopy, the absorption of wavelengths between 200 and 400 nm are examined. In visible spectroscopy, light of wavelength between 400 and 800 nm are examined.

UV-Vis spectra are plots of absorbance (A) against the wavelength of the radiation in nanometres. The wavelength of absorption is measured at the peak, denoted λ_{max}. The concentration of a compound is related to the absorbance by Beer-Lambert Law:

$$A = \log\left(\frac{I^{\circ}}{I}\right) = \varepsilon l c \qquad \text{(Equation 4-2)}$$

Where I° is the intensity of the incident light, I is the intensity of the transmitted light, ε is the molar extinction coefficient in M^{-1} cm^{-1}, l is the length of the cell path in centimetres and c is the sample concentration in M.

Functional groups that absorb UV-Vis light are called chromophores. The proportionality of the absorbance and the concentration is linear over a wide range of concentrations. This makes UV-Vis spectroscopy ideal for determining the concentration of a compound. Quantification is usually done by preparing calibration curves.

4.5 Experimental Design

"Experimental design and optimization are tools that are used to systematically examine different types of problems that arise within, e.g., research, development and production" (Lundstedt et al., 1998, p. 4). When performing an experiment, there are several experimental variables (factors) that may affect the results (responses). There may also be interactions between factors (interaction effects) that affect the final response. To find all the factors and interaction effects that affect the responses, it is essential to plan the experiments so that all the important information is obtained.

In a factorial design the impact of all factors and interaction effects on the response(s) is investigated. If k factors are investigated at two levels, then the factorial design will consist of 2^k experiments. When planning the experimental setup, a table where all the columns are orthogonal is used. It is common to choose two values of each factor investigated, they are usually denoted with a + (plus) for the higher level and a – (minus) for the lower level. When choosing the levels of each factor, the experimental domain is chosen simultaneously. The experimental domain is the experimental area that is investigated, defined by the variation of the factors. At least two centre points are also included, where all the factors are set at their mid value. Centre points are added to minimize the risk of missing non-linear relationships in the middle of the intervals and the repetition also allows for determination of reproducibility and confidence intervals (Lundstedt et al., 1998).

As mentioned earlier, to investigate the effect of k factors in a factorial design, 2^k experiments need to be conducted. If there are seven factors to investigate, there will be $2^7 = 128$ experiments to conduct. It doesn't take many variables before there are more experiments than possible to perform. Fractional factorial designs (FFD) are the solution to this problem. In most cases, the interactions of third order or higher are negligible and can be excluded from the polynomial model. A FFD with 2^{k-p} experiments can then be performed instead, where p denotes the size of the fraction (Gunst and Mason, 2009). Hence, in FFDs many variables can be investigated without an excessive number of experiments.

Use of FFDs is very advantageous when it comes to optimizing a response, e.g., yield of a product. An initial "screening" is first done where the factors and experimental domain is

chosen. The experimental domain should cover a wide enough area so that a maximum of information can be obtained by performing a minimum of experiments. When the influences of the different factors on the response is uncovered from the first design, a second FFD can be planned. The results from the second FFD will decidedly be more satisfactory, due to more knowledge about how the different factors affect the response.

4.5.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a multivariate statistical technique that analyses a data table containing observations described by several quantitative dependent variables. The goal of PCA is to extract the important information from the data table, compress the size of the data set by only keeping the important information, simplify the description of the dataset and analyse the structure of the observations and the variables (Abdi and Williams, 2010). To realize these goals, new variables called principal components are computed. These are obtained as linear combinations of the original variables.

When computing the first principal component (PC1), it is required to contain the largest possible variance. This component will explain or extract the largest part of the variance of the data table. If the dataset contains more variation that is not described by PC1, a second principal component (PC2) can be computed. PC2 needs to explain the largest possible variance of what remains, and it is required to be orthogonal to PC1. Figure 4-13 illustrates an example where the variance of a set of datapoints are explained by two PCs.

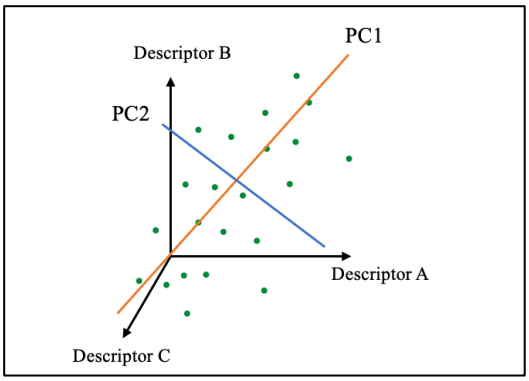


Figure 4-13. Illustrates an example of how two principal components can be computed to explain the variance of a set of datapoints.

When the principal components are computed, perpendicular projections of all the data points, which represent experiments, can be made onto the PC variables. These values are called scores, t. The correlation between a component and a variable is called a loading (Abdi and Williams, 2010).

The principal components can be used to generate two-dimensional plots known as biplots. A biplot combines the PC scores for all the experiments (objects), the PC loadings for all the variables, and measured responses. In a typical biplot there is a cross somewhere in the plot that denotes the origin. A variable's distance from the origin is proportional to the variable's weight on the PC. This means that the most influential variables will be located the furthest away from the origin. A biplot presents a lot of information about the dataset in a convenient way. The more similar different objects, variables or responses are, the closer they will be grouped together in the biplot. In addition, it displays how different objects correlate to each other, different variables, and to the measured responses. An angle of exactly 90° between two points means that no correlation is observed. An angle of less than 90° indicates a positive correlation, and an angle of more than 90° indicates a negative correlation.

4.5.2 Partial Least Squares (PLS) Regression

Partial least squares (PLS) regression, also known as projection on latent structures, is a technique that combines features from PCA and multiple linear regression. The goal of PLS is to analyse and predict a set of dependent variables from a set of independent variables. This is achieved by extracting a set of orthogonal factors from the variables, called latent variables (Abdi, 2010). This technique is especially useful when we need to predict a set of dependent variables from a large set of independent variables.

Let's say we have one matrix, X, containing independent variables (or factors) for a reaction system, and another matrix, Y, containing the dependent variables (responses) for the same reaction system. The goal of PLS is to find components from X that best predict Y. It searches for a set of components (latent vectors) that performs a simultaneous decomposition of X and Y with the constraint that these latent vectors explain as much as possible of the covariance between X and Y. It is then followed by a regression step where the latent vectors obtained from X are used to predict Y (Abdi, 2010). Hence, the results from a PLS analysis can assist in finding the optimal factor settings for X that give the best responses in Y.

An advantage of PLS is that the results can be graphically presented in various plots. This gives the opportunity to extract the correlation coefficient, R, for the model. The more accurate the model describes the system, the closer R is to 1. PLS can also be used to determine weighted correlation coefficients which tells us the relative importance of each factor for a given response. The resulting linear regression model (including second order interactions) can be expressed as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + b_{ij} x_i x_j \qquad \text{(Equation 4-3)}$$

Where y is the dependent variable (the response, e.g., yield), b_0 is constant, b_1, b_2…, b_k is the weighted regression coefficients, b_{ij} is the cross-product coefficients, and x is the value of each variable.

5. Materials, Experimental and Equations

5.1 Materials

All reagents, solvents, and catalysts used were purchased from Sigma-Aldrich and are listed in table 5-1. The waste biomass (cherries) was obtained from Hardanger Fjordfrukt SA.

Table 5-1. List of solvents, reagents, and catalysts used in this thesis organized in alphabetical order.

Chemical	Abbreviation	Chemical formula	Purity (%)
2-Acetylfuran	2-AF	$C_6H_6O_2$	>99.0
Deuterium oxide	D_2O	D_2O	99.9% D w/0.05 wt% TSP
Dimethyl sulfone	$DMSO_2$	$C_2H_6O_2S$	98
Furfural	FUR	$C_5H_4O_2$	>99.0
5-Hydroxymethylfurfural	HMF	$C_6H_6O_3$	>99.0
Levulinic acid	LA	$C_5H_8O_3$	>98.0
5-Methylfurfural	5-MF	$C_6H_6O_2$	>98.5
Methyl-isobutyl-ketone	MIBK	$C_6H_{12}O$	≥99.5
Sodium phosphate dibasic dihydrate		$Na_2HPO_4{\cdot}2H_2O$	≥99.0
Sulfamic acid	SA	H_3NSO_3	≥99.0

5.2 Experimental

5.2.1 Experimental Design

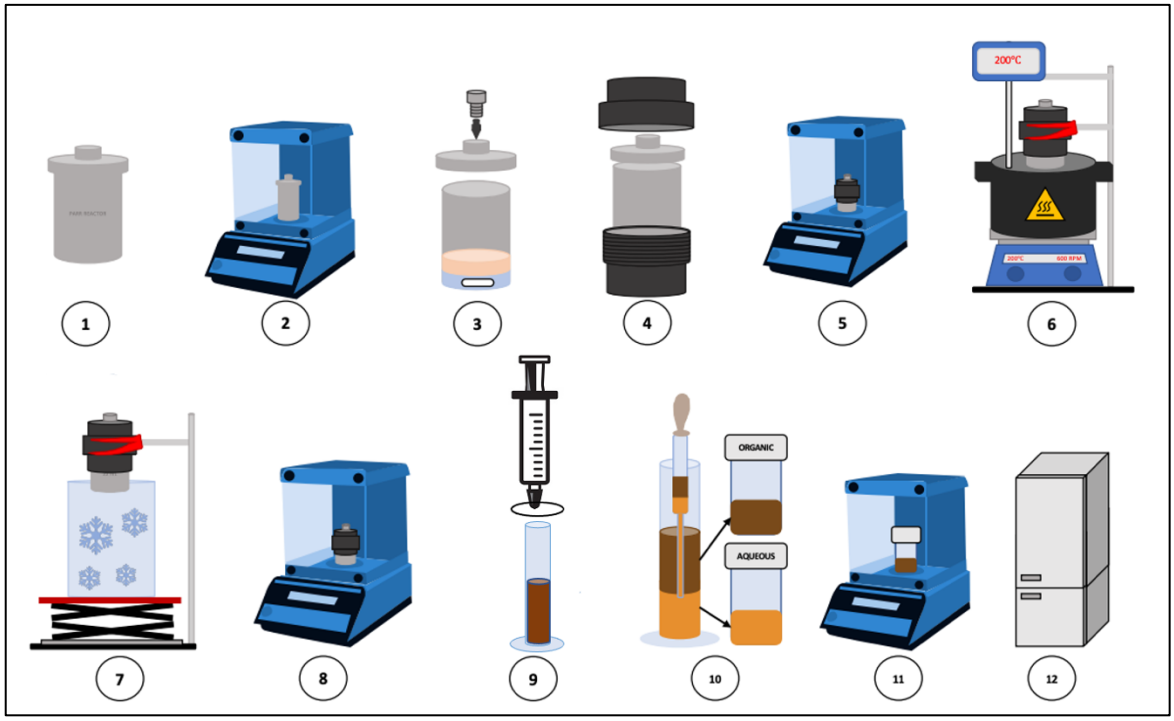
The results obtained from the experimental designs in this thesis (fractional factorial design and full factorial design) were analysed with Sirius PRS 10.0. The multivariate analysis techniques principal component analysis (PCA) and partial least squares regression (PLS) were used to determine correlations between the main factors, factor interactions, and measured response. The results of the multivariate analyses are presented and discussed in section 6.2.

5.2.2 Experimental Procedures – Thermochemical conversion reaction

The conversion reactions were carried out in a Series 4700, 316 Stainless steel batch Parr reactor of size 22mL (picture in figure E-2, Appendix E). Before reaction start, the cherries were deseeded and ground into a semi-homogenous purée with a mortar and pestle. Scheme 5-1 illustrates the procedural steps for the biphasic reaction system (developed by Molnes, 2021). The empty reactor (**1**) was first weighed (**2**) and then the substrate, catalyst, solvents (water and MIBK) and a magnetic stirrer was added to the reactor (**3**). The reactor was subsequently closed (**4**) and weighed again after closing (**5**). Next, it was lowered into an oil bath (**6**) with a temperature ranging from 170-230°C for 30-120 minutes. As soon as the reaction time ended, the reactor was lowered into an ice bath to cool down (**7**). An ice bath of the same size was used each time to ensure a similar cooling rate for each experiment.

When cooled down, the reactor was dried and weighed again (**8**). This was done to check for potential leakage from the reactor. The product solution was then transferred into a pre-weighed 20mL syringe equipped with a 0.45μm filter on the end (**9**). The solution was pressed through the filter and into a pre-weighed 10mL graduated cylinder (**9**). The reactor, syringe

and filter were weighed again after emptying to determine the loss of product. After a few minutes wait, the phases in the graduated cylinder separate. Phase-separation was conducted with a VWR 230mm disposable glass pipette (**10**). The different phases were separated and transferred into pre-weighed 20mL sample vials (**10**). The graduated cylinder was weighed again after emptying to determine the loss of product. The sample vials filled with the different phases were then weighed to determine mass recovery (**11**). Finally, both the phases of the product solution were refrigerated (**12**) until preparation of NMR samples. The NMR samples were always prepared the same day as the conversion reaction was conducted. This was done to ensure that no degradation of compounds took place before analysis.



Scheme 5-1. The procedural steps for the conversion reaction of cherries to FUR and HMF. Adapted and parts redrawn from Molnes (2021).

Because different biomass samples contain varying amounts of moisture, the results (yields) are reported from dry weight. The total solids remaining after drying of the biomass sample was determined by a laboratory analytical procedure developed by the National Renewable Energy Laboratory (Sluiter et al., 2008).

5.2.3 Experimental Procedures – NMR

5.2.3.1 Preparation of 20% D_2O stock solution

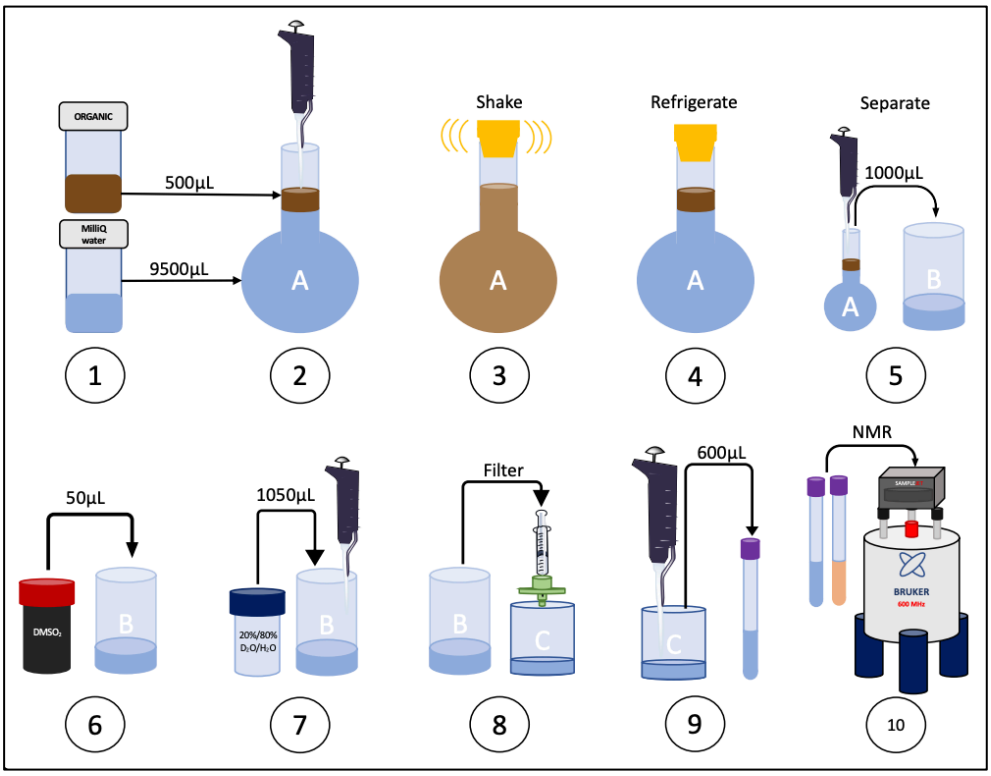
A 20% D_2O stock solution is needed for NMR sample preparation (volume percent). A volume of 250mL solution was prepared. The first step is to weigh and add 0.445g sodium phosphate dibasic dihydrate ($Na_2HPO_4 \cdot 2H_2O$) to the container (concentration of 0.010 M in 250 mL). Then, 120mL distilled water and a magnetic stirrer was added. The solution was stirred and heated until everything was dissolved. Next, 20mL D_2O (containing TSP-sodium salt) was transferred to the container. The solution was pH-adjusted to pH = 7.4 using 1.0 M HCl and 1.0 M NaOH. The pH-adjustment was performed using a Metrohm 798 MPT Titrino automatic titrator. Finally, distilled water was added until the volume reached 250mL.

5.2.3.2 Preparation of NMR samples

To use the NMR sample workup protocol (developed by Løhre et al., 2021) the organic phase had to be extracted with water before sample preparation (Scheme 5-2, steps **1**-**4**). To perform the extraction with water, 500µL of the organic phase was transferred to a 10mL volumetric flask. Then, distilled water was added until the volume reached 10mL (**2**). Both the organic phase and water added was weighed so that an accurate dilution factor could be calculated. The mixture was shaken vigorously (**3**) before refrigeration for at least an hour to ensure adequate phase separation (**4**) (figure E-3, Appendix E). From this point the steps are the same for the aqueous phase and the extracted organic phase.

To a sample vial, 1000µL of the aqueous phase/extracted organic phase was transferred (**5**), followed by 50µL of $DMSO_2$ (used as internal standard, target concentration is 0.101 M) (**6**). Then 1050µL 20% D_2O stock solution (section 5.2.3.1) was added to the sample vial (**7**), making the total percentage of D_2O in the sample 10%. The sample was then filtered using a 0.2µm Pall Acrodisc nylon filter (**8**) and 600µL of the sample was transferred to an NMR tube

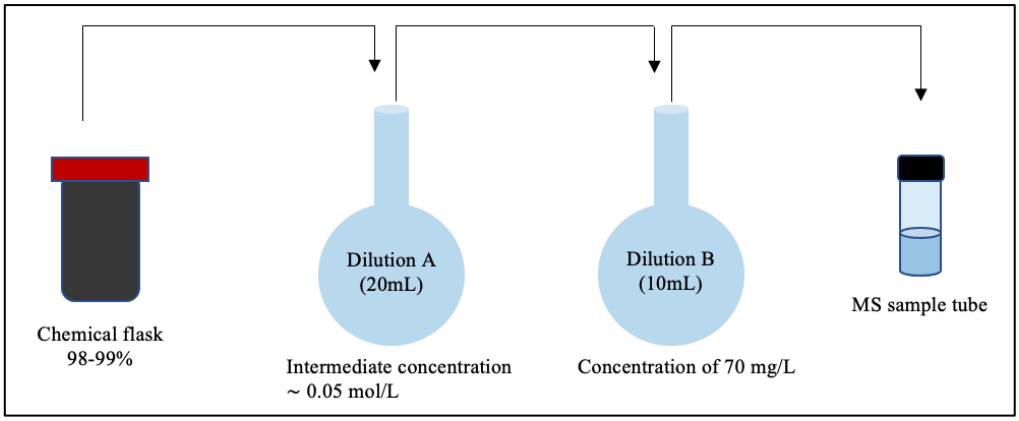
(**9**) ready for NMR analysis (**10**). All the NMR samples in this thesis was analysed with a Bruker BioSpin 600 MHz AVANCE NEO spectrometer equipped with a QCI Cryoprobe.



Scheme 5-2. The procedural steps for NMR sample preparation. Steps 1-4 shows the extraction with water for the organic phase. Steps 5-10 shows sample preparation for aqueous phase and extracted organic phase. Taken from Molnes (2021).

5.2.4 Experimental Procedures – HPLC-MS/UV method development

The first step during the HPLC-MS/UV method development, was to find the correct MS-settings needed to analyse the five different compounds. To do that, a standard solution of suitable concentration for direct MS analysis (70 mg/L) was prepared for each compound. Scheme 5-3 illustrates the dilution steps from highly concentrated chemical flask to suitable concentration for MS analysis. Each dilution step was weighed, and the exact concentration of dilution A and B was calculated for each compound. The compounds were diluted with milliQ water. An intermediate dilution step was done to ensure accurate concentration in dilution B and because the solubility of some of the compounds in water was low.



Scheme 5-3. Illustrates the dilution steps from chemical flask to a concentration of 70 mg/L, suitable for direct MS analysis.

After the correct MS-settings were determined for the five compounds, the next step was to determine the correct settings for separation with HPLC. For this step a mixture of all the five compounds were made, in which the concentration of each compound was 70 mg/L. For this, the already prepared dilution A (scheme 5-3) for each compound was used. The correct volume needed for each compound was calculated before they were transferred from flask A to a new 20mL volumetric flask. The flask was filled to 20mL with milliQ water, and a magnetic stirrer was added. When everything was completely mixed, the solution was transferred to a new MS sample tube and analysed.

When both the correct MS- and HPLC-settings were determined, calibration curves for each compound was made. The solutions for the calibration curves were made in the same way as the mixture for determining HPLC-settings. Five different mixtures of the compounds were made at five different concentrations (5, 60, 120, 180 and 240mg/L). The already prepared dilution A was used to make these further dilutions. Each step was weighed so that the concentrations of each compound at each level could be calculated accurately.

The instrument used was the Agilent 6420A triple quadrupole mass analyser equipped with electrospray ionization (ESI). It is connected to an Agilent 1200 series LC module. The LC column installed is an Agilent ZORBAX SB-C18 RRHT threaded column of size 2.1x50mm. The column is packed with 1.8µm particles which consist of a C18 stationary phase chemically bonded to a high purity ZORBAX porous silica microsphere.

5.3 Equations

Calculation of mass percent HMF yield (m%):

$$\left(\frac{m_{HMF\,(org)}+m_{HMF\,(aq)}}{m_{substrate\,(dry\,weight)}}\right)\times 100 \qquad \text{(Equation 5-1)}$$

Where $m_{HMF\,(org)}$ and $m_{HMF\,(aq)}$ is the mass of HMF in the organic and aqueous phase, respectively, and $m_{substrate\,(dry\,weight)}$ is the dry weight of the substrate added.

Calculation of mass percent FUR yield (m%):

$$\left(\frac{m_{FUR\,(org)}+m_{FUR\,(aq)}}{m_{substrate\,(dry\,weight)}}\right)\times 100 \qquad \text{(Equation 5-2)}$$

Where $m_{FUR\,(org)}$ and $m_{FUR\,(aq)}$ is the mass of FUR in the organic and aqueous phase, respectively, and $m_{substrate\,(dry\,weight)}$ is the dry weight of the substrate added.

6. Results and Discussion

This section covers the results acquired from the objectives in this master's thesis. The results and discussion are divided into two main parts based on the objectives given in section 3. First, the developed HPLC-MS/UV method and the prepared calibration curves will be presented. Then the results from the experimental designs will be given and discussed. The results from the screening fractional factorial design (FFD) quantified with qNMR will be presented, followed by the results from the optimizing factorial design (FD), also quantified with qNMR. Finally, the results of the optimizing FD investigated and quantified with the developed HPLC-MS/UV method will be presented.

6.1 HPLC-MS/UV method development

A literature study on HPLC methods for HMF separation and quantification was performed, leading to the choice of reverse-phase HPLC with positive ESI as ionization method (Salis et al., 2021, Xu et al., 2018). The molecular structure of the five compounds to be analysed are presented in figure 6-1.

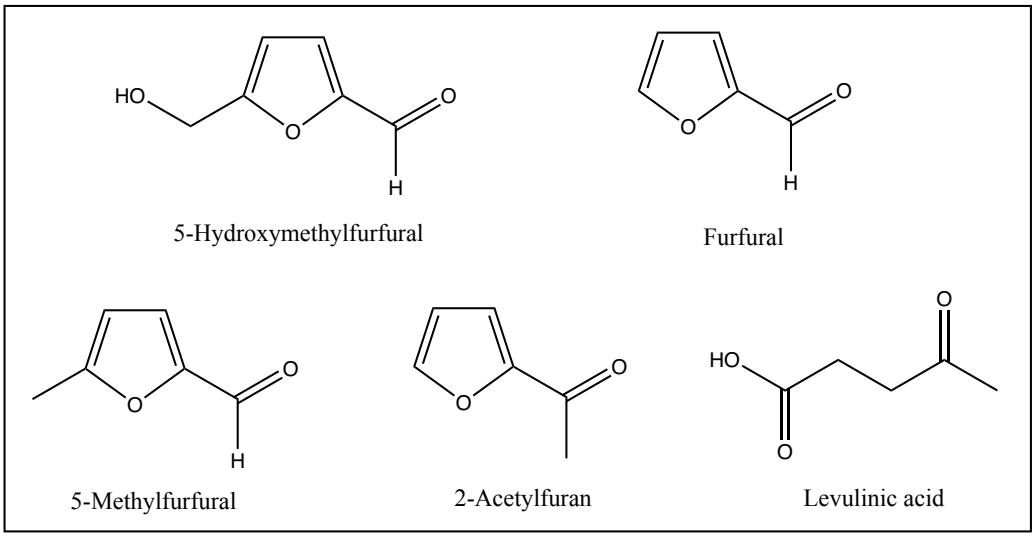


Figure 6-1. The molecular structure of the five compounds separated and analysed with HPLC-MS/UV.

The first step was to find suitable MS-settings for each compound. However, not all the compounds could be analysed with positive ESI due to their structure not being able to stabilize the charge given during ionization. Fortunately, the instrument used was equipped with both a mass detector and a UV-Vis detector. Table 6-1 gives an overview of the compounds and their corresponding MS and/or UV-Vis signals.

Table 6-1. An overview of the compounds, their molar mass, the detection method, and their corresponding MS and/or UV-Vis signals.

Compound:	**Molar mass (g/mol):**	**Detection method:**	**MS signal (m/z):**	**UV-Vis signal, λ_{max} (nm):**
HMF	126.11	MS and UV-Vis	127.1 $[M+H]^+$, 109.0 $[M-OH]^+$	283
FUR	96.08	UV-Vis	-	280
5-Methylfurfural	110.11	MS and UV-Vis	111.1 $[M+H]^+$	290
2-Acetylfuran	110.11	UV-Vis	-	270
Levulinic acid	116.11	MS	117.1 $[M+H]^+$, 139.0 $[M+Na]^+$, 99.0 $[M-OH]^+$	-

A combination of settings that worked for all the compounds detected with MS was found. The settings are displayed in table 6-2. The UV-Vis detector scans in the range 200-850 nm in intervals of 2.0 nm. MS and UV-Vis spectra of all the five compounds, including a background MS spectrum, are given in figure B-1 to B-8 in appendix B.

According to theory, the $[M+H]^+$ ion is generally the signal of highest intensity in spectra obtained with positive ESI. This is the case for HMF and 5-methylfurfural, but for levulinic acid the $[M+Na]^+$ signal is of highest intensity (figure B-8). The MS signals of m/z value 109.0 and 99.0 suggest that both HMF and levulinic acid undergoes fragmentation in the form of losing a hydroxyl group (OH).

Table 6-2. The chosen MS-settings for direct MS analysis.

Parameter:	Specification:
Mobile phase:	MilliQ water
Capillary (positive ESI):	3500 V
Fragmentor:	80 V
Nebulizer:	45 psi
Injection volume:	1 µL
Gas flow:	8 L/min
Mass range:	50-400 m/z

The next step was to determine the HPLC-settings for the compounds. Several of the five compounds to be separated and analysed have quite similar molecular structures (figure 6-1). Consequently, gradient elution must be utilized to separate the signals of the most similar structures. As mentioned earlier, in reverse-phase HPLC the stationary phase is nonpolar. A nonpolar mobile phase will therefore have higher eluent strength than a polar one. For that reason, we'll start out with a mainly polar mobile phase (milliQ water) and then gradually increase the percentage of a nonpolar mobile phase (acetonitrile). Table 6-3 displays the HPLC settings and gradient elution chosen as a first attempt. The sample was a mixture of all the five compounds with a concentration of 70 mg/L each.

Table 6-3. Displays the gradient elution of the first attempt at HPLC-separation of the five compounds. The column temperature was 20°C and the injection volume 2.0 µL.

Time (min):	MilliQ water (%):	Acetonitrile (%):	Flow (mL/min):
0.0	99.0	1.0	0.200
10.0	0.0	100.0	0.200

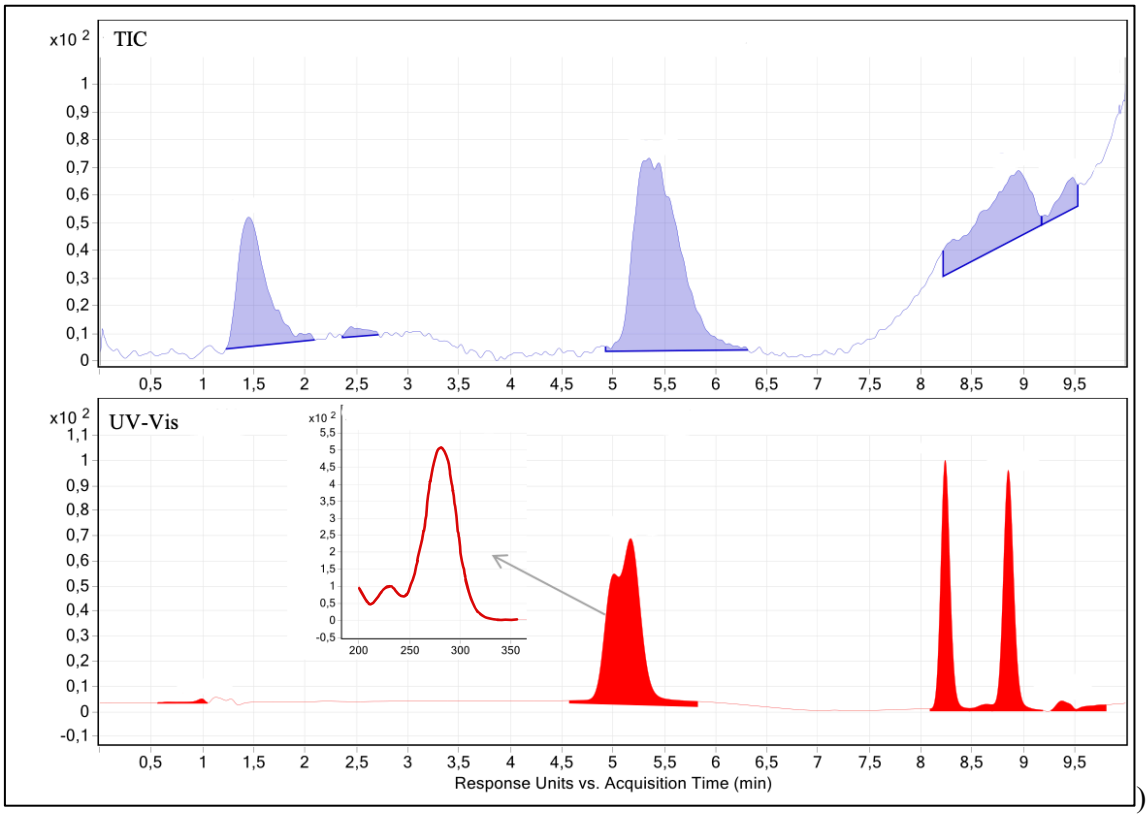


Figure 6-2. Displays the total ion chromatogram (TIC and the UV-Vis signals obtained when using the HPLC settings listed in table 6-3.

As table 6-1 suggests, four UV-Vis signals and three MS signals are expected. The UV signal eluted at 5 minutes is suspected to be two signals overlapping. To separate these signals, we can slow down the gradient elution in this area and increase the temperature to get more narrow peaks. The signal in the TIC at 1.5 minutes contains the MS spectrum of levulinic acid (figure B-8) and the signal at 5.5 minutes contains the MS spectrum of HMF (figure B-2).

The UV-Vis detector in the instrument is physically positioned before the MS detector. The signal obtained with UV-Vis will therefore appear approximately half a minute earlier than the MS signal for the same compound. One of the UV signals at 5 minutes is therefore suspected to belong to the HMF signal at 5.5 minutes. The λ_{max} value of the UV signal is around 280 nm which fits with both HMF and FUR.

It seems that the compound 5-methylfurfural is not easily ionized by positive ESI. Its signal intensity during direct MS analysis is generally weaker than expected for the concentration, and during this sample acquisition (figure 6-2), its signal was completely overlooked. To accommodate this, the MS scan type was changed to selected ion monitoring (SIM). From here

on, all the HPLC-MS spectra will be obtained with the MS SIM scan type. This scan type will monitor ions of specific m/z values chosen by the operator (table 6-4) and discard all other signals.

Table 6-4. Displays the m/z values of the chosen ions to scan after with the MS SIM scan type.

Compound:	**m/z values:**
HMF:	127
5-Methylfurfural:	111
Levulinic acid:	139, 117, 99

Furthermore, the instrument experienced some leakage at a flow rate of 0.200 mL/min. This issue stemmed from a part in the instrument that needed replacement. To overcome this problem, the flow rate was reduced. Reducing the flow rate reduces the pressure needed to force the mobile phase through the packed column, which helps minimize the leakage. Table 6-5 displays the final HPLC settings and gradient elution chosen after several attempts and adjustments. Figure 6-3 displays the HPLC-MS/UV results when using the settings given in table 6-5 and SIM scan.

Table 6-5. Displays the final gradient elution chosen for HPLC-separation of the five compounds. The column temperature was 60°C and the injection volume 2.0 µL.

Time (min):	**MilliQ water (%):**	**Acetonitrile (%):**	**Flow (mL/min):**
0	99.0	1.0	0.180
5	50.0	50.0	0.180
7	40.0	60.0	0.180
12	0.0	100.0	0.180
13	0.0	100.0	0.180

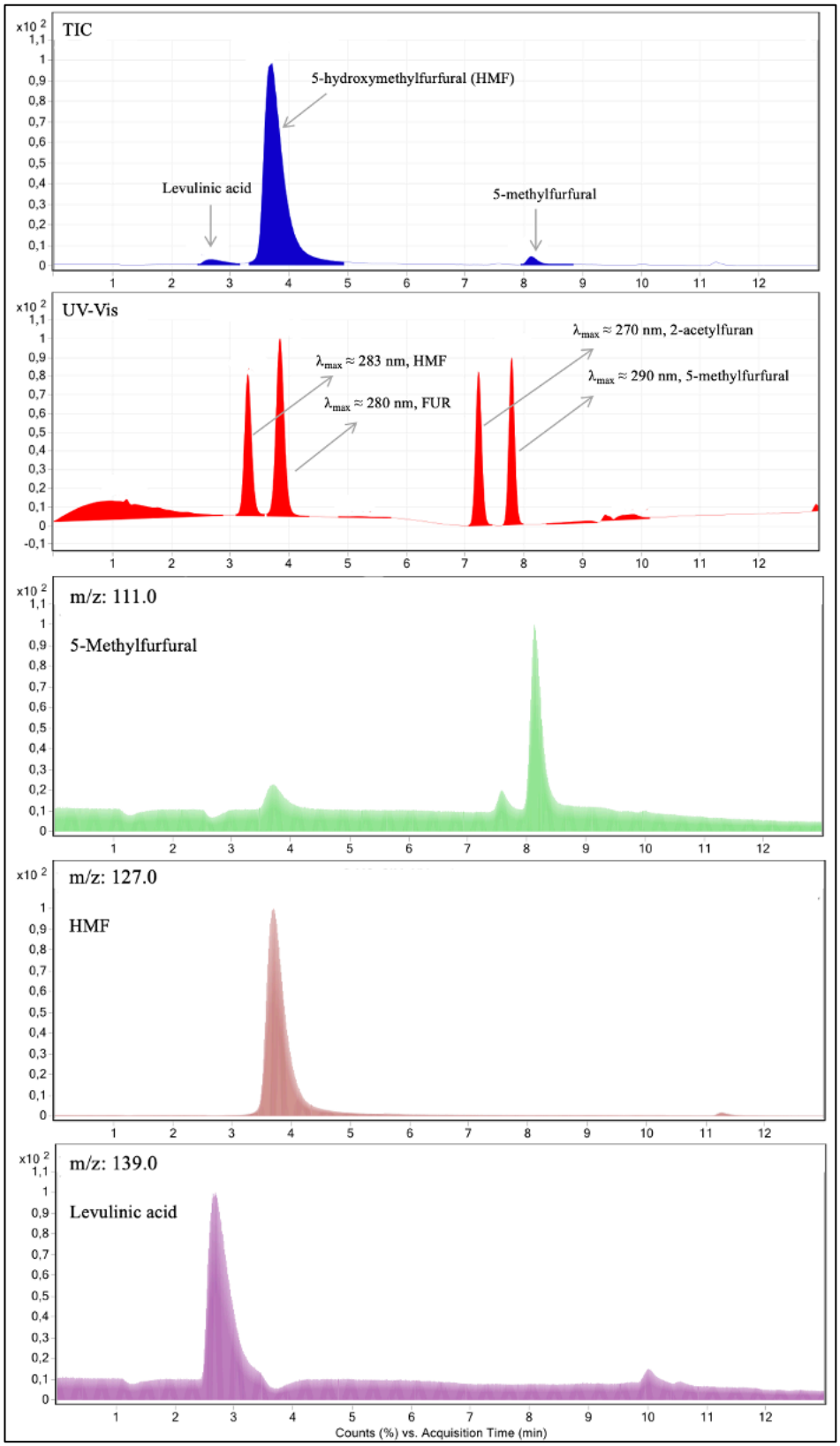


Figure 6-3. Displays the obtained TIC, UV-Vis signals, and m/z values monitored with MS SIM scan when using the HPLC settings listed in table 6-5.

When inspecting the UV-Vis signals in figure 6-3, it seems that the signals of FUR and HMF did indeed overlap in figure 6-2. In figure 6-3 they elute closely, but completely separated. The signal of HMF is of high intensity in the TIC, while the signals of levulinic acid and 5-MF are only displayed as small bumps in the chromatogram. However, when searching for the m/z values monitored by the SIM scan their signals have the same intensity as HMF (1.0 x 10^2). In the three lower chromatograms (m/z values of 111, 127 and 139) there are a few extra unexpected bumps (figure 6-3). These bumps may arise from the fact that the leakage causes varying pressure in the LC column, this will affect the elution and retention times of the compounds. They may also arise from the elution of system contaminants with the same m/z value as the main components.

The λ_{max} values of the UV-Vis signals are marked in the spectrum. The maximum values of the absorptions fit with the theoretical values of FUR, HMF, 2-AF and 5-MF. But a UV-Vis signal doesn't contain as much identifying information as a mass spectrum. The signals of FUR and 2-AF were therefore confirmed by running their individual standard solutions on the HPLC-MS/UV method and checking the retention times.

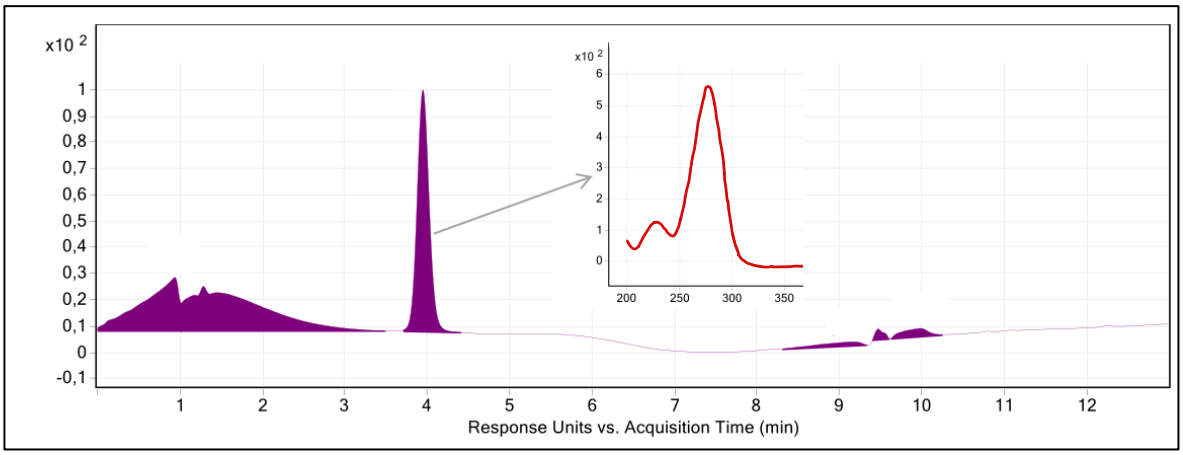


Figure 6-4. Displays the UV-Vis spectrum obtained when running the standard solution of FUR (70mg/L) on the developed HPLC-MS/UV method.

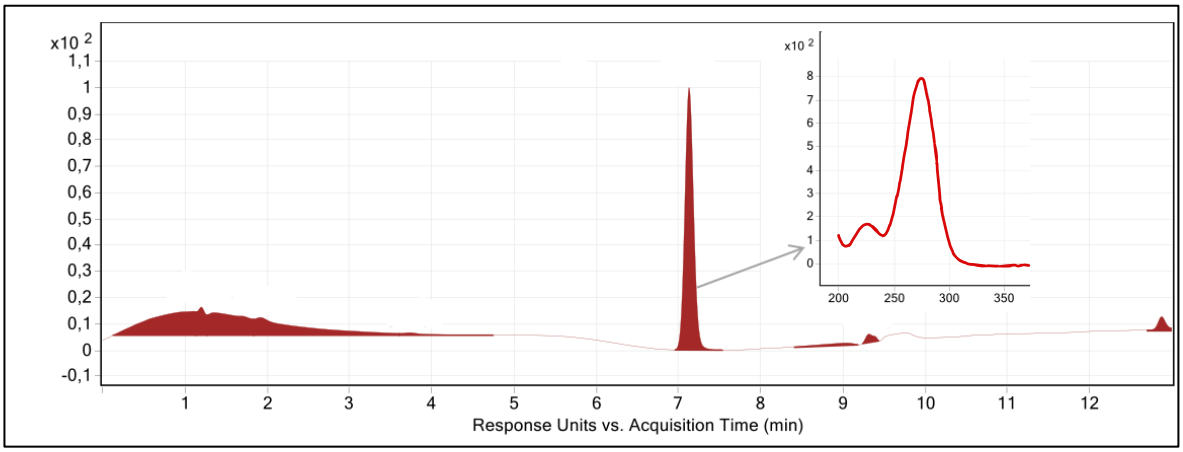


Figure 6-5. Displays the UV-Vis spectrum obtained when running the standard solution of 2-Acetylfuran (70mg/L) on the developed HPLC-MS/UV method.

Figure 6-4 and 6-5 displays the resulting UV-Vis spectra when running the standard solutions of FUR and 2-AF on the developed HPLC-MS/UV method, respectively. When comparing the retention times of the signals suspected to be FUR and 2-AF in figure 6-3 with the retention times in figure 6-4 and 6-5 we can confirm that the signals are indeed obtained from FUR and 2-AF.

With this confirmation, a HPLC-MS/UV method that can both separate and analyse the five chosen compounds has been established.

6.1.1 Calibration Curves

Calibration curves were prepared for each of the five compounds. For the compounds with both MS and UV-Vis signal, two calibration curves were prepared. These are given in figure 6-6 to figure 6-12 below.

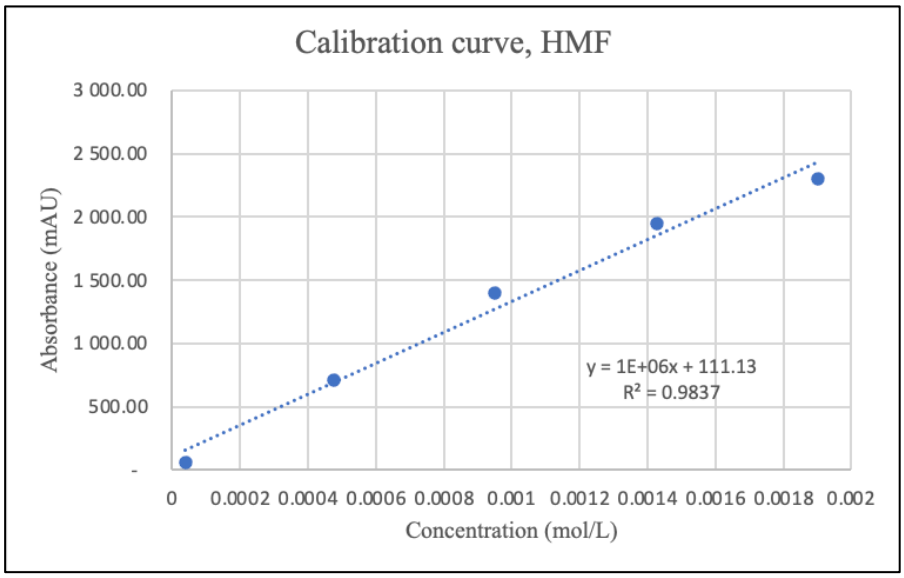


Figure 6-6. UV-Vis calibration curve for 5-hydroxymethylfurfural. Absorbance (mAU) vs. concentration (M).

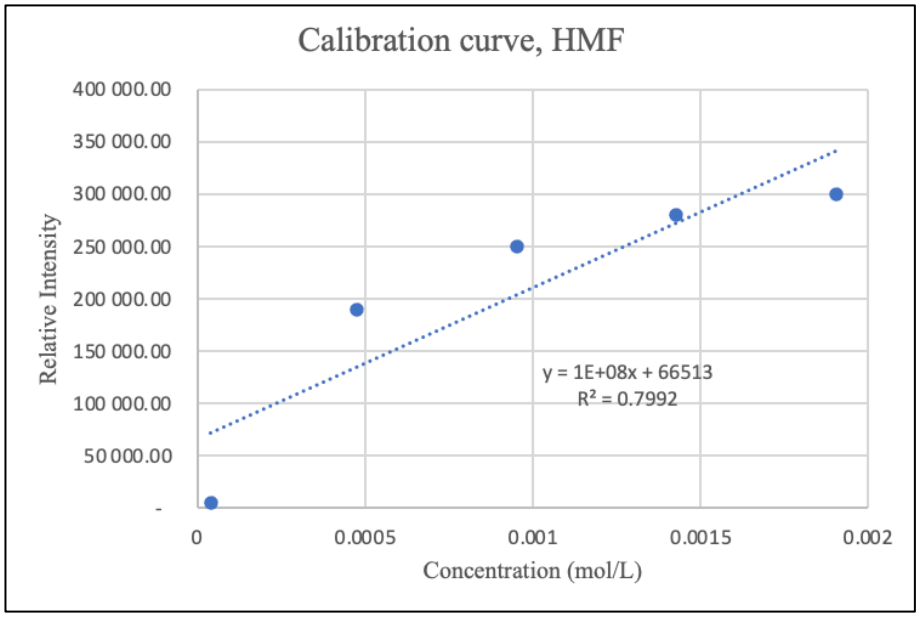


Figure 6-7. MS Calibration curve for 5-hydroxymethylfurfural. Relative Intensity vs. concentration (M).

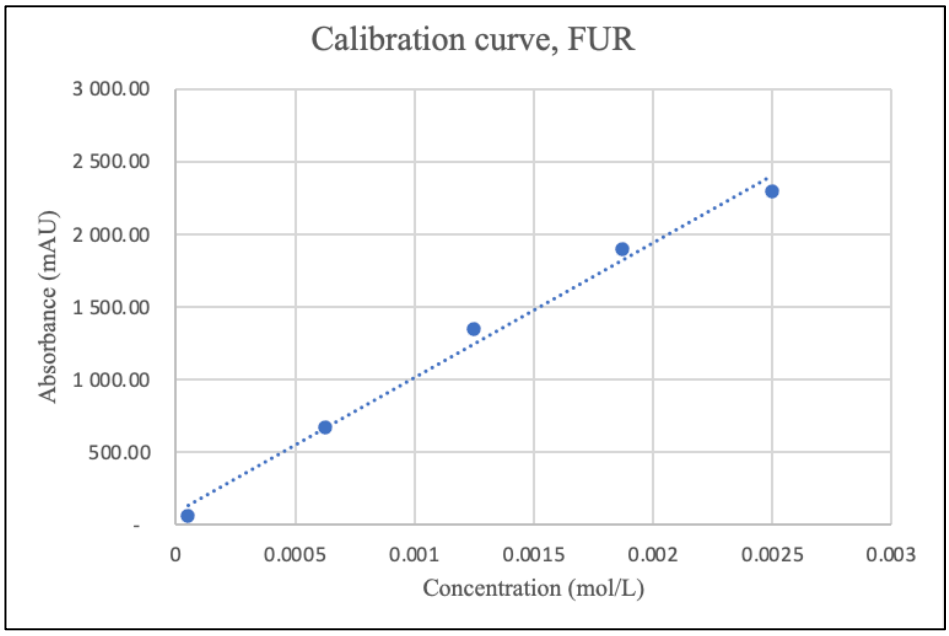


Figure 6-8. UV-Vis calibration curve for furfural.

Absorbance (mAU) vs. concentration (M).

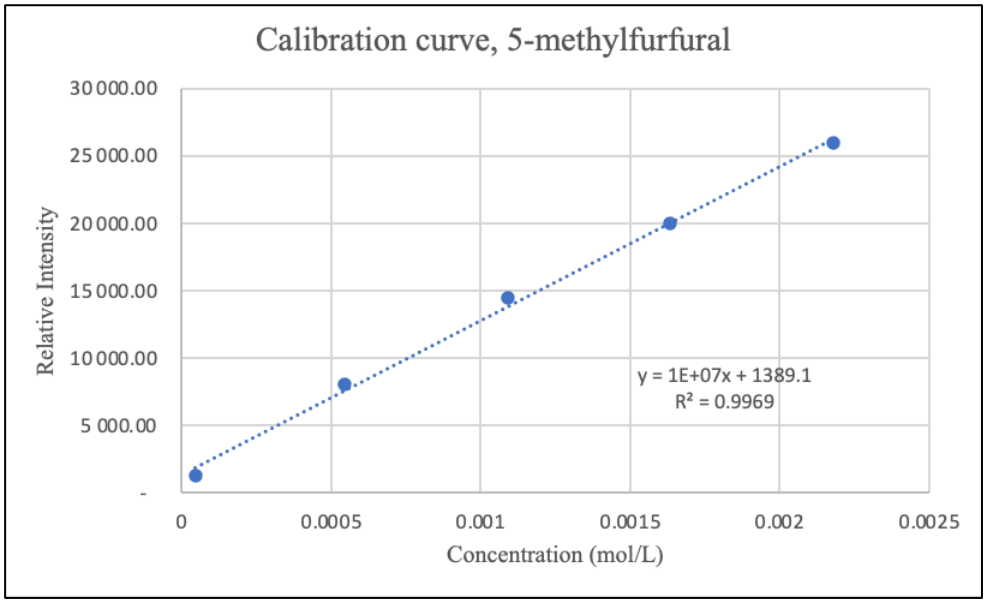


Figure 6-9. MS calibration curve for 5-methylfurfural.

Relative intensity vs. concentration (M).

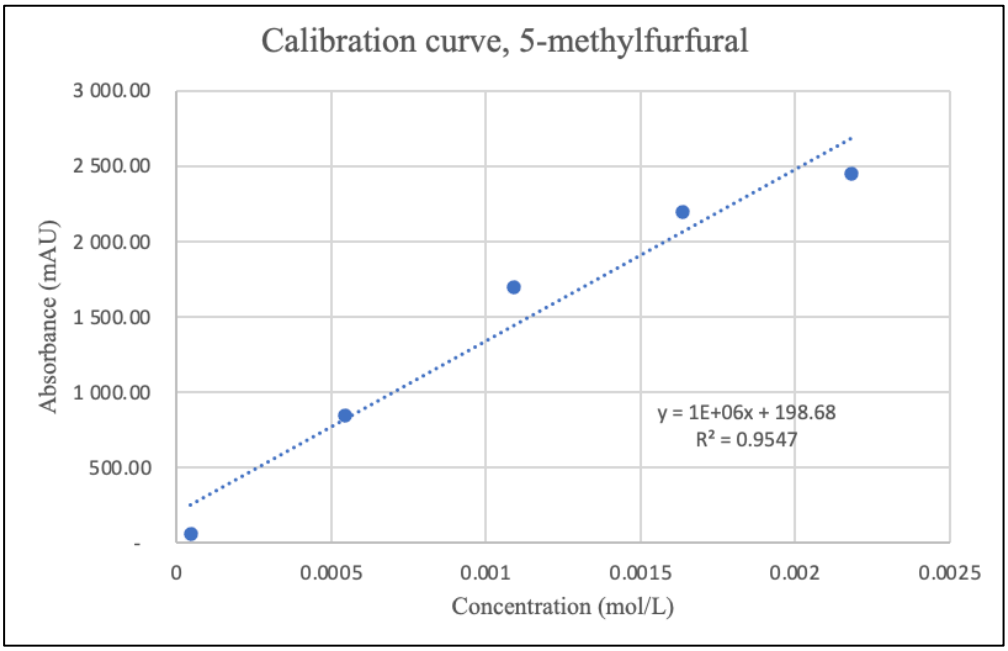


Figure 6-10. UV-Vis calibration curve for 5-methylfurfural.

Absorbance (mAU) vs. concentration (M):

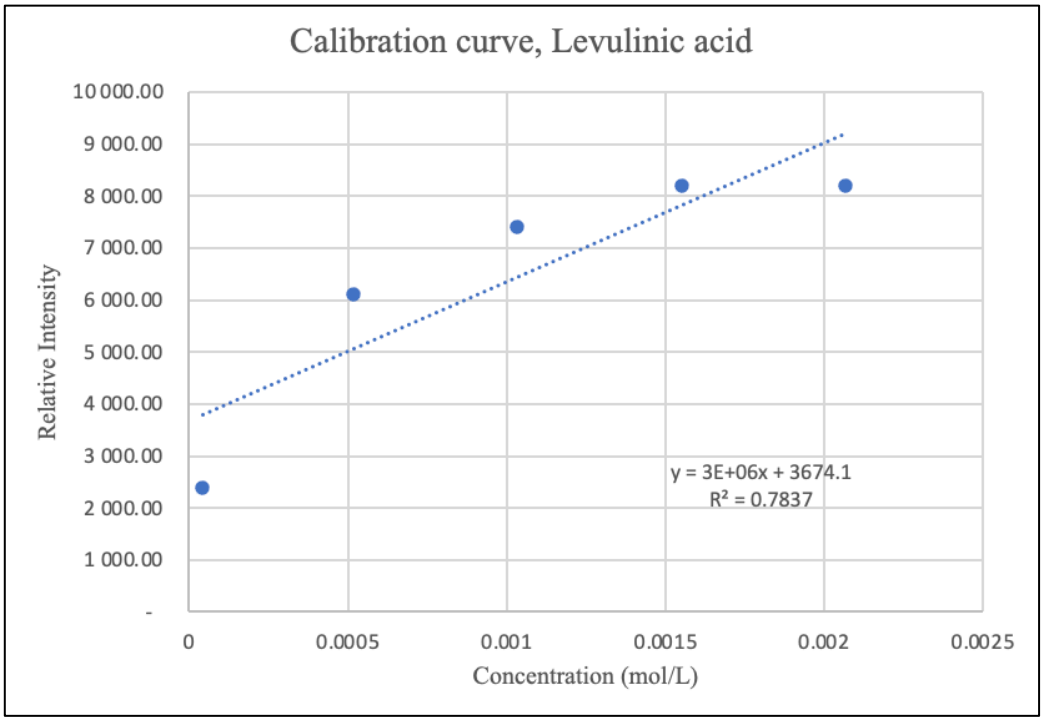


Figure 6-11. MS calibration curve for levulinic acid.

Relative intensity vs. concentration (M).

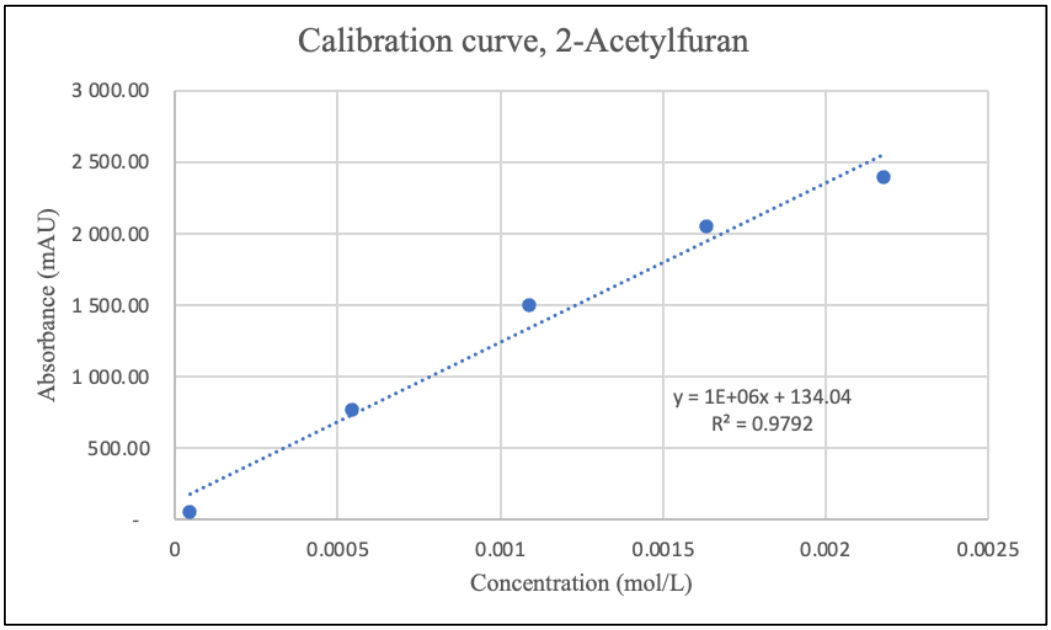


Figure 6-12. UV-Vis calibration curve for 2-acetylfuran.
Absorbance (mAU) vs. concentration (M).

The maximum concentration levels of the calibration curves were chosen with consideration for the MS-detector. Too high sample concentrations can overload the detector. The levels chosen were 5, 60, 120, 180, and 240 mg/L. The concentrations were converted to exact molar concentrations of each compound before preparation of calibration curves. The product solutions to be quantified with the developed HPLC-MS/UV method will need to be diluted before analysis.

The maximum concentrations in the MS calibration curves for HMF and levulinic acid (figure 6-7 and figure 6-11) are past the linear response range for mass spectrometry. At this point, increasing the concentrations of the analytes will not give a response of higher relative intensity. The maximum concentrations in the UV-Vis calibration curves for 5-MF and 2-AF (figure 6-10 and figure 6-12) are also past the linear response range for UV-Vis spectroscopy. This can be seen to a lesser extent in the UV-Vis calibration curves for HMF and FUR as well (figure 6-6 and figure 6-8).

6.2 Experimental Design

6.2.1 Results from the first experimental design (qNMR)

A 2^{4-1} fractional factorial design (FFD) was used during the screening of the biphasic reaction system with the chosen main factors and cherry rejects as substrate. The four main factors investigated were: reaction temperature, residence time, amount of aqueous solvent (water), and amount of organic solvent (MIBK).

The high and low levels of the main factors were selected based on literature and the results from Molnes, 2021. The chosen high, low, and centre point (CP) levels of the main factors are displayed in table 6-6. The layout of the experimental design is presented in table 6-7, the experiments were randomized before execution. The complete layout of the experimental design, including constant parameters, is presented in Appendix C (table C-1). The results from the screening FFD, quantified with qNMR, are presented in table 6-8. The concentrations of FUR and HMF in the aqueous and organic product solutions are given in Appendix C (table C-2).

Table 6-6. The high, low, and centre point values of the four main factors investigated in the screening FFD. T is temperature, t is residence time, Vaq is volume of aqueous solvent (water), and Vorg is volume of organic solvent (MIBK).

Level:	**T (°C):**	**t (min):**	**Vaq (mL):**	**Vorg (mL):**
-	170	30	1.0	1.0
0	190	75	2.5	2.5
+	210	120	4.0	4.0

Table 6-7. The layout of the experimental design for the screening FFD.

Experiment #:	T (°C):	t (min):	Vaq (mL):	Vorg (mL):
1	170	30	4.0	1.0
2	210	30	4.0	1.0
3	170	120	4.0	1.0
4	210	120	4.0	1.0
5	170	30	1.0	4.0
6	210	30	1.0	4.0
7	170	120	1.0	4.0
8	210	120	1.0	4.0
CP1	190	75	2.5	2.5
CP2	190	75	2.5	2.5
CP3	190	75	2.5	2.5
CP4	190	75	2.5	2.5

Table 6-8. The calculated mass percent (m%) yield of FUR and HMF for the screening FFD, quantified with qNMR.

Experiment #:	FUR yield (m%):	HMF yield (m%):
1	0.02	1.23
2	1.16	18.43
3	0.26	8.54
4	0.77	14.85
5	0.03	4.30
6	0.59	19.40
7	0.29	12.17
8	1.12	12.67
CP1	1.02	19.14
CP2	0.43	16.63
CP3	0.56	16.81
CP4	0.50	17.21

Figure 6-13 displays a selection of pictures of the product solutions prior to phase separation. There seems to be a trend where higher concentration of HMF (and to some degree FUR) gives a darker colour to the product solution. Pictures of all the product solutions from the screening FFD is presented in figure C-1 in Appendix C. Figure 6-14 and 6-15 presents the NMR spectra of the aqueous product solution and the organic product solution of experiment 8, respectively. All the NMR spectra obtained from the screening FFD are presented in Appendix C figure C-2 to figure C-25.

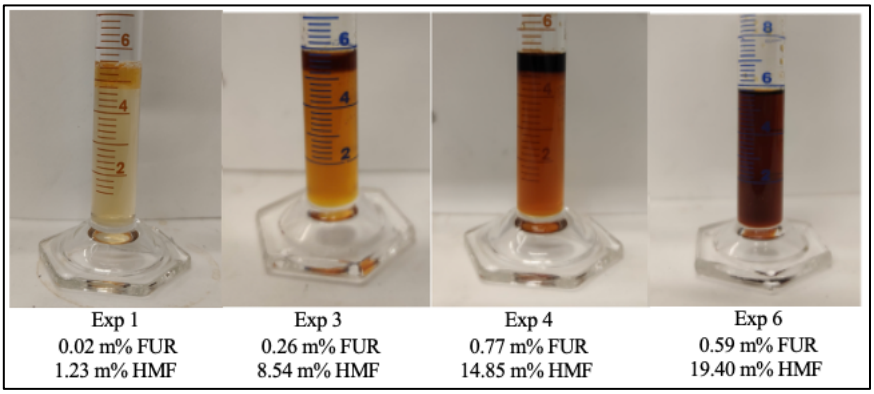


Figure 6-13. Pictures of four selected product solutions from the screening FFD, prior to phase separation. The product solutions are ordered from lowest to highest concentration of HMF.

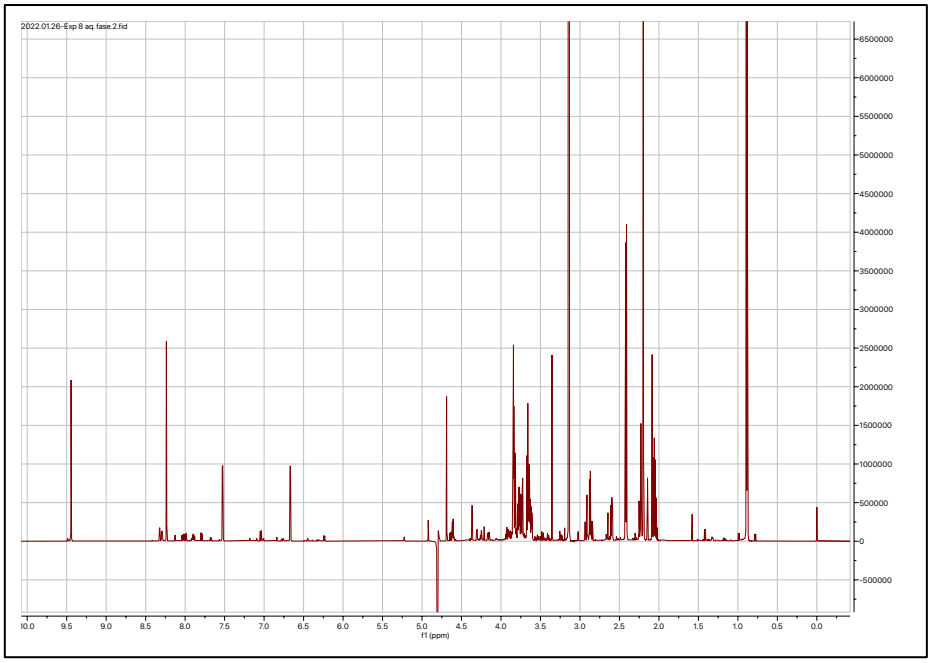


Figure 6-14. The NMR spectrum of the aqueous product solution of experiment 8.

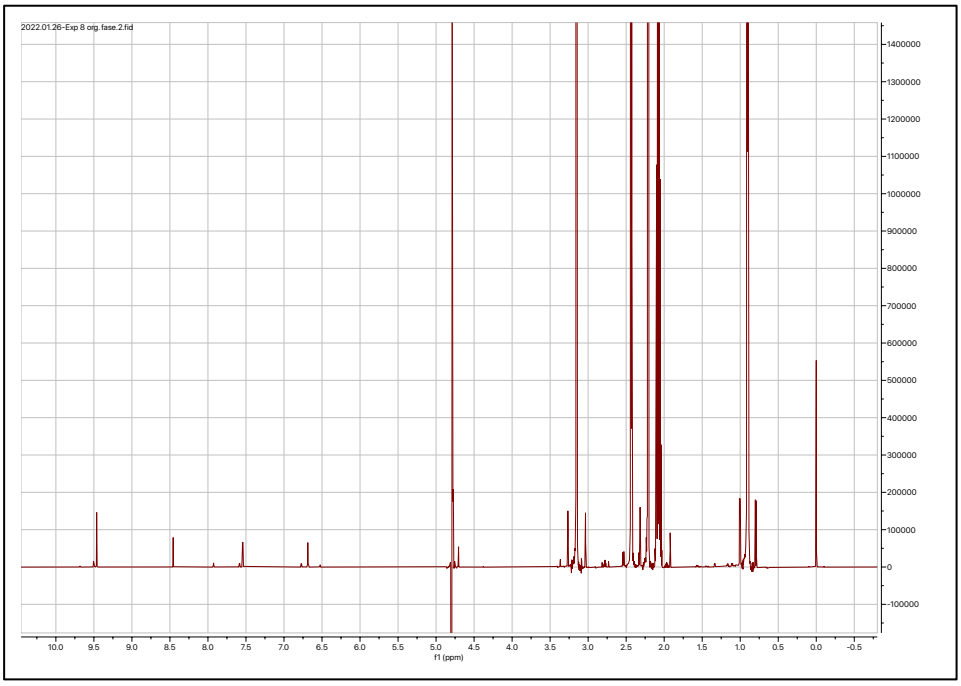


Figure 6-15. The NMR spectrum of the organic product solution of experiment 8.

Figure 6-16 displays the biplot obtained from PCA analysis of the screening FFD. The dataset is standardised, and the analysis employs mean centring. The centre points (marked in blue) should lie exactly on top of each other, as the experiment conditions are the exact same for all the centre points. This is the case for all the centre point experiments except CP1. The CP1 experiment is therefore suspected to be an outlier and removed from the dataset. With the exception of CP1, there is an acceptable reproducibility between the centre point experiment results. Figure 6-17 displays the biplot obtained from PCA analysis of the screening FFD, with the object CP1 excluded.

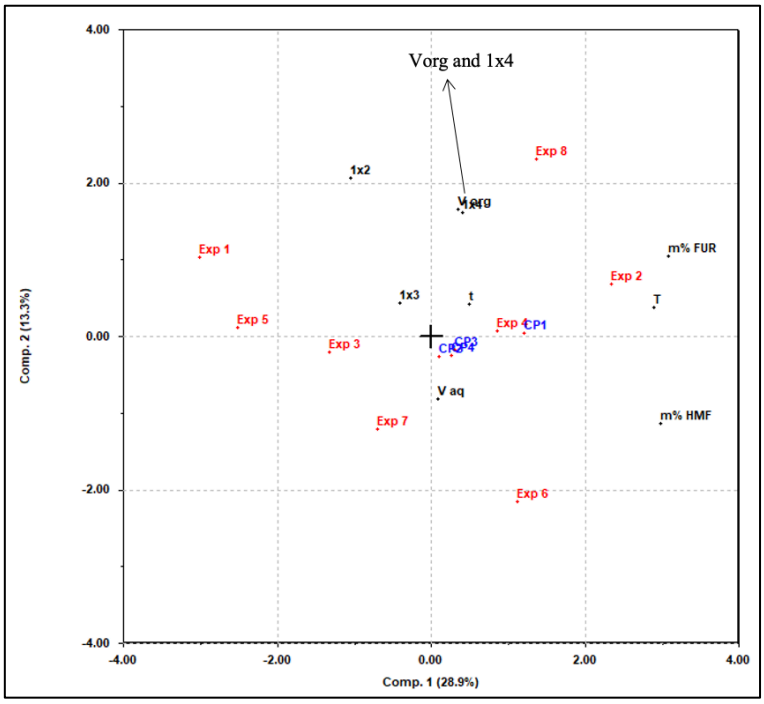


Figure 6-16. Displays the biplot obtained from PCA analysis of the screening FFD. The experiments are marked in red, the centre points in blue, and the main factors, two-factor interactions, and responses are marked in black.

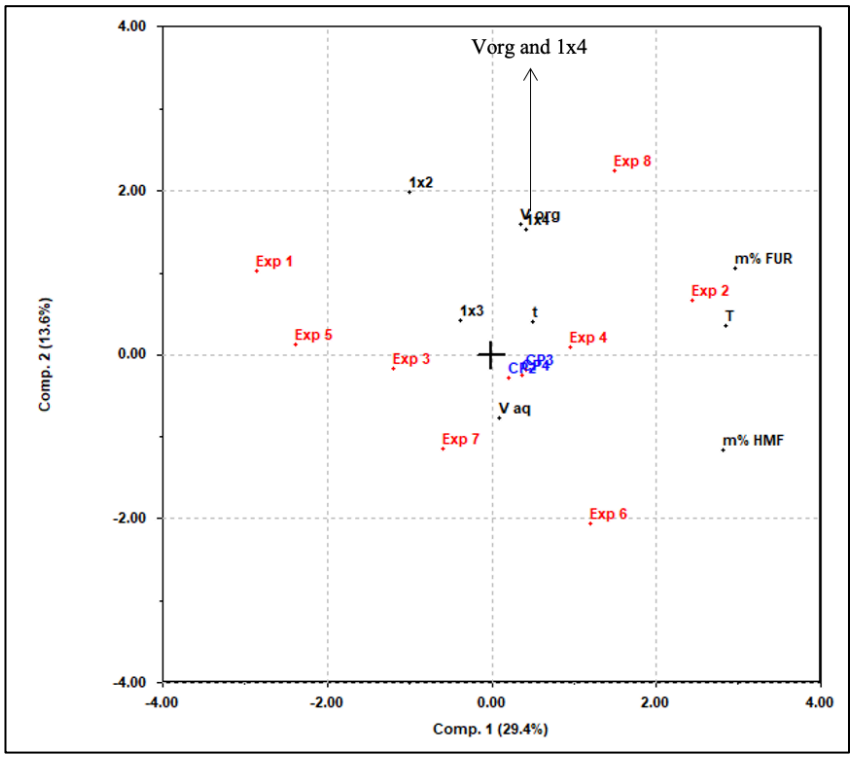


Figure 6-17. Displays the biplot obtained from PCA analysis of the screening FFD, the object CP1 is excluded. The experiments are marked in red, the centre points in blue, and the main factors, two-factor interactions, and responses are marked in black.

When investigating the biplot in figure 6-17, it is clear that both HMF and FUR yield is positively correlated with a higher temperature (T). Residence time (t) is also correlated with HMF and FUR yield. There is a negative correlation between volume of aqueous and organic solvent. The experiments most correlated with yield will have the highest yields. Experiment 2, 4, 6, 8, and the centre point experiments are therefore the experiments with the highest yield of HMF and FUR. This fits with the results listed in table 6-8.

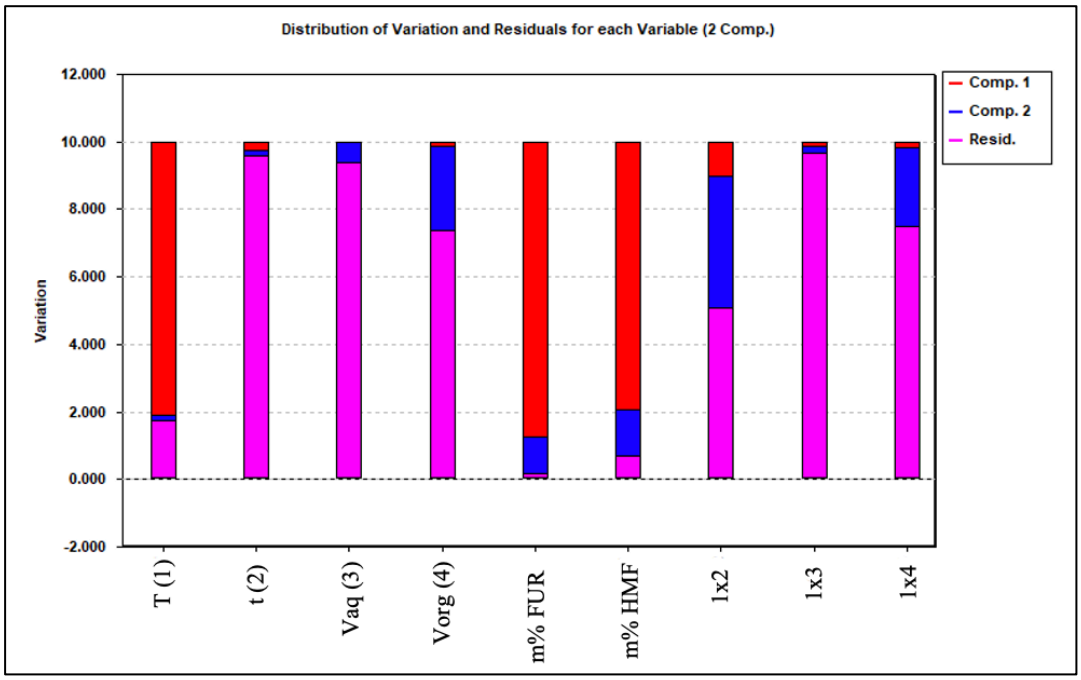


Figure 6-18. Illustrates the distribution of variation and residuals for each variable and two-factor interaction on each of the two principal components.

The distribution of variation and residuals for each variable on each principal component (PC) is illustrated in figure 6-18. The main factors residence time, volume of aqueous solvent, and volume of organic solvent contribute the least to the model. In other words, a small share of the variation of the variables are explained by the PCs. Temperature is the most important main factor to the model and the two-factor interaction between temperature and time is the most important interaction.

The yields of FUR were generally low while the HMF yields were quite high. This suggests that the reaction system, conditions, and/or substrate is more suitable for HMF production. The variation described by each PC for FUR and HMF yield are quite different. Consequently, separate models need to be made for each of the responses.

6.2.1.1 Regression model HMF

The purpose is to find a mathematical expression that can estimate the yield of HMF when using the selected biphasic reaction system. The results from the PLS analysis are presented in figure 6-19 to figure 6-25 below.

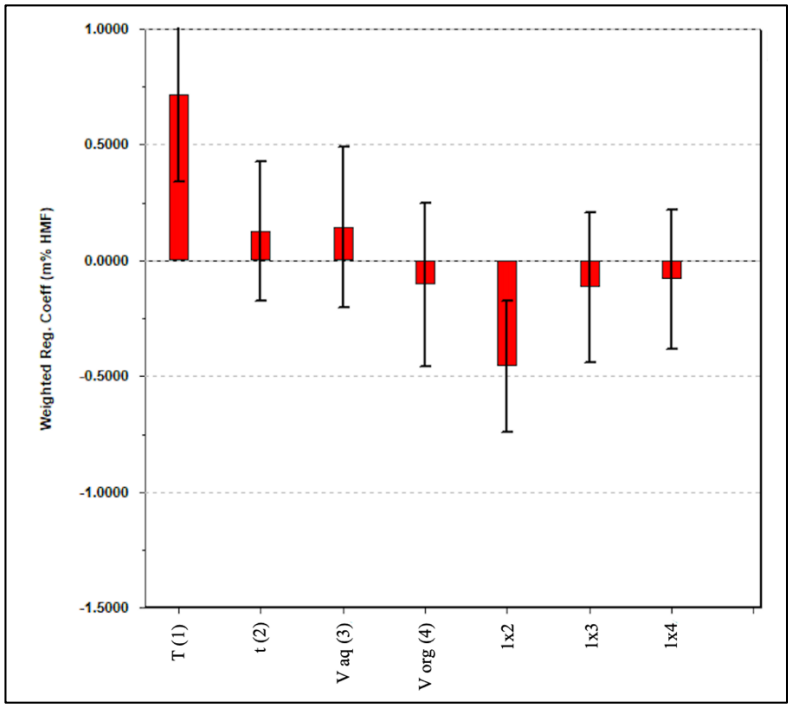


Figure 6-19. Presents weighted regression coefficients for the main factors and two-factor interactions. The number behind each main factor corresponds to the number in each two-factor interaction.

Figure 6-19 presents the weighted regression coefficients for the obtained linear regression model. The figure confirms that temperature is the most important main factor regarding HMF yield. The interaction between temperature and time (1x2) is the most important two-factor interaction with a considerable impact on the model. The importance of this two-factor interaction is further investigated in figure 6-20 with the use of an interaction graph.

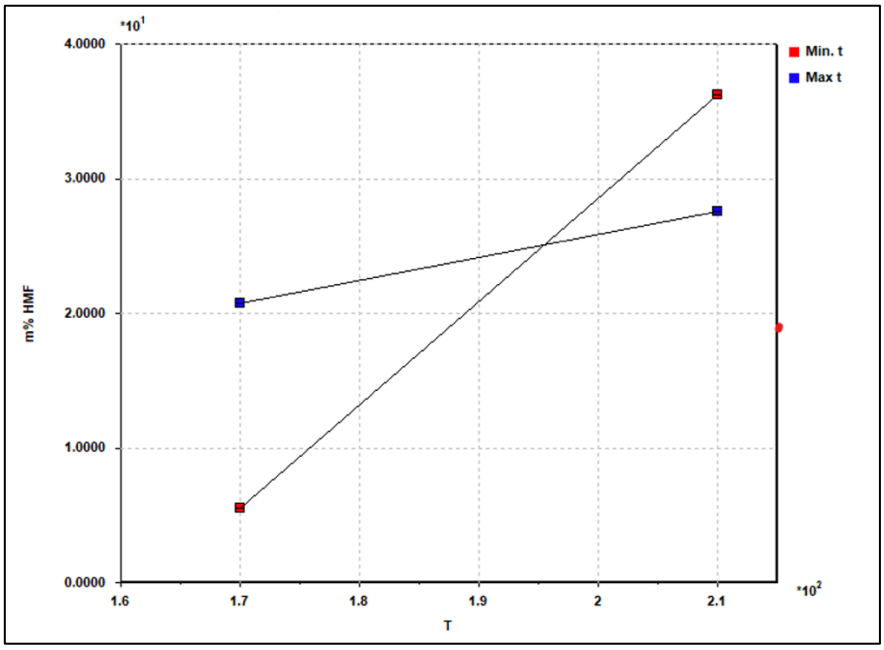


Figure 6-20. Interaction graph displaying the effect of minimum and maximum values of temperature and time on HMF yield.

The two lines in figure 6-20 are far from parallel, this confirms that the two-factor interaction between temperature and time has a significant effect on the HMF yield. According to the figure, increasing the temperature from 170°C to 210°C has a small positive effect on the HMF yield for experiments conducted with a 120 minute residence time. This can be seen by the gentle slope of the line between the two blue boxes. However, increasing the temperature from 170°C to 210°C for experiments conducted with a 30 minute residence time, has a substantial positive effect on HMF yield. This can be seen by the steep slope of the line between the two red boxes.

Figure 6-21 and 6-22 displays plots of predicted vs. measured HMF yield with and without the centre point experiments, respectively. When comparing the plots, it can be seen that including CPs makes the regression model non-linear. In figure 6-21 the CP experiments are situated on a horizontal line. This is because the model predicts the same value for each CP experiment. However, the measured yields of each CP varied to some degree, causing them to spread along the x-axis. This indicates that the response surface of HMF yield is curved, meaning that quadratic terms are needed in the model to give a more accurate prediction of HMF yield. The curved response surface of HMF yield is further verified with the contour plot in figure 6-23.

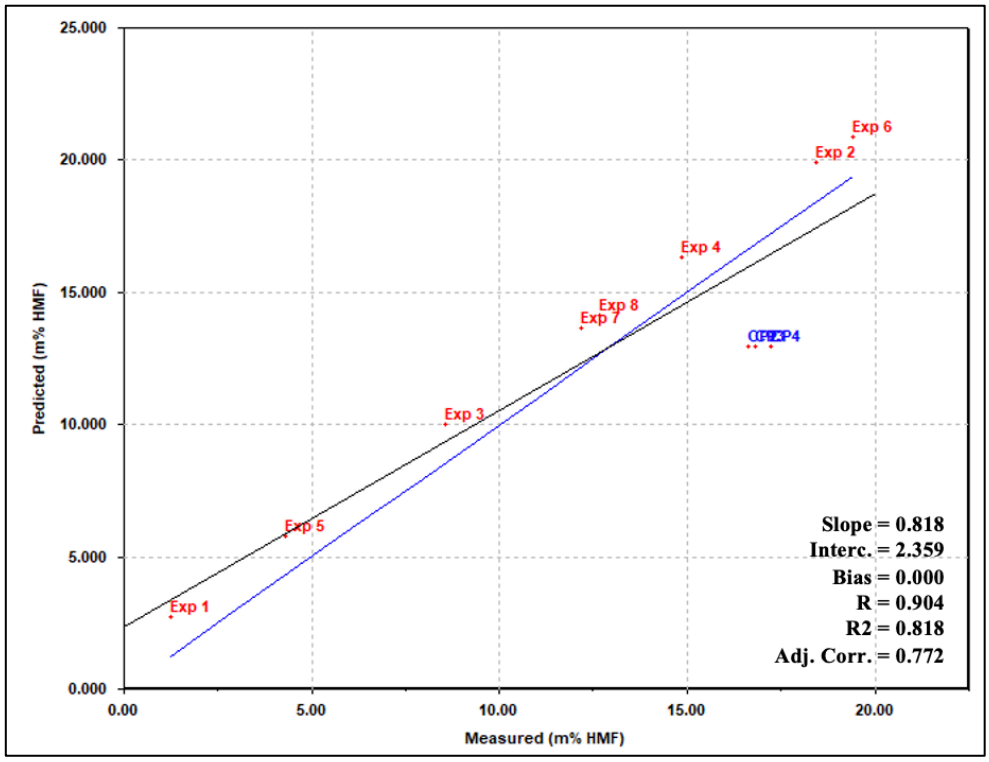


Figure 6-21. Plot of predicted vs. measured values of HMF yield (m%), predicted by the obtained linear regression model.

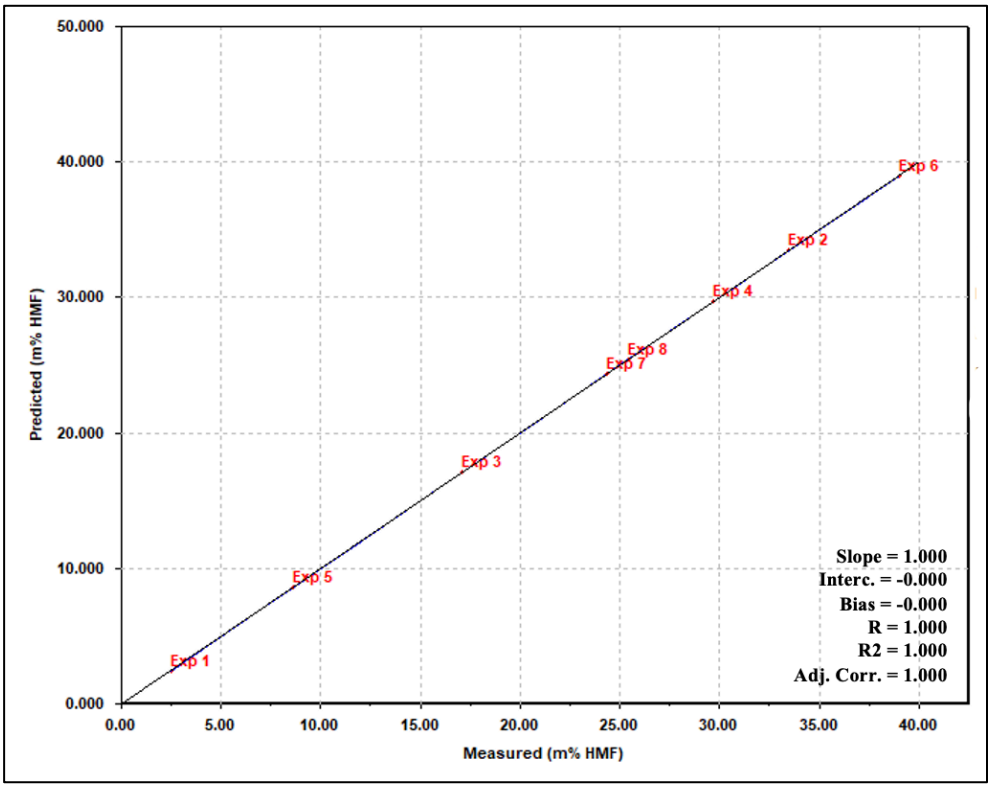


Figure 6-22. Plot of predicted vs. measured values of HMF yield (m%), predicted by the obtained linear regression model. Excluding the centre point experiments.

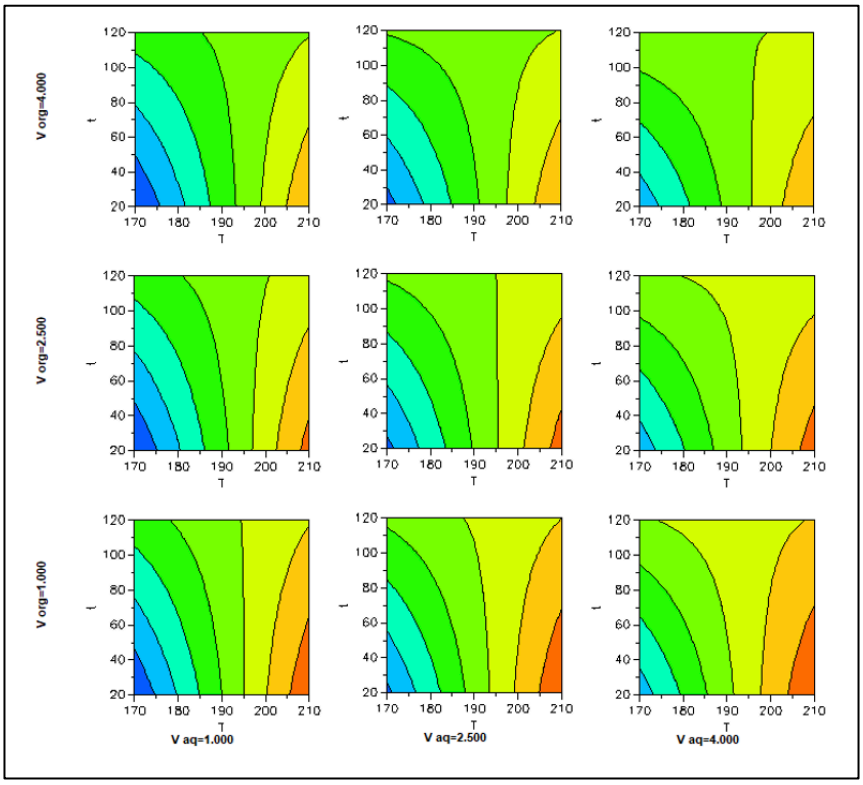


Figure 6-23. Contour plots showing the curved response surface of HMF yield within the experimental domain.

Figure 6-23 displays contour plots of the response surface of HMF yield. The lines are clearly curved. This confirms that HMF does indeed have a curved response surface, and the model for HMF yield should include quadratic terms.

Figure 6-24 presents the weighted regression coefficients, including quadratic terms. The quadratic terms included is TxT (1x1) and txt (2x2). The figure shows the importance of the TxT quadratic term for the model predicting HMF yield. This further confirms that the temperature is the most important and influential variable for achieving high HMF yields.

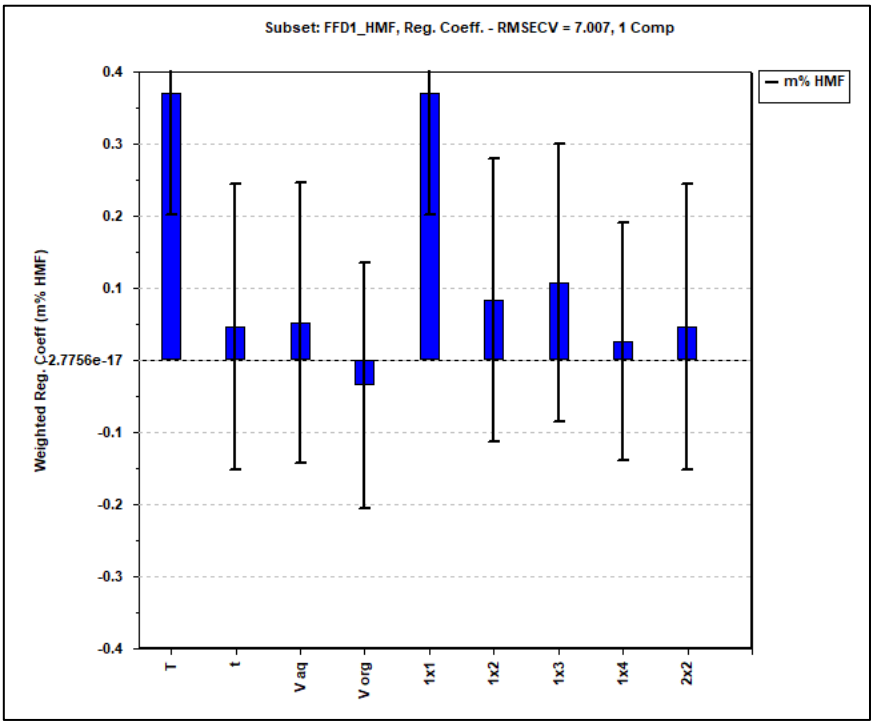


Figure 6-24. Presents weighted regression coefficients for the main factors, two-factor interactions, and quadratic terms. (1x1 = T x T, 2x2 = t x t).

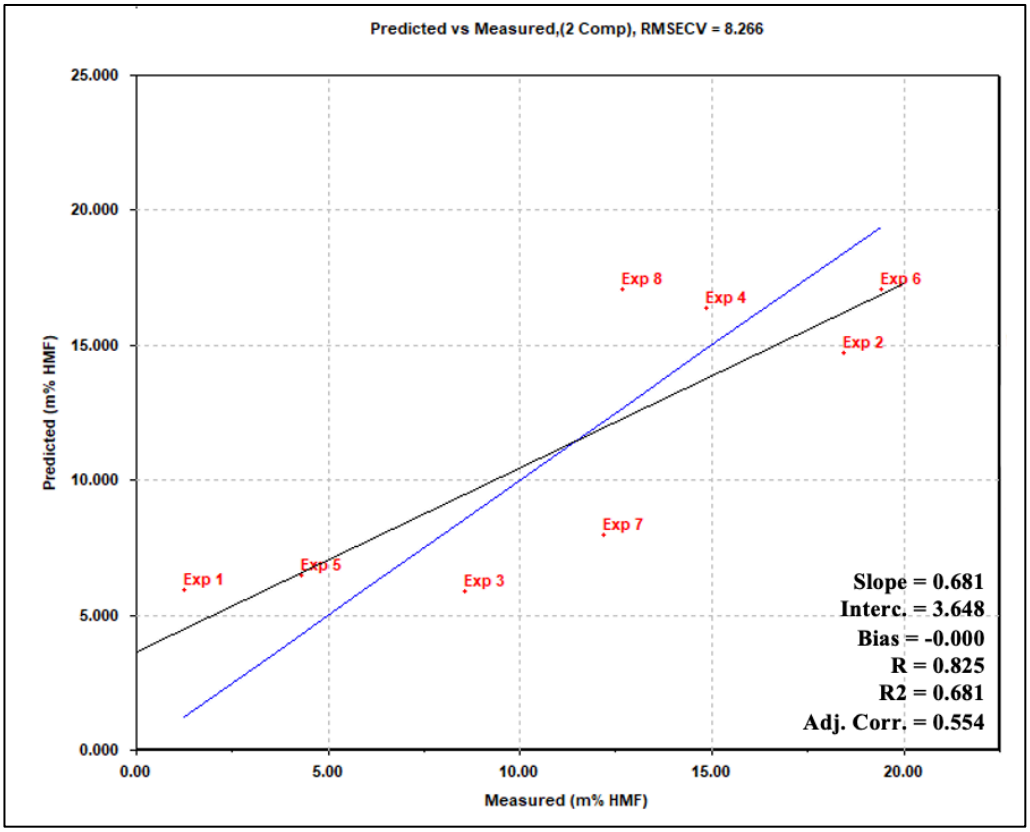


Figure 6-25. Plot of predicted vs. measured values of HMF yield (m%), predicted by the final regression model. Excluding the centre point experiments.

Figure 6-25 displays the plot of predicted vs. measured values of HMF yield obtained by the final regression model. The black line is the fitted regression line for the model. The coefficient of correlation for the regression model is 0.681. The model does not describe the system with complete accuracy, in which case R^2 would be equal to 1. This indicates that other factors not included in the model could affect the formation of HMF. The mathematical expression for the regression model (Equation 6-1) is given below.

$$Y\ (m\%\ HMF) = -3.685 + 0.396T + 0.016t + 0.035V_{aq} - 0.066V_{org} + 0.396T^2 + 0.034Txt + 0.086TxV_{aq} - 0.002TxV_{org} + 0.016t^2 \quad \text{(Equation 6-1)}$$

Where T is temperature, t is residence time, V_{aq} is volume of aqueous solvent, V_{org} is volume of organic solvent, TxV_{aq} and TxV_{org} are the two-factor interactions between temperature and volume of solvent, Txt is the two-factor interaction between temperature and residence time, T^2 is the quadratic term for temperature, and t^2 is the quadratic term for residence time.

6.2.1.2 Regression model FUR

The purpose of this section is to find a mathematical expression that can estimate the yield of FUR when using the selected biphasic reaction system. The results from the PLS analysis are presented in figure 6-26 to figure 6-30 below.

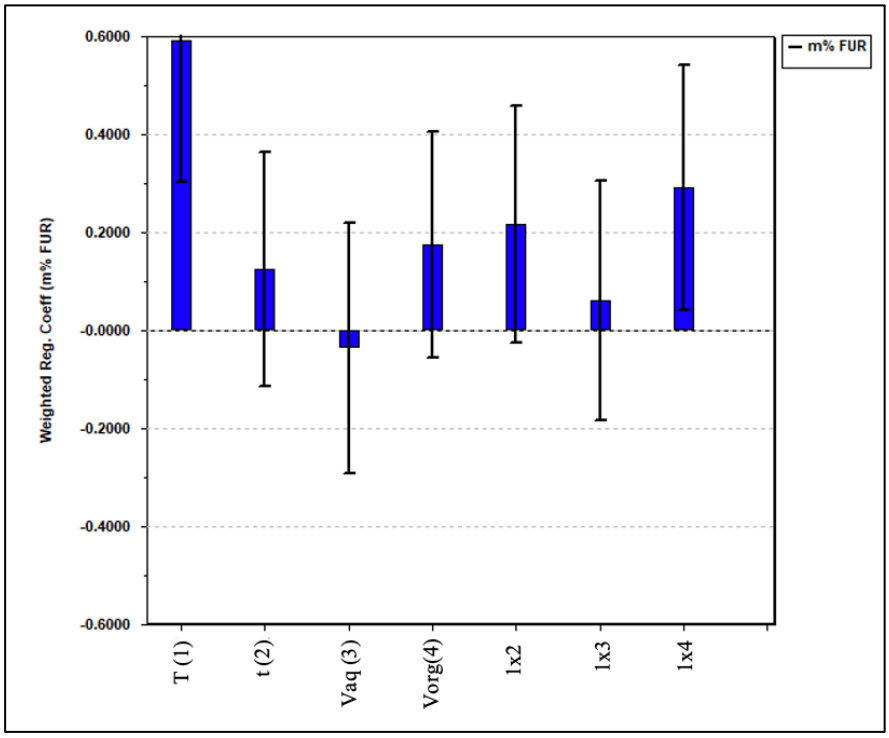


Figure 6-26. Presents weighted regression coefficients for the main factors and two-factor interactions. The number behind each main factor corresponds to the number in each two-factor interaction.

Figure 6-26 presents the weighted regression coefficients for the obtained linear regression model. The figure confirms that temperature is the most important main factor regarding FUR yield as well. Residence time is a more important main factor for FUR yield than HMF. The interaction between temperature and volume of organic solvent (1x4) is the most significant two-factor interaction. The importance of this two-factor interaction is further investigated in figure 6-27 with the use of an interaction graph.

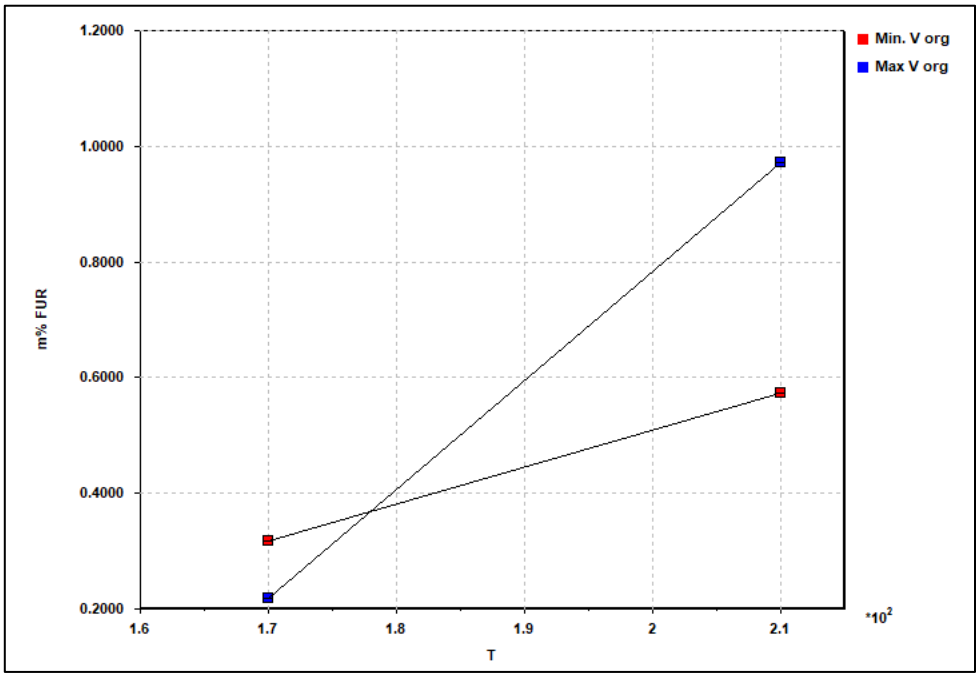


Figure 6-27. Interaction graph displaying the effect of minimum and maximum values of temperature and volume of organic solvent on FUR yield.

The two lines in figure 6-27 are not parallel. This means that the interaction between temperature and volume of organic solvent has a significant effect on the FUR yield. This two-factor interaction needs to be included in the model.

Figure 6-28 and figure 6.29 displays plots of predicted vs. measured FUR yield with and without the centre point experiments, respectively. The centre point experiments doesn't affect the model for FUR yield as much as the model for HMF. The regression coefficient R^2 is 0.908, quite close to 1, with the CPs included. The current regression model, without any quadratic terms, describes the system with an adequate accuracy. However, the CPs are positioned on a horizontal line and do still cause some non-linearity to the model ($R^2 = 0.915$ without the CPs). The response surface of FUR is therefore investigated with the help of contour plots (figure 6-30).

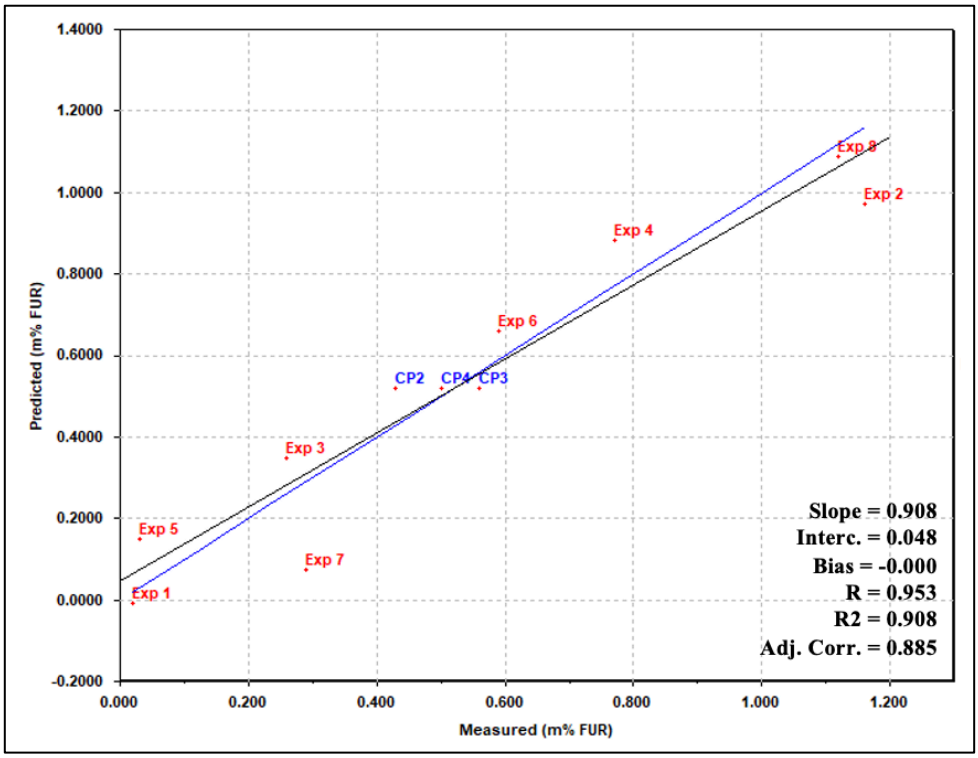


Figure 6-28. Plot of predicted vs. measured values of FUR yield (m%), predicted by the obtained linear regression model.

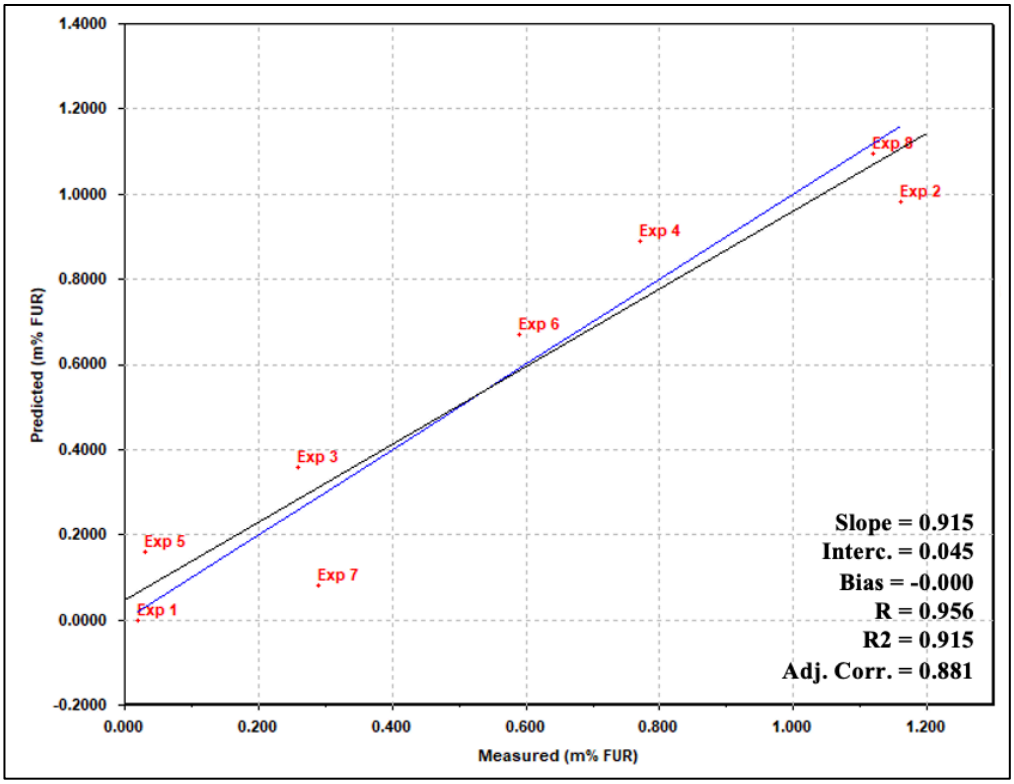


Figure 6-29. Plot of predicted vs. measured values of FUR yield (m%), predicted by the obtained linear regression model. Excluding the centre point experiments.

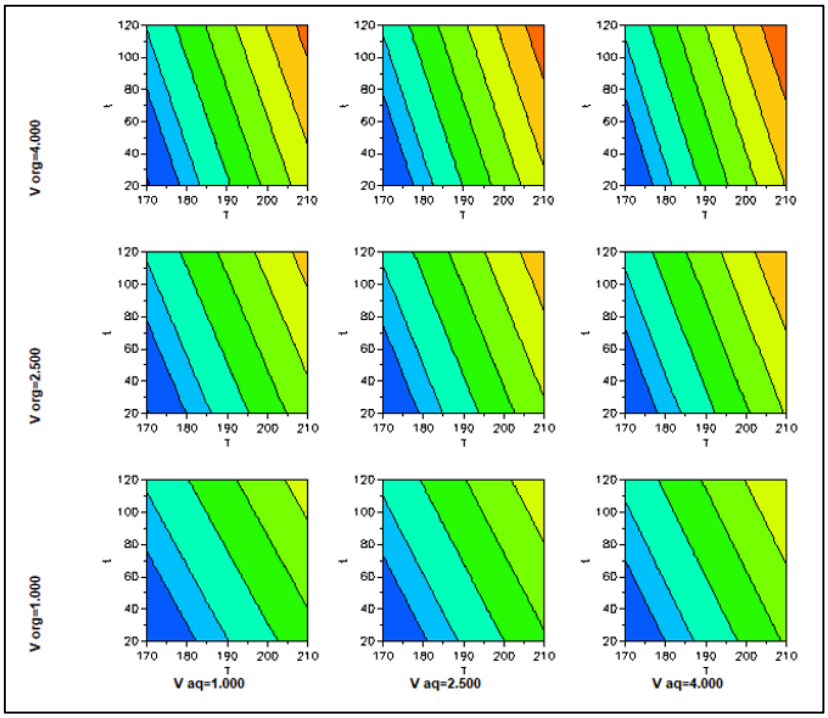


Figure 6-30. Contour plots showing the response surface of FUR yield within the experimental domain.

The contour plots presented in figure 6-30 consists of a series of near-parallel lines. The response surface of FUR yield is therefore not curved, and the regression model doesn't need any quadratic terms. Thus, we have a linear regression model without quadratic terms to describe and predict FUR yield. The mathematical expression for the linear regression model (Equation 6-2) is given below.

$$Y\ (m\%\ FUR) = -6.750 + 0.822T + 0.031t - 0.098V_{aq} + 0.052V_{org} + 0.153Txt + 0.033TxV_{aq} + 0.233TxV_{org} \quad \text{(Equation 6-2)}$$

Where T is temperature, t is residence time, V_{aq} is volume of aqueous solvent, V_{org} is volume of organic solvent, Txt is the two-factor interaction between temperature and time, and TxV_{aq} and TxV_{org} are the two-factor interactions between temperature and volume of solvent (both aqueous and organic).

The responses of FUR and HMF yield are described by two different models, in consequence we need to choose one of the responses to optimize the next design for. As mentioned earlier, the yields of HMF are substantially higher than the yields of FUR. This suggests that the reaction system and/or substrate is more suitable for HMF production. Therefore, future optimization of the FFD will be focused on HMF yield.

The two experiments with the highest HMF yields had residence times of 30 and 75 minutes and temperatures of 210°C and 190°C, respectively (table 6-8). It appears that the optimal reaction conditions for HMF production is an intermediate residence time with a higher temperature. It seems that with longer residence times, HMF undergoes further reactions and/or degradation. Considering the theory of mechanistic pathways to HMF given in section 2.5.1, it can be speculated that with higher residence times, HMF degrades into levulinic acid and formic acid (figure 2-9).

The two main factors volume of aqueous solvent and volume of organic solvent proved to not be very important when modelling HMF yield. Their contributions to the model are opposite directed (Equation 6-1), which corresponds to the negative correlation between them. They can therefore be combined to a ratio and held constant, in order to reduce the number of experiments needed in the optimizing FFD.

6.2.2 Results from the second experimental design (qNMR)

Based on the results from the first fractional factorial design (FFD) the reaction system was optimized with the use of a 2^2 full factorial design. The two main factors investigated was temperature (T) and residence time (t).

The high and low levels of the factors were selected based on the results from the screening FFD. Centre point levels of the factors were regrettably not included in this design due to lack of time. The chosen high and low levels of the main factors are displayed in table 6-9. The layout of the experimental design is presented in table 6-10, the experiments were randomized before execution. The complete layout of the experimental design, including constant parameters, is presented in Appendix D (table D-1) The results from the optimizing design, quantified with qNMR, are presented in table 6-11. The concentrations of FUR and HMF in the aqueous and organic product solutions are given in Appendix D (table D-2).

Table 6-9. The chosen high and low levels of the two main factors investigated in the optimizing full factorial design.

Level:	**T (°C):**	**t (min):**
-	190	30
+	230	75

Table 6-10. The layout of the experimental design for the optimizing factorial design.

Experiment #:	**T (°C):**	**t (min):**
2.1	190	30
2.2	230	30
2.3	190	75
2.4	230	75

Table 6-11. The calculated mass percent (m%) yield of FUR and HMF for the optimizing factorial design, quantified with qNMR.

Experiment #:	FUR yield (m%):	HMF yield (m%):
2.1	0.16	8.93
2.2	0.86	20.04
2.3	0.50	17.21
2.4	0.86	6.23

Figure 6-31 displays pictures of the four product solutions from the optimizing factorial design, before phase separation. Figure 6-32 and 6-33 presents the NMR spectra of the aqueous product solution and the organic product solution of experiment 2.2, respectively. All the NMR spectra obtained from the optimizing full factorial design are presented in Appendix D figure D-2 to figure D-9.

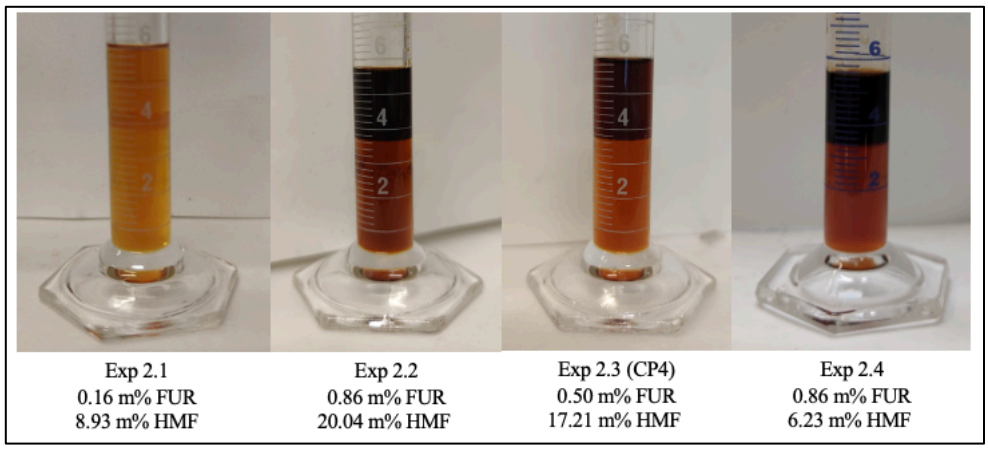


Figure 6.31. Pictures of the four product solutions from the optimizing factorial design, before phase separation.

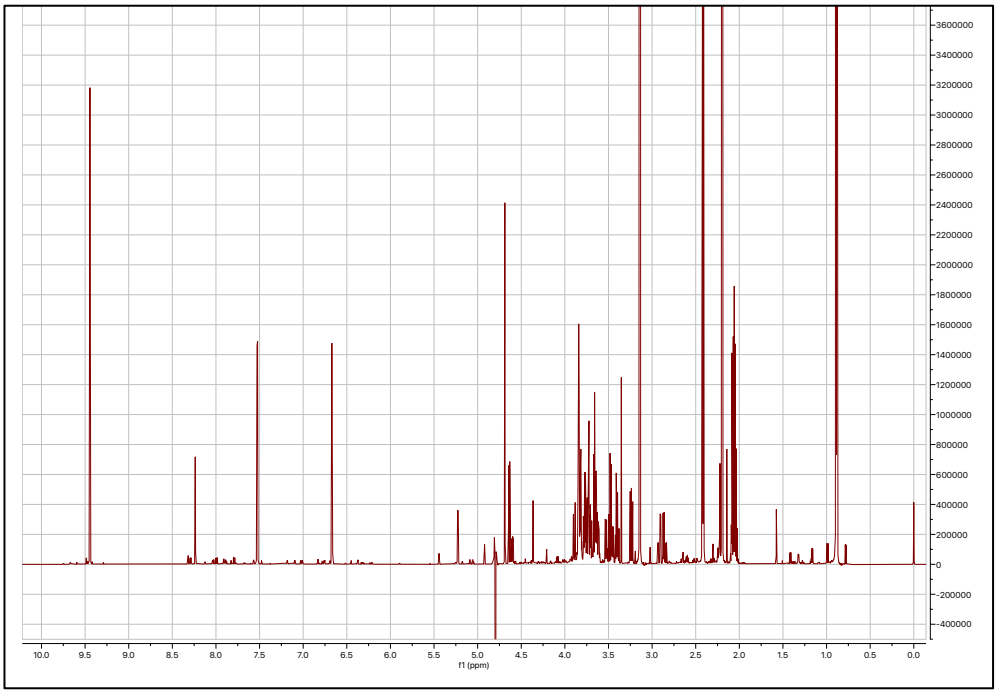


Figure 6-32. The NMR spectrum of the aqueous product solution of experiment 2.2.

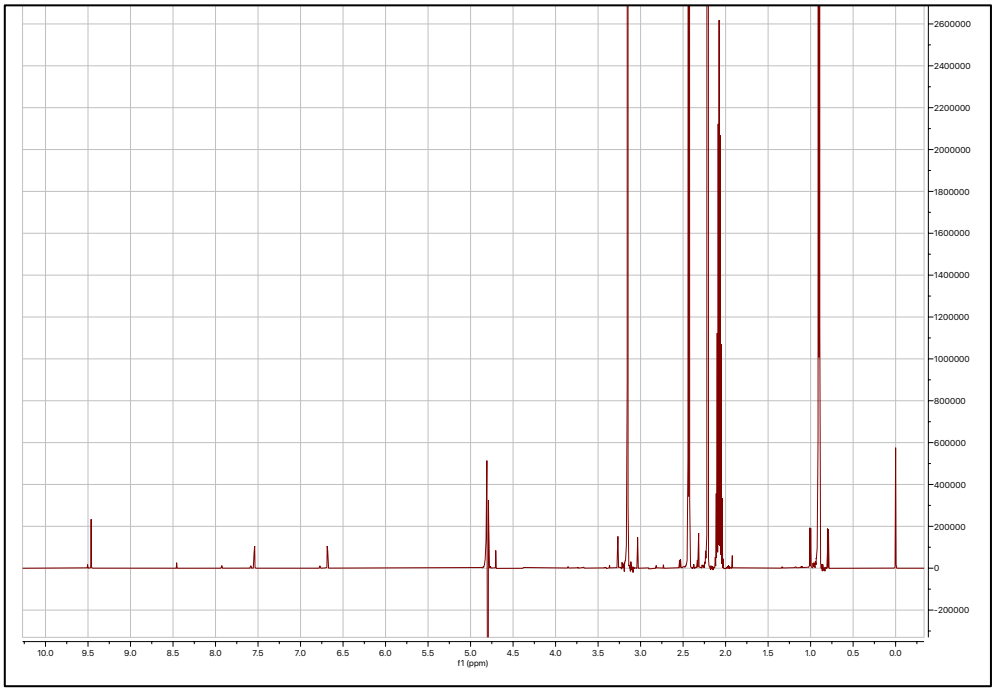


Figure 6-33. The NMR spectrum of the organic product solution of experiment 2.2.

Figure 6-34 displays the biplot obtained from PCA analysis of the optimizing factorial design. The dataset is standardised, and the analysis employs mean centring.

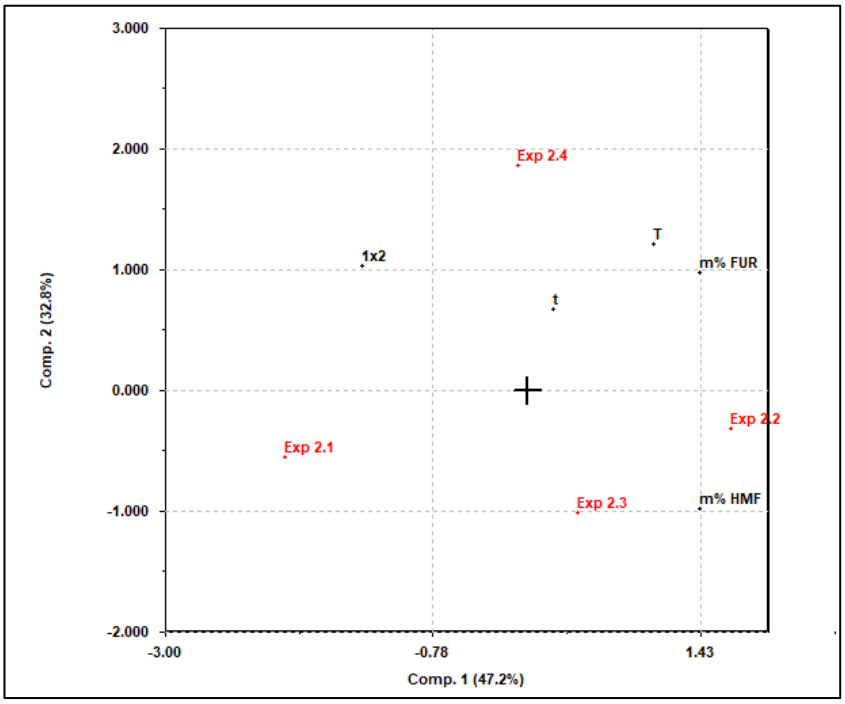


Figure 6-34. The biplot obtained from PCA analysis of the optimizing factorial design. The experiments are marked in red, and the main factors and responses are marked in black.

The biplot in figure 6-34 shows that both FUR and HMF yield are positively correlated with temperature. However, FUR is more correlated with temperature than HMF. Residence time is positively correlated with FUR, but negatively correlated with HMF (the angle between HMF and t is >90°). The two-factor interaction 1x2 is negatively correlated with both HMF and FUR yield. According to the biplot, experiment 2.2 gives the highest yield of HMF, and experiment 2.3 is a close second. This fits with the results given in table 6-11.

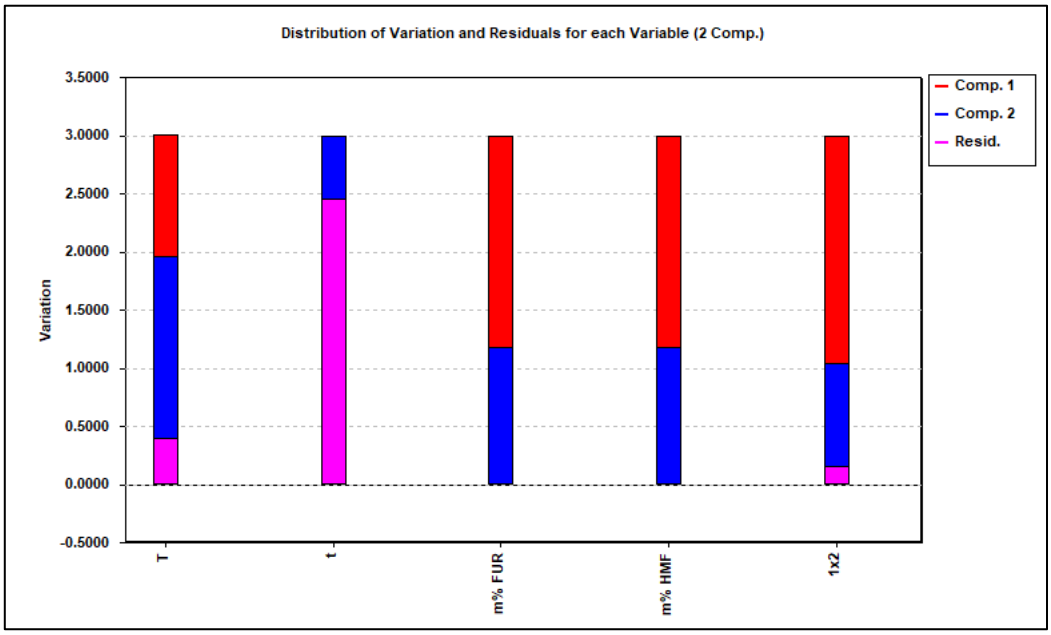


Figure 6-35. Illustrates the distribution of variation and residuals for each variable on each of the two principal components.

The distribution of variation and residuals for each variable on each principal component (PC) is illustrated in figure 6-35. The main factor residence time contributes the least to the model, this means that a small share of the variation in the variable is explained by the PCs. Temperature and the two-factor interaction between temperature and time are the most important factors to the model. The variation described by each PC for FUR and HMF yield are quite similar in this model. However, the correlations between FUR and HMF and other factors in the biplot were quite different. For this reason, separate models will be made for each of the responses.

6.2.2.1 Regression model HMF

The purpose is to find a mathematical expression that can estimate the yield of HMF when using the selected biphasic reaction system. The results from the PLS analysis are presented in figure 6-36 to figure 6-42 below.

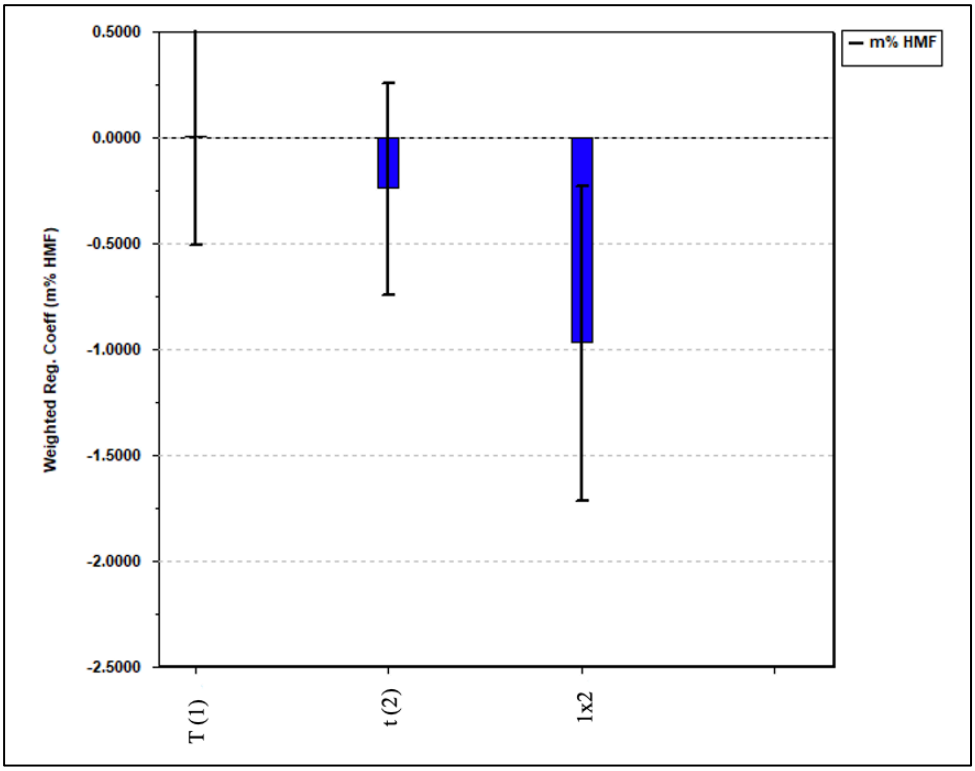


Figure 6-36. The weighted regression coefficients for the main factors and the two-factor interaction. The number behind each main factor corresponds to the number in the two-factor interaction.

Figure 6-36 presents the weighted regression coefficients for the obtained linear regression model. According to the figure, the two-factor interaction between temperature and time is the most important factor concerning HMF yield. The interaction between the main factors is more important than the main factors themselves. The importance of the two-factor interaction is further investigated in figure 6-37 with the use of an interaction graph.

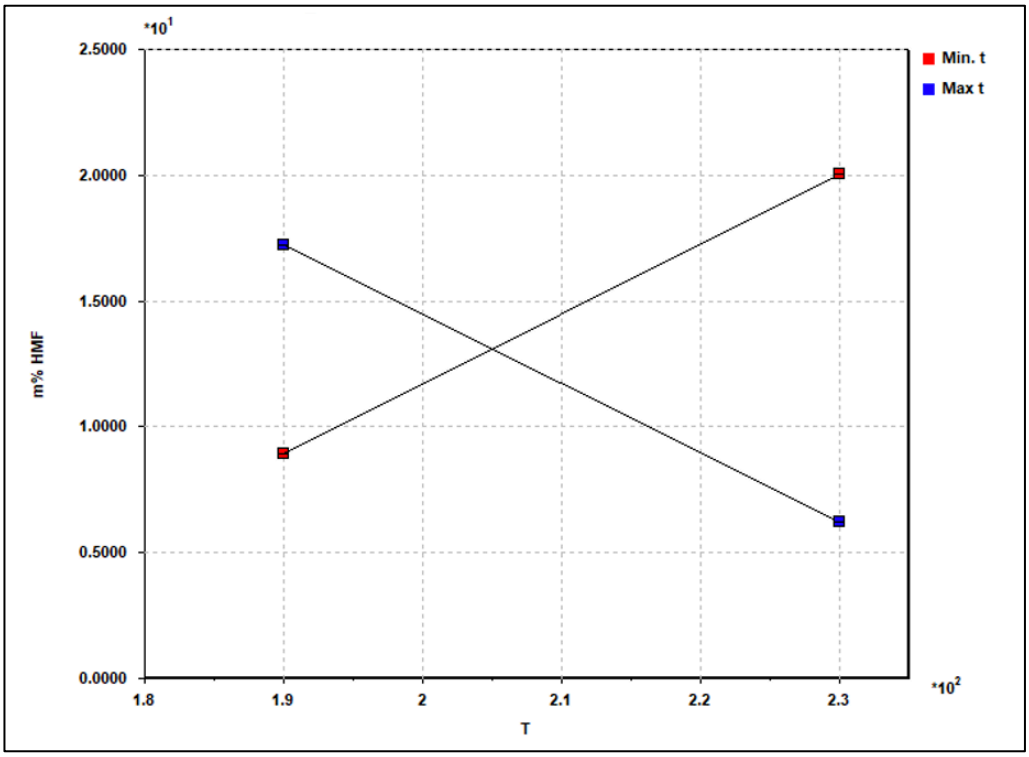


Figure 6-37. Interaction graph displaying the effect of minimum and maximum values of temperature and time on HMF yield.

The two lines in figure 6-37 are far from parallel. This confirms that the two-factor interaction between temperature and time has a significant effect on the HMF yield. According to the figure, increasing the temperature from 190° to 230° for experiments conducted with a 75 minute residence time, has a substantial negative effect on HMF yield. This can be seen by the steep downwards slope of the line between the two blue boxes. However, increasing the temperature from 190° to 230° for experiments with a 30 minute residence time, has a substantial positive effect on HMF yield. This can be seen by the steep upwards slope of the line between the two red boxes.

Figure 6-38 displays a plot of predicted vs. measured HMF yield. The model seems to describe the system very well with a correlation coefficient R^2 of 1. There are regrettably no centre point experiments in this design that can give any indication of a curved response surface for HMF yield. A contour plot is therefore used to investigate the response surface (figure 6-39).

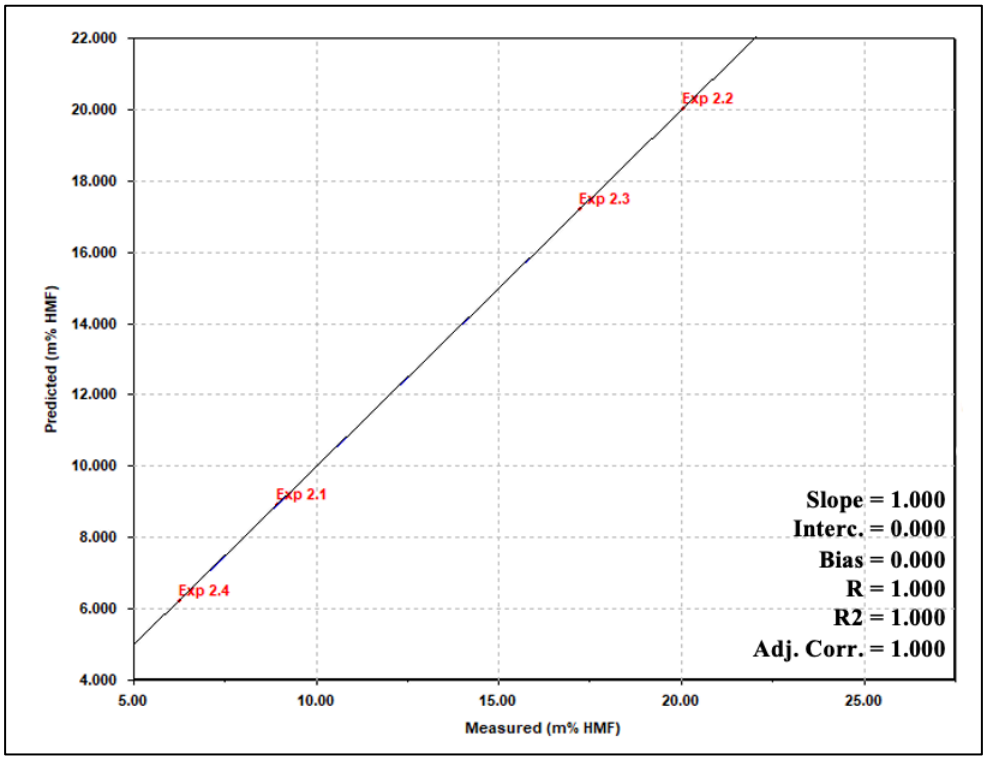


Figure 6-38. Plot of predicted vs. measured values of HMF yield (m%), predicted by the obtained linear regression model.

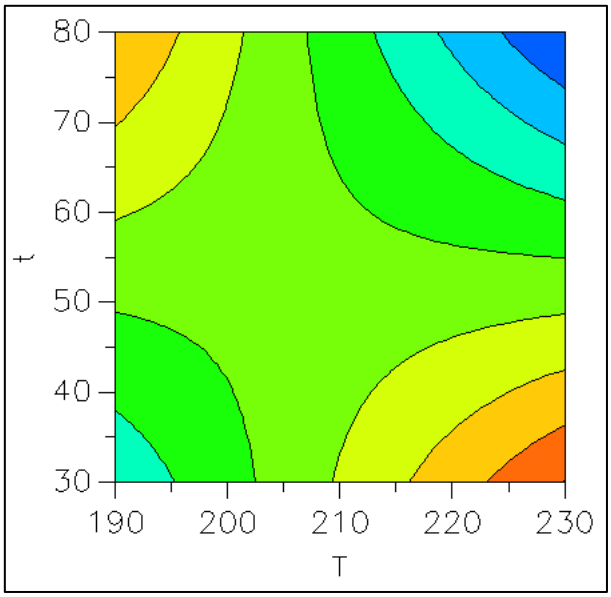


Figure 6-39. Contour plot showing the curved response surface of HMF yield within the experimental domain.

Figure 6-39 displays the contour plot of the response surface of HMF yield. The lines are clearly curved. This confirms that HMF does indeed have a curved response surface in the explored domain, and the model for HMF yield should include quadratic terms.

Figure 6-40 presents the weighted regression coefficients, including quadratic terms. The quadratic terms included is TxT (1x1) and txt (2x2). The figure shows that the quadratic term txt is approximately as important to the model as the main factor residence time and the interaction between temperature and time (Txt). The interaction between temperature and time is still the most important factor. This confirms that quadratic terms need to be included in the model for prediction of HMF yield.

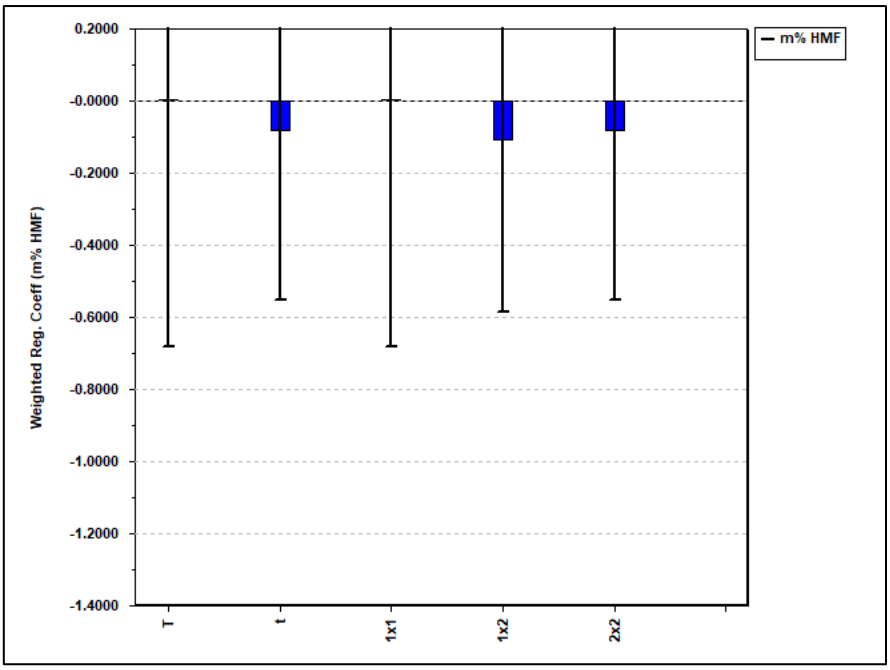


Figure 6-40. The weighted regression coefficients for the main factors, the two-factor interaction, and the quadratic terms.

The plot of predicted vs. measured values of HMF yield obtained by the regression model described by two latent vectors is displayed in figure 6-41. The black line is the fitted regression line. This model describes the system poorly with a regression coefficient R^2 of 0.101. All the other regression models have been described by two latent vectors. However, in this case the model needs three latent vectors to adequately predict HMF yield. Figure 6-42 displays a plot of predicted vs. measured values of HMF yield obtained by the final regression model described by three latent vectors.

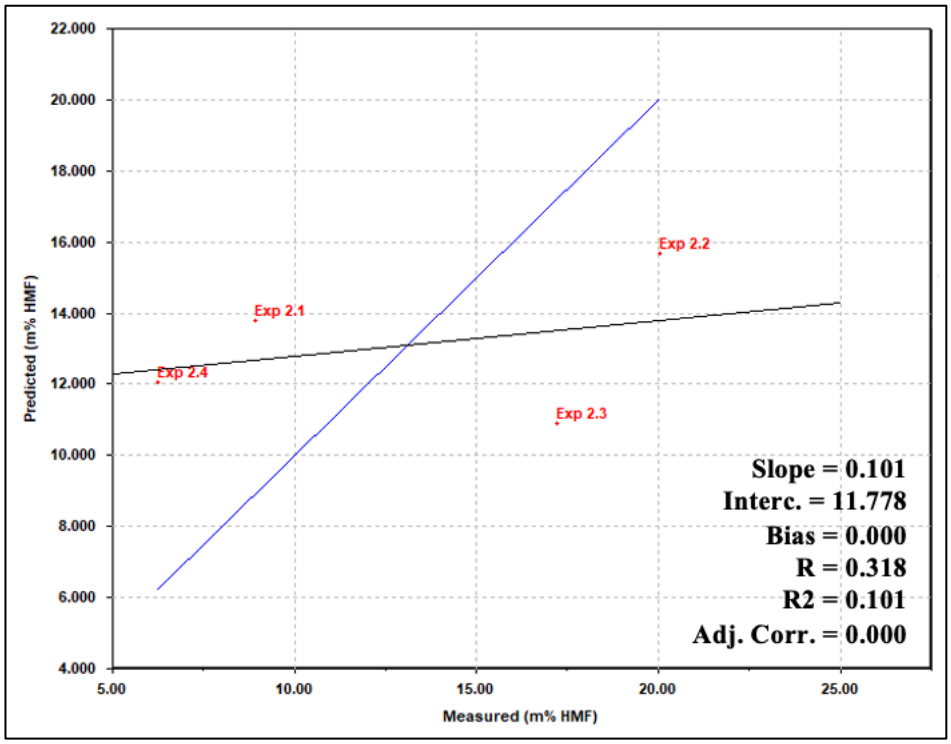


Figure 6-41. Plot of predicted vs. measured values of HMF yield (m%), predicted by the regression model described by two latent vectors.

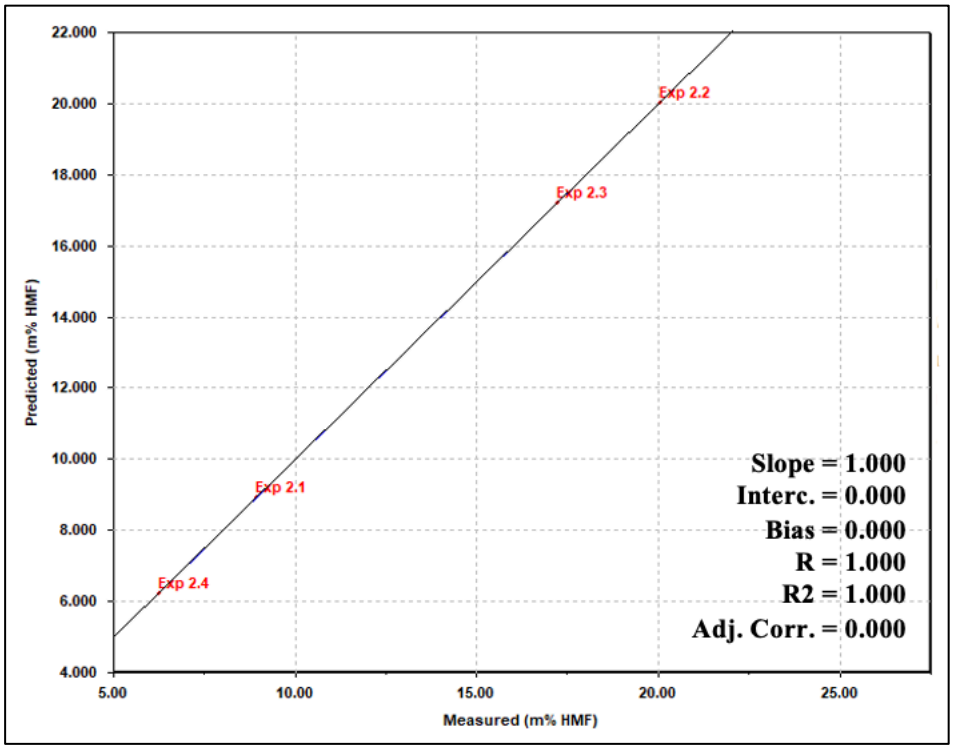


Figure 6-42. Plot of predicted vs. measured values of HMF yield (m%), predicted by the final regression model described by three latent vectors.

The regression model illustrated in figure 6-42 describes the system well. The regression line has a correlation coefficient R^2 of 1.0. There should however be more experiments and centre points to validate the model. The mathematical expression for the regression model (Equation 6-3) is given below.

$$Y\ (m\%\ HMF) = -8.938 + 1.135T + 4.971t + 1.135T^2 - 10.48Txt + 4.971t^2 \quad \text{(Equation 6-3)}$$

Where T is temperature, t is residence time, Txt is the two-factor interaction between temperature and time, and T^2 and t^2 are the quadratic terms for temperature and time, respectively.

6.2.2.2 Regression model FUR

The purpose is to find a mathematical expression that can estimate the yield of FUR when using the selected biphasic reaction system. The results from the PLS analysis are presented in figure 6-43 to figure 6-46 below.

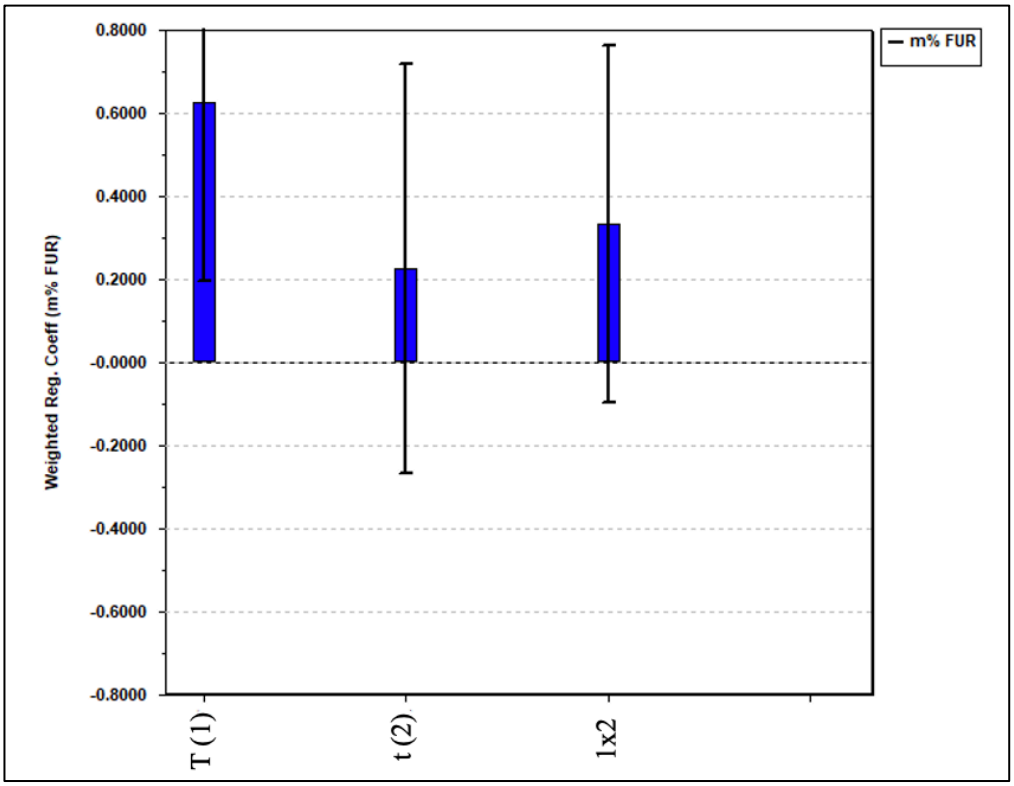


Figure 6-43. The weighted regression coefficients for the main factors and the two-factor interaction. The number behind each main factor corresponds to the number in each two-factor interaction.

The weighted regression coefficients for the obtained linear regression model are presented in figure 6-43. The figure confirms the observation from the biplot (figure 6-34) that temperature is the most important main factor regarding FUR yield. Residence time and the interaction between temperature and time (1x2) are also important to the model. The importance of the two-factor interaction is further investigated in figure 6-44 with the use of an interaction graph.

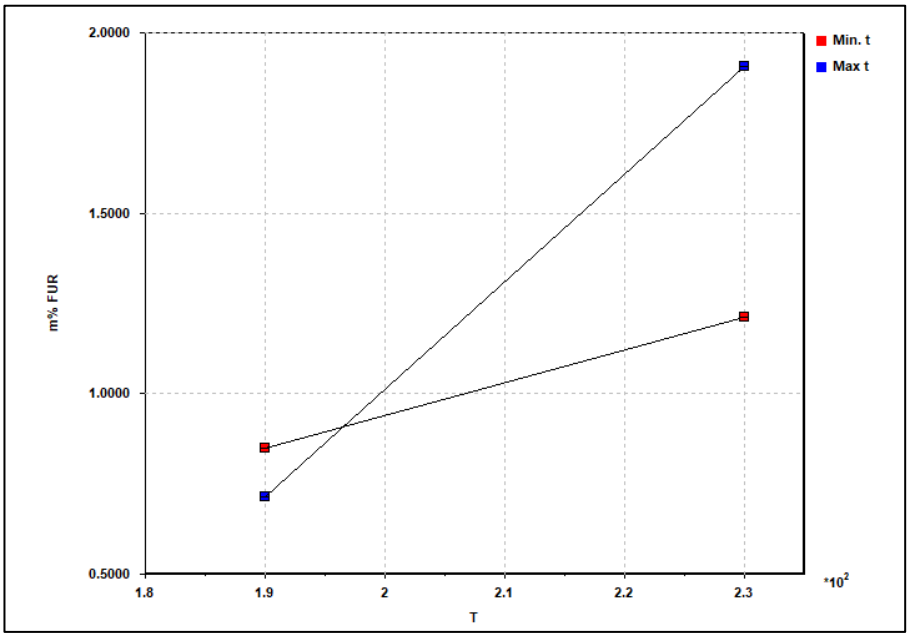


Figure 6-44. Interaction graph displaying the effect of minimum and maximum values of temperature and time on FUR yield.

The two lines in figure 6-44 are not parallel, this confirms that the two-factor interaction between temperature and time has a significant effect on the FUR yield. According to the figure, increasing the temperature from 190° to 230° has a small positive effect on FUR yield for experiments conducted with a 30 minute residence time. This can be seen by the gentle upwards slope of the line between the two red boxes. However, increasing the temperature from 190° to 230° for experiments conducted with a 75 minute residence time, has a large positive effect on FUR yield. This can be seen by the steep upwards slope of the line between the two blue boxes.

A plot of predicted vs. measured values of FUR yield is presented in figure 6-45. The regression model seems to describe the system adequately with a correlation coefficient R^2 of 0.903. There are regrettably no centre point experiments in this design that can give any indication of a curved response surface for FUR yield. A contour plot is therefore used to investigate the response surface (figure 6-46).

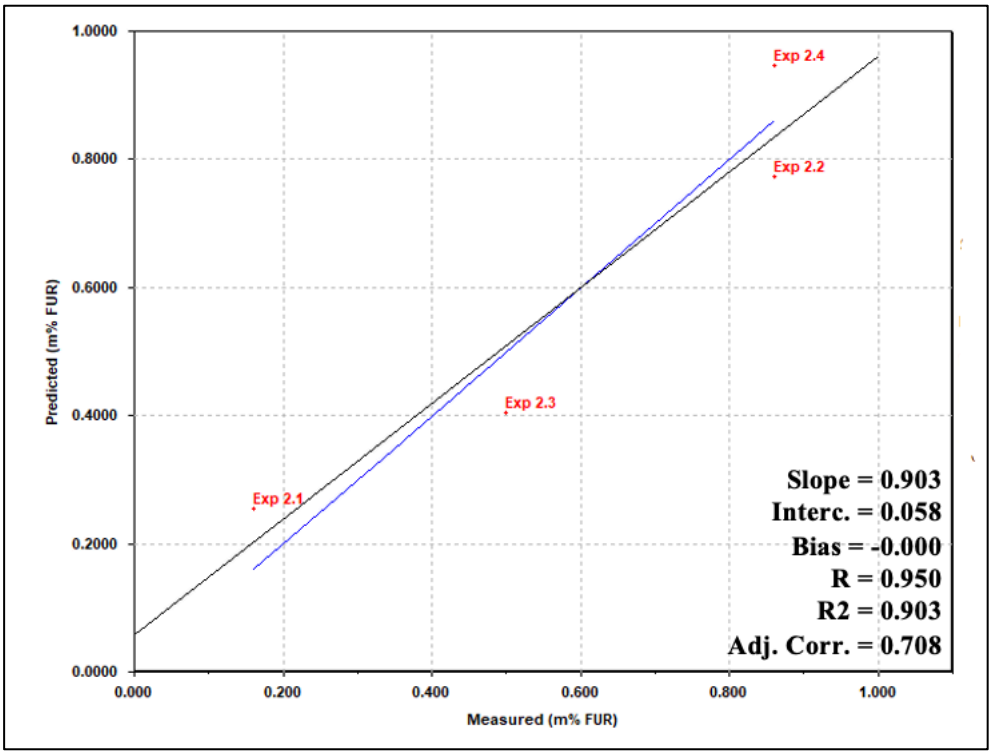


Figure 6-45. Plot of predicted vs. measured values of FUR yield (m%), predicted by the obtained linear regression model.

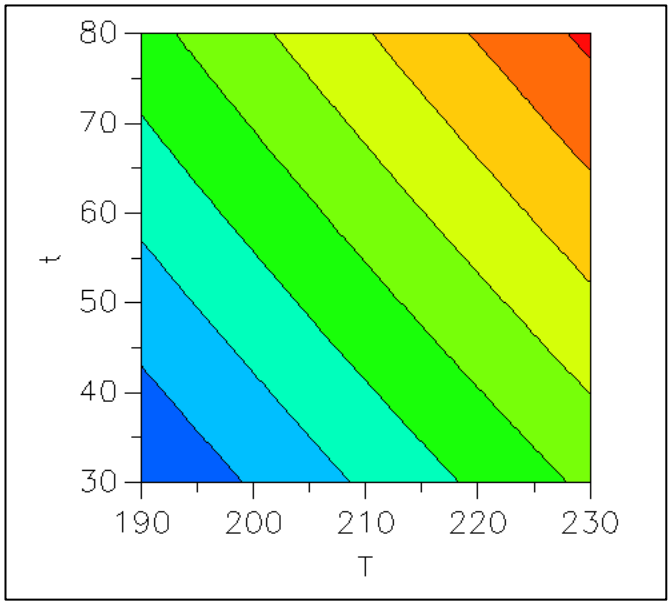


Figure 6-46. A contour plot showing the response surface of FUR yield within the experimental domain.

The lines in the contour plot (figure 6-46) are all near-parallel. This means that the response surface of FUR yield is not curved, and the regression model doesn't need any quadratic terms. Thus, we have a linear regression model without quadratic terms to describe and predict FUR yield. As the correlation coefficient of the model is 0.903, the model doesn't describe the system with complete accuracy. There may be other factors not included in the model that could affect the formation of FUR. The mathematical expression for the regression model (Equation 6-4) is given below.

$$Y\ (m\%\ FUR) = -6.633 + 0.863T + 0.071t + 0.212Txt \qquad \text{(Equation 6-4)}$$

Where T is temperature, t is residence time and Txt is the two-factor interaction between temperature and time.

6.2.3 Results from the optimizing experimental design (HPLC-MS/UV)

As can be seen from the NMR spectra presented in figure 6-14 and 6-32 (the spectra of the aqueous solutions) there is a big area of overlapping signals in the range 3.0-4.0 ppm. The NMR spectrum of cherries and water in figure E-1 (Appendix E) shows that these overlapping signals mostly originate from compounds present in cherries. The overlapping signals make quantification of compounds with signals in this area difficult. This is a weakness of NMR. A wish to investigate and quantify other compounds developed during the reaction, or compounds FUR and HMF may degrade to, is the reason why a HPLC-MS/UV method was developed. The mechanistic pathway to HMF described in figure 2-9 (section 2.5.1) for instance, shows that HMF can degrade into levulinic acid.

The three compounds chosen to investigate (in addition to FUR and HMF) is levulinic acid, 5-methylfurfural, and 2-acetylfuran. The aqueous solutions from the optimizing full factorial design were analysed with the developed HPLC-MS/UV method and then attempted quantified with the prepared calibration curves. The results are presented in table 6-12.

Table 6-12. Presents the results from the quantification of the aqueous product solutions of the optimizing full factorial design by the calibration curves obtained from the developed HPLC-MS/UV method. The quantification of HMF in the aqueous product solutions by qNMR are also presented for comparison. The concentrations are given in millimolar (mM).

Exp #:	cLevulinic acid, MS (mM):	cHMF, UV-Vis (mM):	cHMF, MS (mM):	cHMF, qNMR (mM):
2.1	-	14.95	17.12	18.58
2.2	-	51.11	59.57	60.05
2.3	-	41.07	41.89	47.85
2.4	1.26	14.55	16.73	18.60

Unfortunately, levulinic acid and HMF were the only compounds found to be present in the aqueous solutions from the optimizing FD. If any of the other compounds were present, their concentrations were under the detection limit of the instrument. From the qNMR analysis we

know that FUR is present in these product solutions, but at low concentrations. The concentrations of FUR in the aqueous product solutions are given in table D-2 in Appendix D. Ergo, the concentration of FUR (in the diluted product solutions) was under the detection limit of the instrument.

A limitation of the MS detector is that it can be overloaded by samples of too high concentrations. When analysing the product solutions, the samples needed to be diluted so that the MS-active compound of highest concentration was diluted to an appropriate concentration. The dilutions were first diluted more than needed to be sure that no MS-active compound of higher concentration than HMF was present. After doing this in several dilution steps, it was found that HMF was the MS-active compound of highest concentration in the solutions. Consequently, all the product solutions were diluted so that HMF had the highest concentration possible without compromising the MS detector.

Levulinic acid was found to be present in the aqueous product solution of experiment 2.4. This is also the experiment of lowest HMF yield in the optimizing FD (table 6-11). This experiment was conducted with a residence time of 75 minutes and temperature of 230°C. The interaction graph displayed in figure 6-37 showed that the HMF yield decreased with a high temperature and high residence time. Meanwhile low residence time and high temperature increased HMF yield. It is reasonable to postulate that levulinic acid is one of the compounds HMF degrades to when its yield decreases, this coheres with the degradation pathways for HMF shown in figure 2-9 (section 2.5.1).

During analysis of the product solutions with the developed HPLC-MS/UV method, it became clear that the solutions contained a plethora of unknown compounds. Figure 6-47 and 6-48 presents the signals obtained from analysis of the aqueous product solution of experiment 2.1 with the developed HPLC-MS/UV method. The UV-signal with an approximate retention time of 3.5 minutes is of especial interest as it was present in most of the product solutions and in quite high concentrations, comparable to the HMF concentrations. Its λ_{max} value was roughly 220 nm. Unfortunately, a UV signal on its own doesn't contain enough identifying information to determine the compound present.

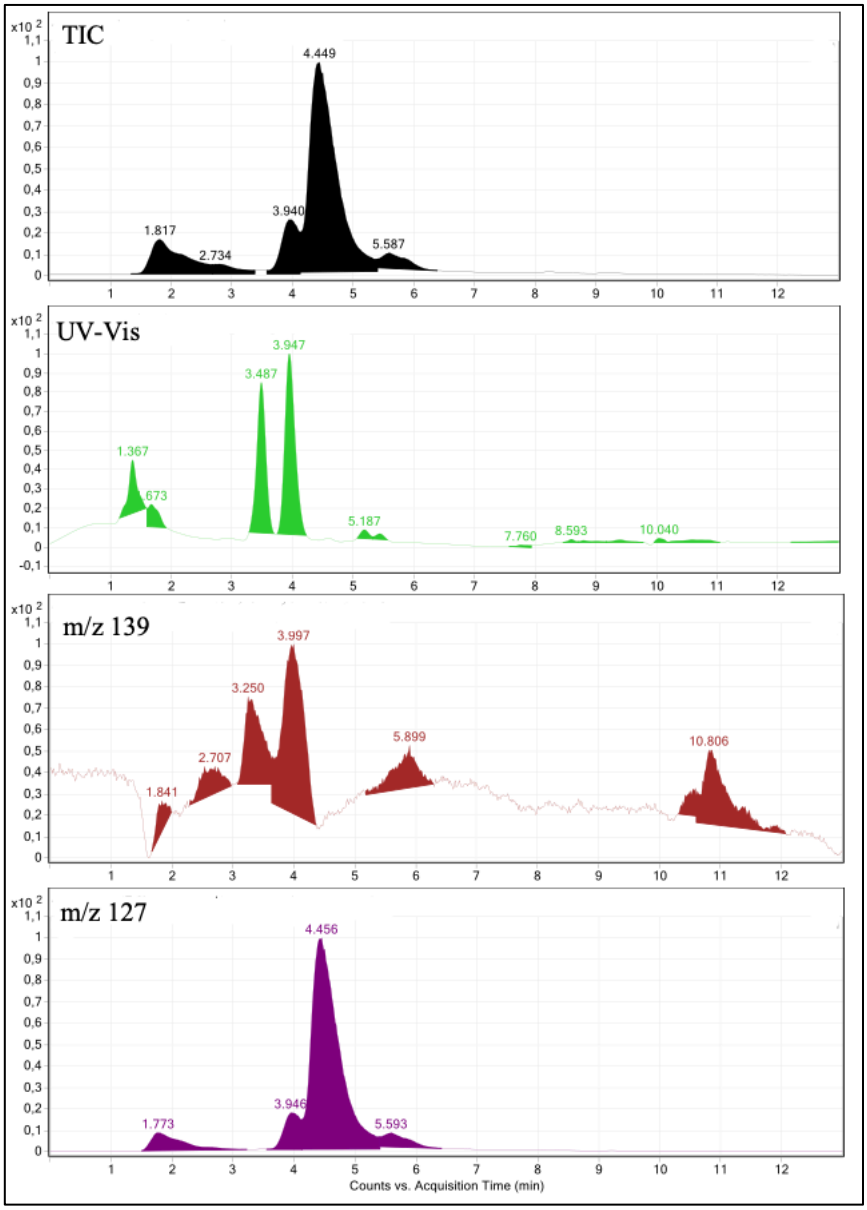


Figure 6-47. The TIC, UV-Vis signals, and SIM signals of m/z values 139 and 127 obtained for the aqueous solution of experiment 2.1 when analysed with the developed HPLC-MS/UV method. The number above each signal is the integrated value of the signal.

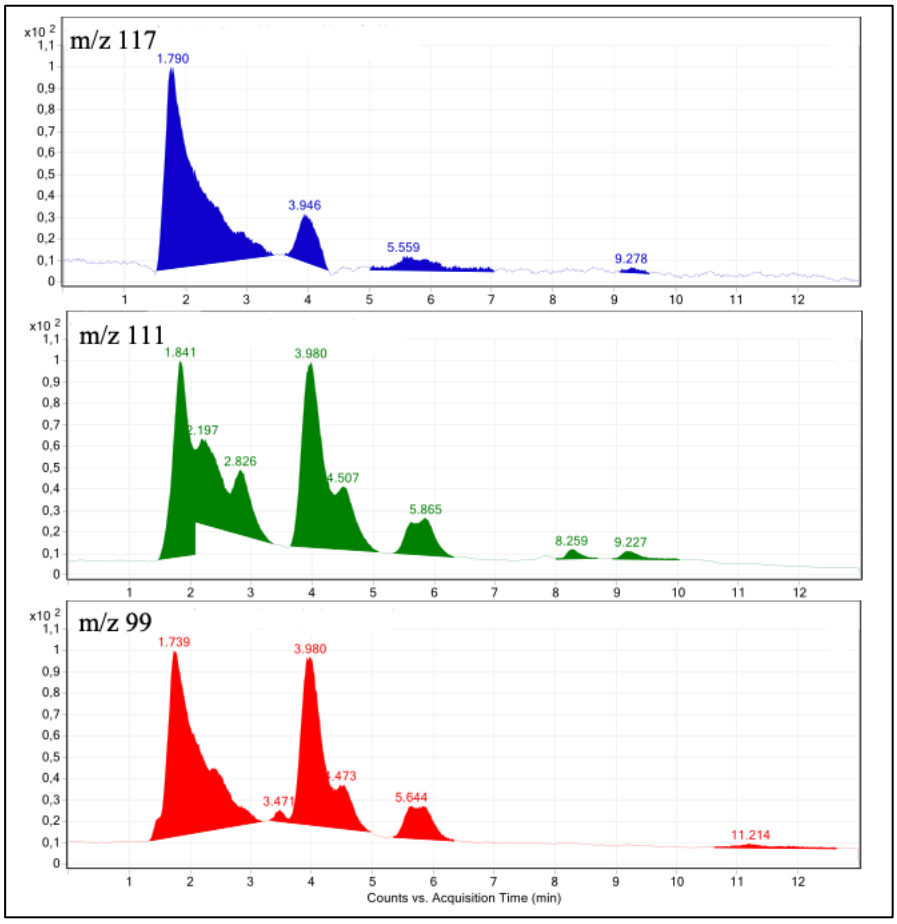


Figure 6-48. The SIM signals of m/z values 117, 111, and 99 obtained for the aqueous solution of experiment 2.1 when analysed with the developed HPLC-MS/UV method. The number above each signal is the integrated value of the signal.

7. Conclusion

The primary objective of this thesis was, as stated in section 3, to optimize a thermochemical conversion reaction of carbohydrates in waste biomass (cherry rejects) into renewable platform chemicals like HMF and FUR using a green biphasic reaction system. The optimization was done to evaluate whether cherry rejects can be used as a feedstock in biorefineries for the large-scale production of FUR and HMF. To accomplish this, the thesis was divided into four main tasks.

In the first task a HPLC-MS/UV method capable of separating and quantifying FUR, HMF, 5-methylfurfural, 2-acetylfuran, and levulinic acid was successfully developed. The mass spectrometry settings for detection are given in table 6-2 and the HPLC settings for the separation are given in table 6-5.

In the second task the screening FFD resulted in relatively high yields of HMF and relatively low yields of FUR. The highest yields in the screening were 19.40 m% and 1.16 m% for HMF and FUR, respectively. Temperature and time were found to be the most significant main factors regarding both HMF and FUR yield. Because the yield of HMF was substantially higher than the FUR yield, the reaction system and substrate seemed to be more suitable for HMF production. Consequently, a choice was made to optimize the next FD based on the yield of HMF only.

The volumes of organic and aqueous solvents did influence HMF yield. The volume of aqueous solvent had a positive effect on HMF yield while the volume of organic solvent had a negative effect. The response surface for HMF yield was curved, hence its regression model had to include quadratic terms. The quadratic term T^2 had a substantial effect on HMF yield, this confirmed that temperature was the most influential variable for achieving high yields of HMF. The two experiments with highest yields were conducted with residence times of 30 and 75 minutes and temperatures of 210°C and 190°C, respectively (table 6-8). These results suggested that an intermediate residence time with a higher temperature would increase the HMF yield. Additionally, it was hypothesised that HMF undergoes further reactions and/or degradation with longer residence times.

In the third task the main factors temperature and time were investigated further. Most importantly, the temperature was increased to a high level of 230°C. The optimizing was successful with a record HMF yield of 20.04 m%, obtained with a residence time of 30 minutes and temperature of 230°C. As the yield is given as mass percent of the start material (containing many other compounds in addition to carbohydrates) this is a satisfactory result. The two-factor interaction between temperature and time was found to be the most influential factor on HMF yield. The interaction graph between temperature and time (figure 6-37) found that HMF yield increased when the temperature was increased for experiments with a residence time of 30 minutes. However, the HMF yield decreased when the temperature was increased for experiments with a residence time of 75 minutes. This observation supports the hypothesis that HMF undergoes further reactions and/or degradation with longer residence times.

It is clear that the main factors temperature and time need to be investigated further in order to find a combination of them that will optimize HMF yield. More research and development of the reaction system needs to be done if cherry rejects are to be used as feedstock in the large-scale production of FUR and HMF. However, the results obtained in this thesis suggest that significant improvements in terms of yield, especially for HMF, can be obtained if the reaction system is optimized. Hence, utilization of biomass waste (in this case cherry rejects) could be a helping hand in the ongoing transition from fossil-derived to renewable energy sources.

The fourth task investigated the presence of three compounds (in addition to FUR and HMF) in the aqueous product solutions obtained from the optimizing factorial design. These compounds were 2-methylfurfural, 2-acetylfuran, and levulinic acid. Unfortunately, levulinic acid was the only compound of the three found in any of the product solutions. Levulinic acid was only found in one of the product solutions, belonging to the experiment with the lowest yield of HMF. This experiment was conducted with a residence time of 75 minutes and temperature of 230°. As mentioned earlier, the interaction graph for temperature and time (figure 6-37) showed that HMF yield decreased with a high temperature and residence time. It is therefore reasonable to postulate that when HMF yield decreases with higher residence times, it is because some of the HMF is degraded to levulinic acid among other compounds. This coheres with the degradation pathways for HMF shown in figure 2-9 (section 2.5.1).

8. Further work

- Perform separate fractional factorial designs for HMF and FUR yield in addition to including more main factors. E.g., Substrate and catalyst amount.

- Conduct a life cycle assessment (LCA) for the reaction system to investigate the impacts of the inputs and outputs. If needed, adjust chemicals/solvent to make the reaction system even greener.

- Test the reaction system with different biomass types as substrate.

- Look for other possible compounds produced during the reaction, or compounds FUR and HMF degrade to, and establish a new HPLC-MS/UV method for these compounds. It would be interesting to find out all the different unknown chemicals present in the product solutions.

9. Bibliography

ABDI, H. 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews, Computational Statistics*, 2, 97-106.

ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews, Computational Statistics*, 2, 433-459.

ACHEAMPONG, M., ERTEM, F. C., KAPPLER, B. & NEUBAUER, P. 2017. In pursuit of Sustainable Development Goal (SDG) number 7: Will biofuels be reliable? *Renewable and Sustainable Energy Reviews,* 75, 927-937.

AHMAD, T., KENNE, L., OLSSON, K. & THEANDER, O. 1995. The formation of 2-furaldehyde and formic acid from pentoses in slightly acidic deuterium oxide studied by ^{1}H NMR spectroscopy. *Carbohydrate Research*, 276, 309-320.

ANASTAS, P. T. & WARNER, J. C. 1998. *Green Chemistry: Theory and Practice,* Oxford, Oxford University Press.

ANTAL JR., M. J., LEESOMBOON, T., MOK, W. S. & RICHARDS, G. N. 1991. Mechanisms of formation of 2-furaldehyde from D-xylose. *Carbohydrate Research*, 217, 71-85.

BP 2022. *Statistical Review of World Energy 2022.* BP: London.

BAJPAI, P. 2016. Pretreatment of Lignocellulosic Biomass for Biofuel Production. Chapter 2: Structure of Lignocellulosic Biomass. Singapore: Springer, p. 7.

CHERUBINI, F. 2010. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management*, 51, 1412-1421.

CHISTI, Y. 2007. Biodiesel from microalgae. *Biotechnology Advances*, 25, 294-306.

DANON, B., MARCOTULLIO, G. & DE JONG, W. 2014. Mechanistic and kinetic aspects of pentose dehydration towards furfural in aqueous media employing homogeneous catalysis, *Green Chemistry*, 16, 39-54.

ESEYIN, A. E. & STEELE P. H. 2015. International Journal of Advanced Chemistry. *An overview of the applications of furfural and its derivatives*, 3, 42-47.

GUNST, R. F. & MASON, R. L. 2009. Fractional factorial design. *Wiley Interdisciplinary Reviews, Computational Statistics*, 1, 234-244.

GUPTA, V.K. & TUOHY, M. G. 2013. *Biofuel technologies: recent developments*, Springer Science and Business Media.

HARRIS, D. C. 2016. *Quantitative Chemical Analysis.* USA: W. H. Freeman and Company.

HUBER, G. W., IBORRA, S., & CORMA, A. 2006. Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering. *Chemical Reviews,* 106, 4044-4098.

HÜLSEY, M. J. 2018. Shell biorefinery: A comprehensive introduction. *Green Energy & Environment*, 3, 318-327.

IEA 2020. *World Energy Outlook*. 2020. OECD publishing: Paris.

IEA 2021. *Renewables 2021.* IEA: Paris.

JEONG, S., LEE, J. 2015 Chapter 5 – Hydrothermal Treatment. *Pretreatment of Biomass*, 61-74.

KUMAR, B. & VERMA, P. 2021. Biomass-based biorefineries: An important architype towards a circular economy. *Fuel*, 288.

LEVITT, M. H. 2008. *Spin Dynamics: basics of nuclear magnetic resonance,* Chichester: Wiley.

LIU, J., XU, G., DONG, W., XU, N., XIN, F., MA, J., FANG, Y., ZHOU, J. & JIANG, M. 2018. Letters in Applied Microbiology. *Biodegradation of diethyl terephthalate and polyethylene terephthalate by a novel identified degrader Delftia sp. WL-3 and its proposed metabolic pathway*, 67, 254-261.

LIU, C., WEI, L., YIN, X., WEI, M., XU, J., JIANG, J. & WANG, K. 2020. Selective conversion of hemicellulose into furfural over low-cost metal salts in a γ-valerolactone/water solution, *Industrial Crops and Products,* 147.

LUNDSTEDT, T., SEIFERT, E., ABRAMO, L., THELIN, B., NYSTROM, Å., PETTERSEN, J. & BERGMAN, R. 1998. Experimental design and optimization. *Chemometrics and Intelligent Laboratory Systems*, 42, 3-40.

LØHRE, C., UNDERHAUG, J., BRUSLETTO, R. & BARTH, T. 2021. A Workup Protocol Combined with Direct Application of Quantitative Nuclear Magnetic Resonance Spectroscopy of Aqueous Samples from Large-Scale Steam Explosion of Biomass. *ACS Omega,* 6, 6714-6721.

MIKA, L. T., CSEFALVAY, E. & NEMETH, A. 2018. Chemical Reviews. *Catalytic Conversion of Carbohydrates to Initial Platform Chemicals: Chemistry and Sustainability*, 118, 505-613.

MILLAN, G. G., HELLSTEN, S., LLORCA, J., LUQUE, R., SIXTA, H. & BALU, A. M. 2019. Recent Advances in the Catalytic Production of Platform Chemicals from Holocellulosic Biomass. *Chemcatchem*, 11, 2022-2042.

MOHRIG, J. R., ALBERG, D. G., HOFMEISTER, G. E., SCHATZ, P. F. & HAMMOND, C. N. 2014. *Laboratory Techniques in Organic Chemistry.* USA: W. H. Freeman and Company.

MOLNES, J. L. 2021. *Fueling the future with sustainable chemistry: Furfural Synthesis from Xylose, Plums and Cherries in a Biphasic Reaction System for Renewable Fuel and Chemicals Production*. Master Thesis in Renewable Energy. Bergen: University of Bergen.

PAVIA, D. L., LAMPMAN, G. M., KRIZ, G. S. & VYVYAN J. A. 2014. *Introduction to Spectroscopy*. Stamford: Cengage Learning.

QIAN, X. 2012. Mechanisms and energetics for brønsted acid-catalyzed glucose condensation, dehydration and isomerization reactions. *Topics in Catalysis*, 55, 218-226.

QIAN, X., NIMLOS, M. R., DAVIS, M., JOHNSON, D. K. & HIMMEL, M. E. 2005. Ab initio molecular dynamics simulations of β-D-xylose degradation mechanisms in acidic aqueous solution. *Carbohydrate Research*, 340, 2319-2327.

RASMUSSEN, H., SØRENSEN, H. R. & MEYER, A. S. 2014. Formation of degradation compounds from lignocellulosic biomass in the biorefinery: sugar reaction mechanisms. *Carbohydrate Research*, 385, 45-57.

SALIS, S., SPANO, N., CIULU, M., FLORIS, I., PILO, M. I. & SANNA, G. 2021. Electrochemical Determination of the "Furanic Index" in Honey. *Molecules*, 26, 4115.

SEVENICH, R., RAUH, C., BELKOVA, B. & HAJSLOVA, J. 2020. Present and Future of High-Pressure Processing. *Chapter 6 – Effect of high-pressure thermal sterilization (HPTS) on the reduction of food processing contaminants (e.g., furan, acrylamide, 3-MCPD-esters, HMF)*, 139-172.

SIGMA-ALDRICH. 2022. Safety data sheet for furfural and 5-hydroxymethylfurfural.

SLUITER, A., HAMES, B., HYMAN, D., PAYNE, C., RUIZ, R., SCARLATA, C., SLUITER, J., TEMPLETON, D. & WOLFE, J. 2008. *Determination of Total Solids in Biomass and Total Dissolved Solids in Liquid Process Samples*. Technical Report NREL/TP-510-42621. Colorado: National Renewable Energy Laboratory for the U.S. Department of Energy.

SUNDARAM, S. 2015. Biorefineries and chemical processes: design, integration, and sustainability analysis. *Green Processing and Synthesis*, 4, 65-66.

WANG, B., SUN, Y. & SUN. R. 2019. Journal of Leather Science and Engineering. *Fractionational and structural characterization of lignin and its modification as biosorbents for efficient removal of chronium from wastewater: a review*, 1.

WERPY, T., PETERSEN, G., ADEN, A., BOZELL, J., HOLLADAY, J., WHITE, J. & MANHEIM, A. 2004. Top Value Added Chemicals From Biomass. Volume I: Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Washington DC: U.S. Department of Energy.

XU, X., FENG, T., ZHANG, J., MA, Y. & SONG, L. 2018. A rapid clean-up method for the quantification of 5-hydroxymethyl-2-furaldehyde in thermally treated abalone (Haliotis discus) muscle by HPLC-MS/MS. *Analytical Methods*, 10, 5091.

XU, C., PAONE, E., RODRIGUEZ-PADRON, D., LUQUE, R. & MAURIELLO, F. 2020. Chemical Society Reviews. *Recent catalytic routes for the preparation and the upgrading of biomass derived furfural and 5-hydroxymethylfurfural*, 49, 4273-4306.

YANG, G., PIDKO, E. A. & HENSEN, E. J. M. 2012. Mechanism of Brønsted acid-catalyzed conversion of carbohydrates. *Journal of Catalysis*, 295, 122-132.

Appendices

Appendix A
Physical and chemical properties of HMF and FUR

Table A-1. Physical and chemical properties of furfural (FUR). Values are taken from Sigma-Aldrich (2022).

Property:	Value:
Chemical formula:	$C_5H_4O_2$
CAS number:	98-01-1
Appearance:	Light brown, clear, viscous liquid
Molar mass (g/mol):	96.08
Density (g/cm3):	1.16 at 25°C
Melting point/freezing point (°C):	-36
Boiling point (°C):	162
Vapor pressure (hPa):	2.3 at 20°C
Solubility in water (g/100mL):	8.3

Table A-2. Physical and chemical properties of 5-hydroxymethylfurfural (HMF). Values are taken from Sigma-Aldrich (2022).

Property:	Value:
Chemical formula:	$C_6H_6O_3$
CAS number:	67-47-0
Appearance:	Beige or yellow solid
Molar mass (g/mol):	126.11
Density (g/cm3):	1.243 at 25°C
Melting point/freezing point (°C):	28-34
Boiling point (°C):	115 at 1 mm Hg
Vapor pressure (mm Hg):	0.00528
Solubility in water (g/mL):	7.0×10^5 at 25°C

Appendix B
Raw data from HPLC-MS/UV method development

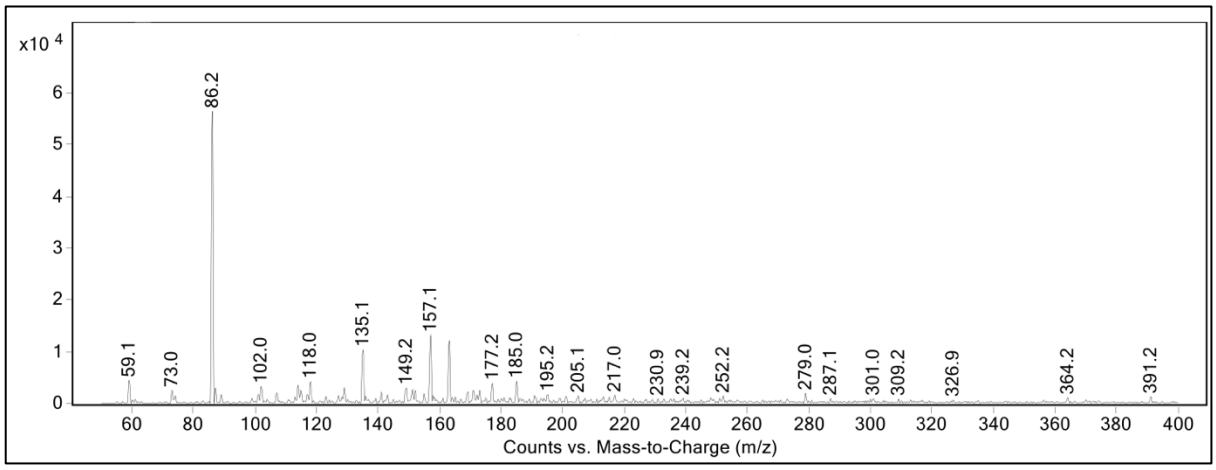


Figure B-1. Displays the MS background spectrum. Obtained with settings listed in table 6-2.

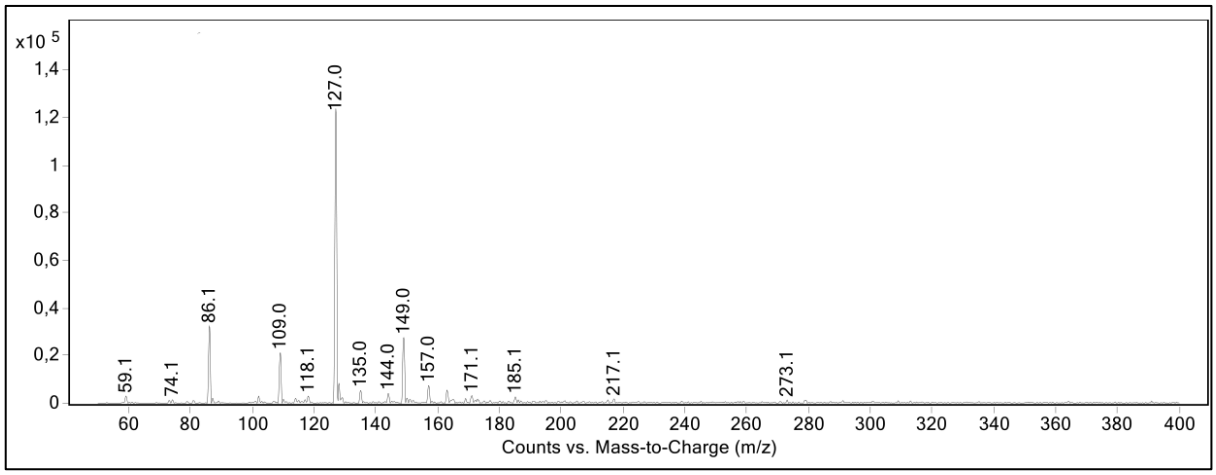


Figure B-2. Displays the MS spectrum of HMF. Obtained with settings listed in table 6-2.

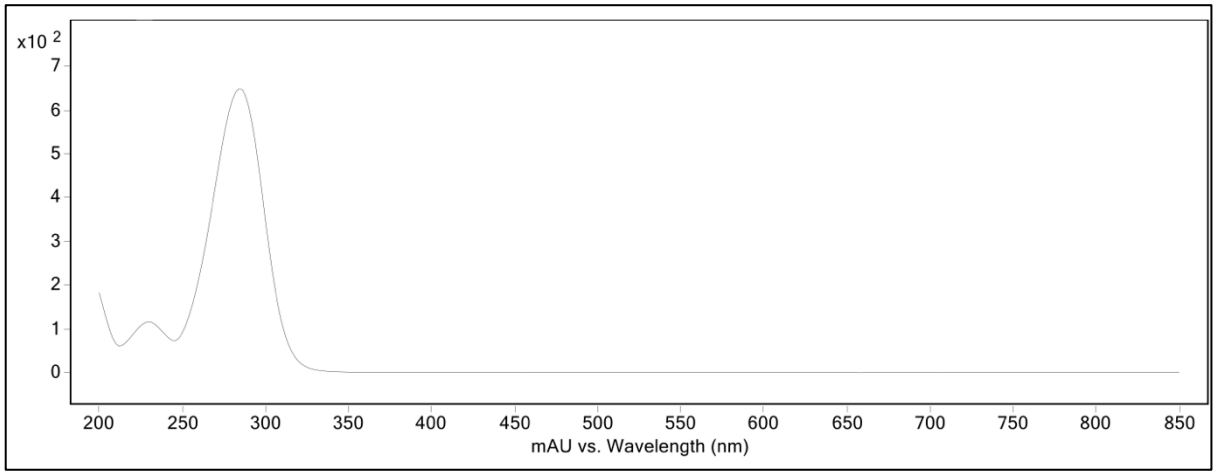


Figure B-3. Displays the UV-Vis spectrum of HMF.

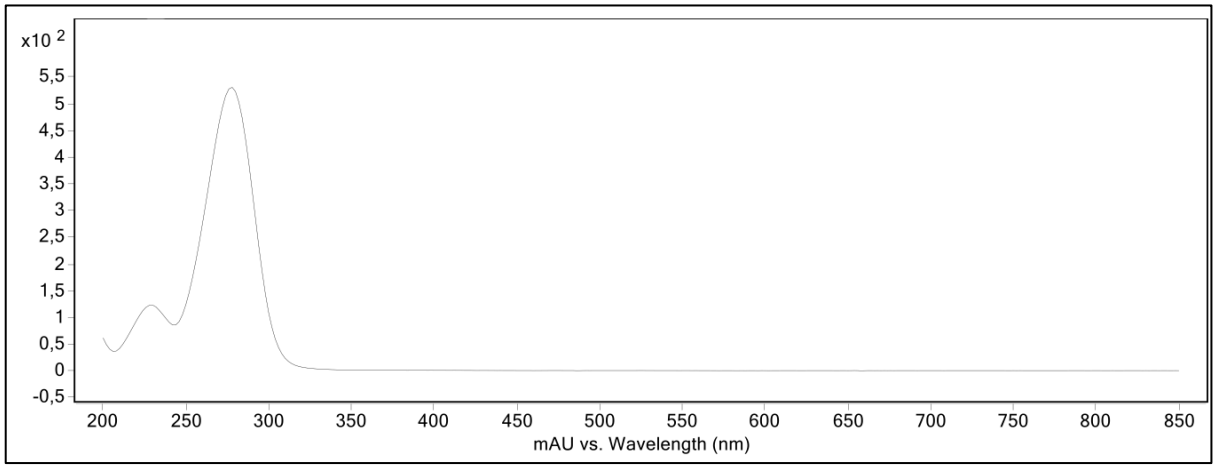


Figure B-4. Displays the UV-Vis spectrum of FUR.

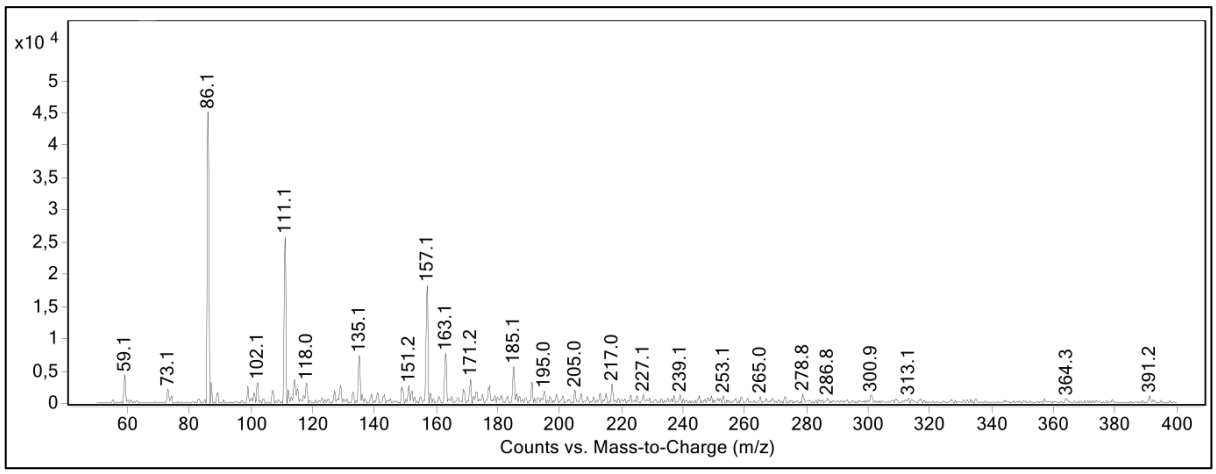


Figure B-5. Displays the MS spectrum of 5-Methylfurfural. Obtained with settings listed in table 6-2.

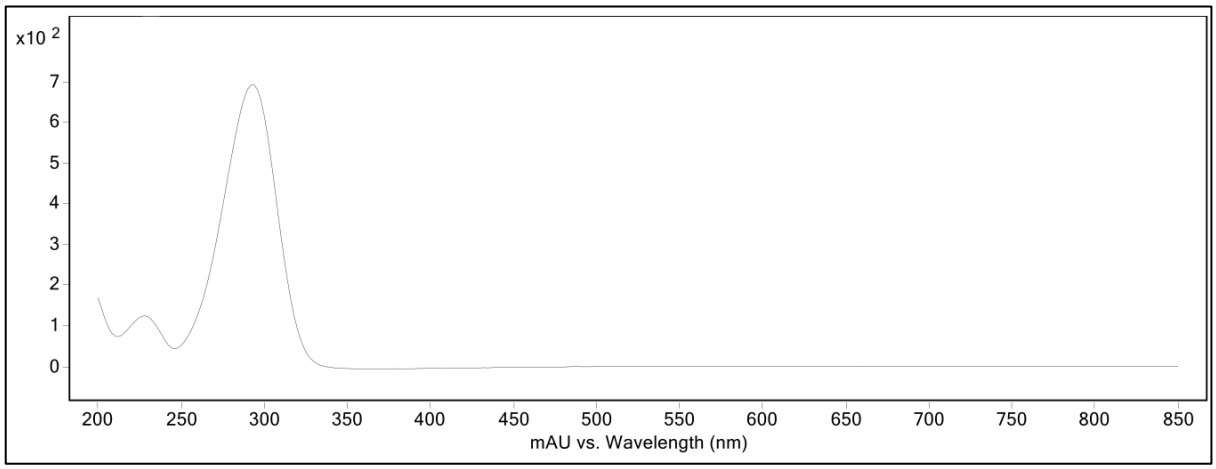


Figure B-6. Displays the UV-Vis spectrum of 5-Methylfurfural.

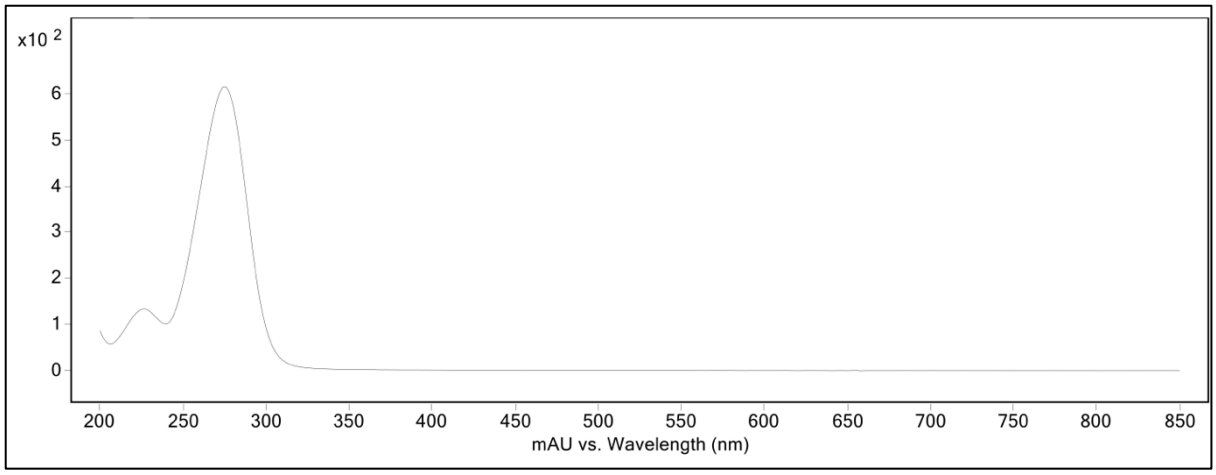


Figure B-7. Displays the UV-Vis spectrum of 2-Acetylfuran.

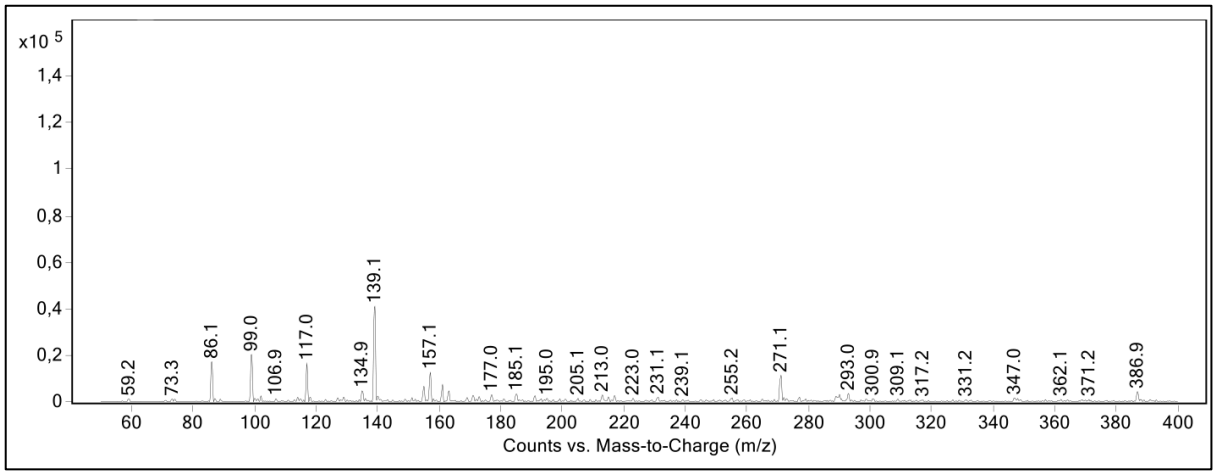


Figure B-8. Displays the MS spectrum of levulinic acid. Obtained with settings listed in table 6-2.

Appendix C
Raw data from the screening fractional factorial design, and additional results

Table C-1. All the variables and constants for the screening fractional factorial design. Substrate wet is the weight of the raw cherries added to the reactor and substate dry is the calculated dry weight.

Exp #:	T (°C):	t (min):	Vaq (mL):	Vorg (mL):	Catalyst (g):	Substrate wet (g):	Substrate dry (g):
1	170	30	4	1	0.025	1.4131	0.2
2	210	30	4	1	0.025	1.4131	0.2
3	170	120	4	1	0.025	1.4131	0.2
4	210	120	4	1	0.025	1.4131	0.2
5	170	30	1	4	0.025	1.4131	0.2
6	210	30	1	4	0.025	1.4131	0.2
7	170	120	1	4	0.025	1.4131	0.2
8	210	120	1	4	0.025	1.4131	0.2
CP1	190	75	2.5	2.5	0.025	1.4131	0.2
CP2	190	75	2.5	2.5	0.025	1.4131	0.2
CP3	190	75	2.5	2.5	0.025	1.4131	0.2
CP4	190	75	2.5	2.5	0.025	1.4131	0.2

Table C-2. The results from the screening fractional factorial design, including the concentrations of FUR and HMF in the aqueous and organic product solutions. “Traces” means the signals were over the detection limit but under the quantification limit.

Exp #:	cFur (aq) (mM):	cHMF (aq) (mM):	cFur (org) (mM):	cHMF (org) (mM):	FUR yield (m%):	HMF yield (m%):
1	0.11	4.07	traces	3.27	0.02	1.23
2	1.96	49.84	29.74	104.15	1.16	18.43
3	0.79	25.15	3.23	25.95	0.26	8.54
4	2.45	47.92	11.26	47.88	0.77	14.85
5	0.28	11.90	traces	12.11	0.03	4.30
6	1.16	53.75	2.64	55.39	0.59	19.40
7	0.40	34.70	1.44	36.94	0.29	12.17
8	1.87	37.58	5.76	39.87	1.12	12.67
CP1	1.97	61.14	9.33	64.71	1.02	19.14
CP2	0.97	47.99	2.89	48.72	0.43	16.63
CP3	1.06	46.44	3.86	48.06	0.56	16.81
CP4	0.93	47.85	3.19	47.95	0.50	17.21

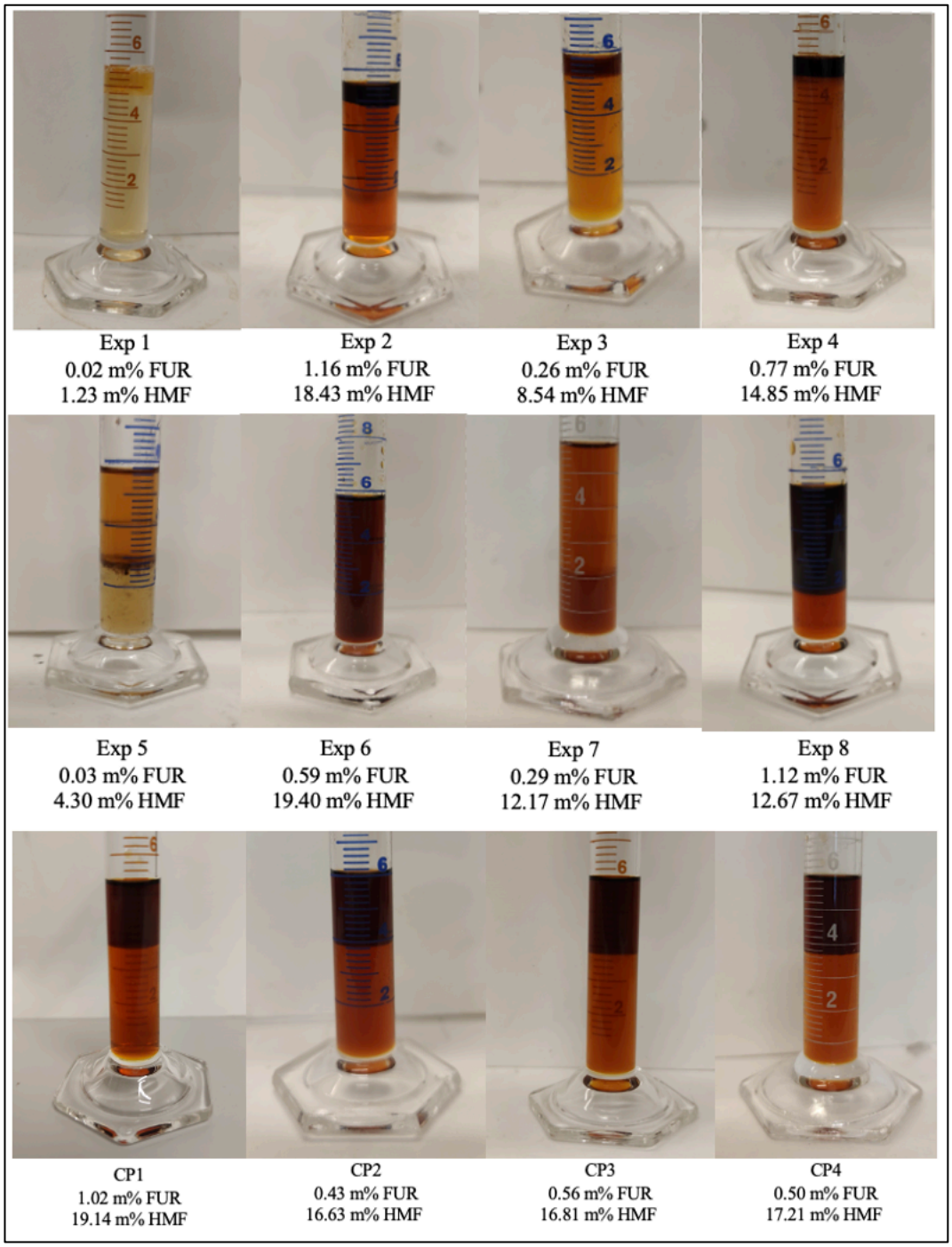


Figure C-1. Pictures of all the product solutions belonging to each experiment in the screening FFD, before phase separation.

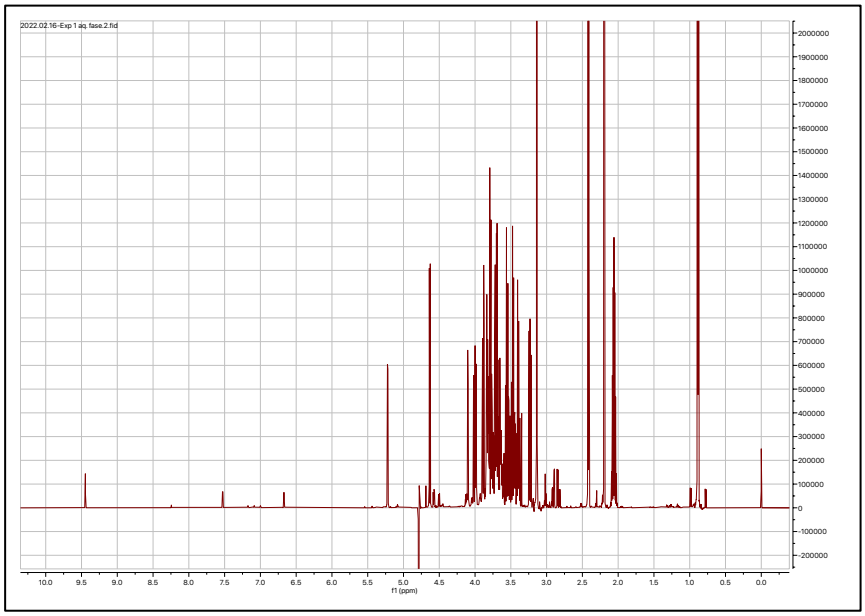


Figure C-2. The NMR spectrum of the aqueous product solution of experiment 1.

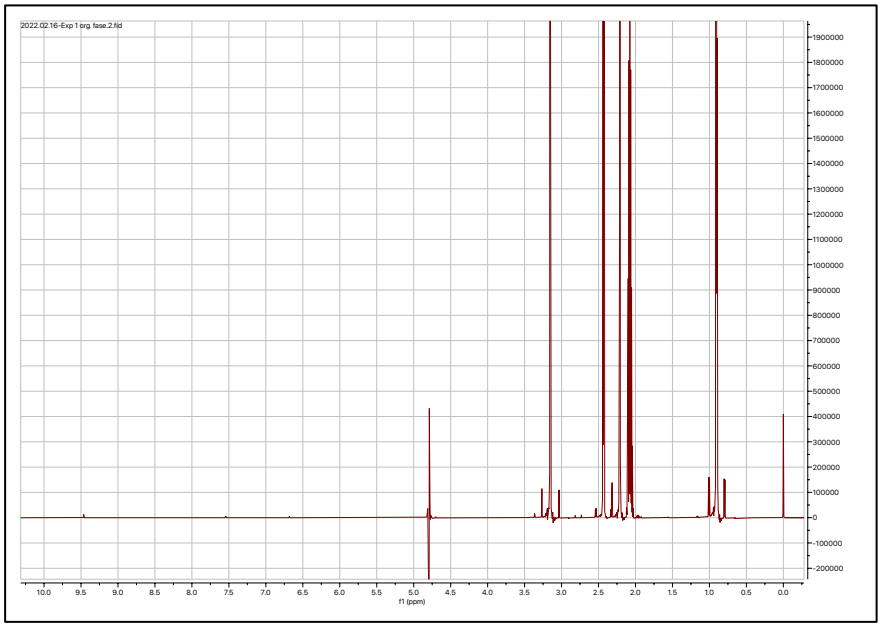


Figure C-3. The NMR spectrum of the organic product solution of experiment 1.

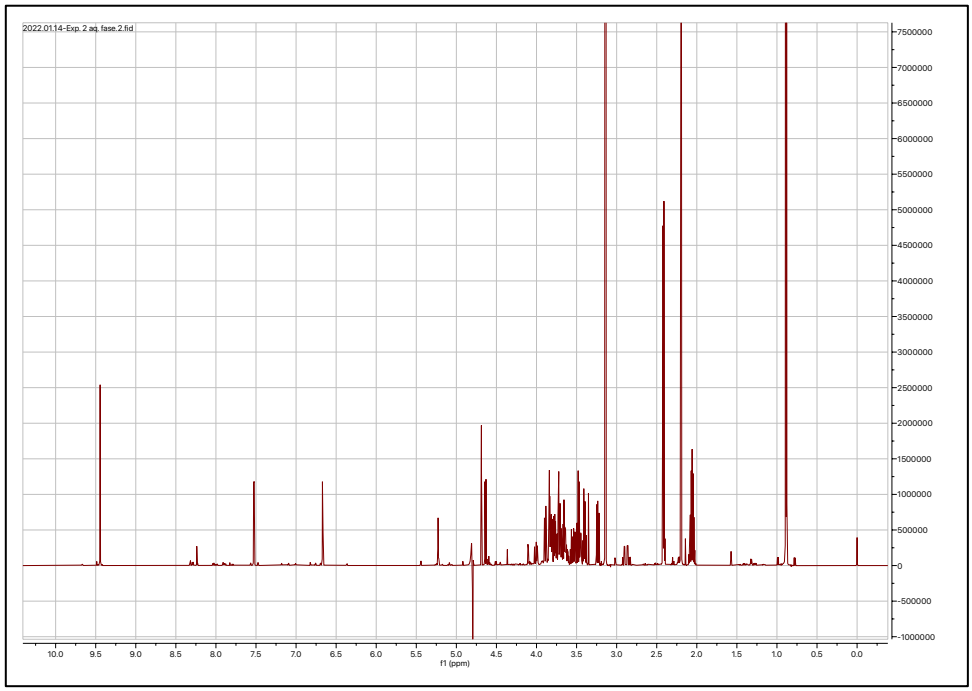


Figure C-4. The NMR spectrum of the aqueous product solution of experiment 2.

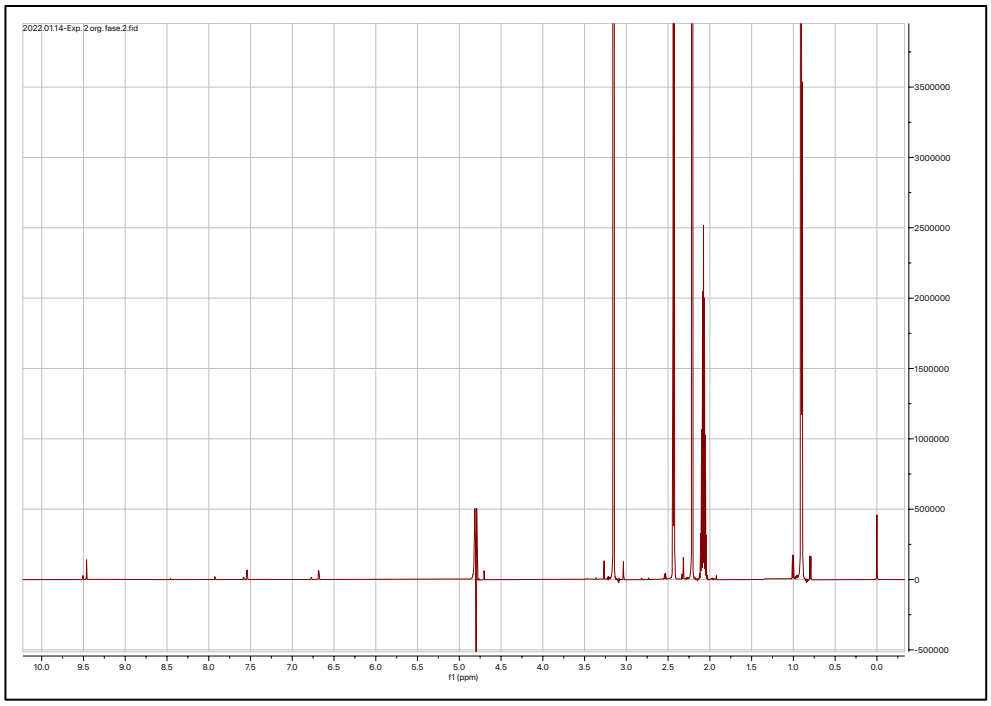


Figure C-5. The NMR spectrum of the organic product solution of experiment 2.

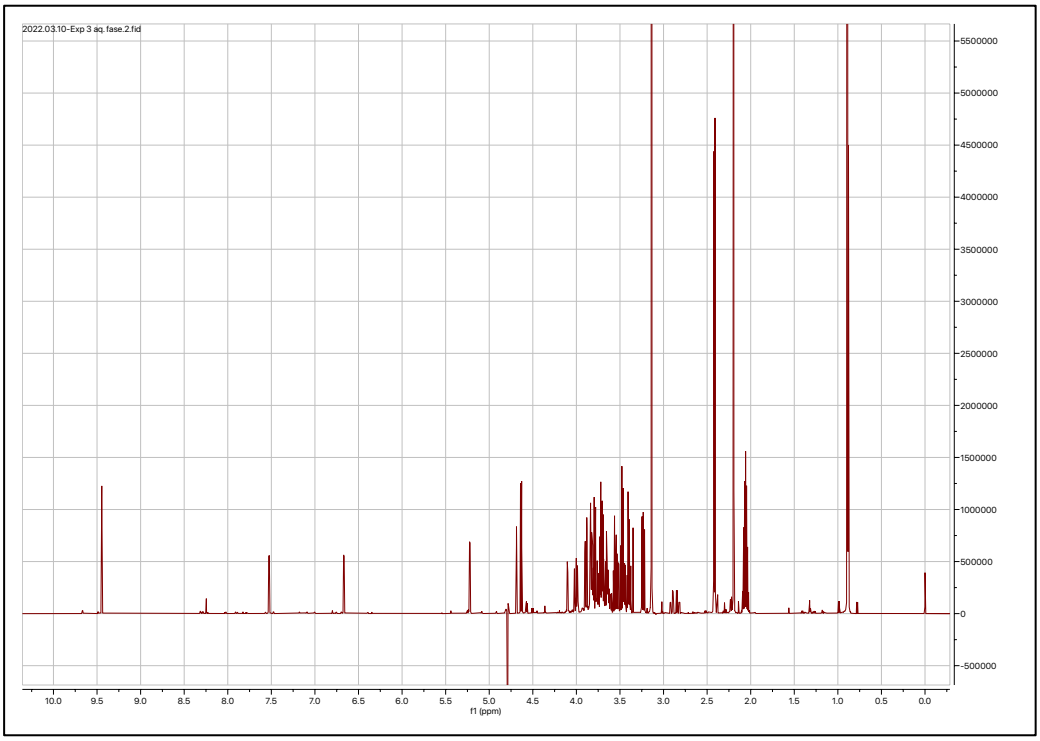


Figure C-6. The NMR spectrum of the aqueous product solution of experiment 3.

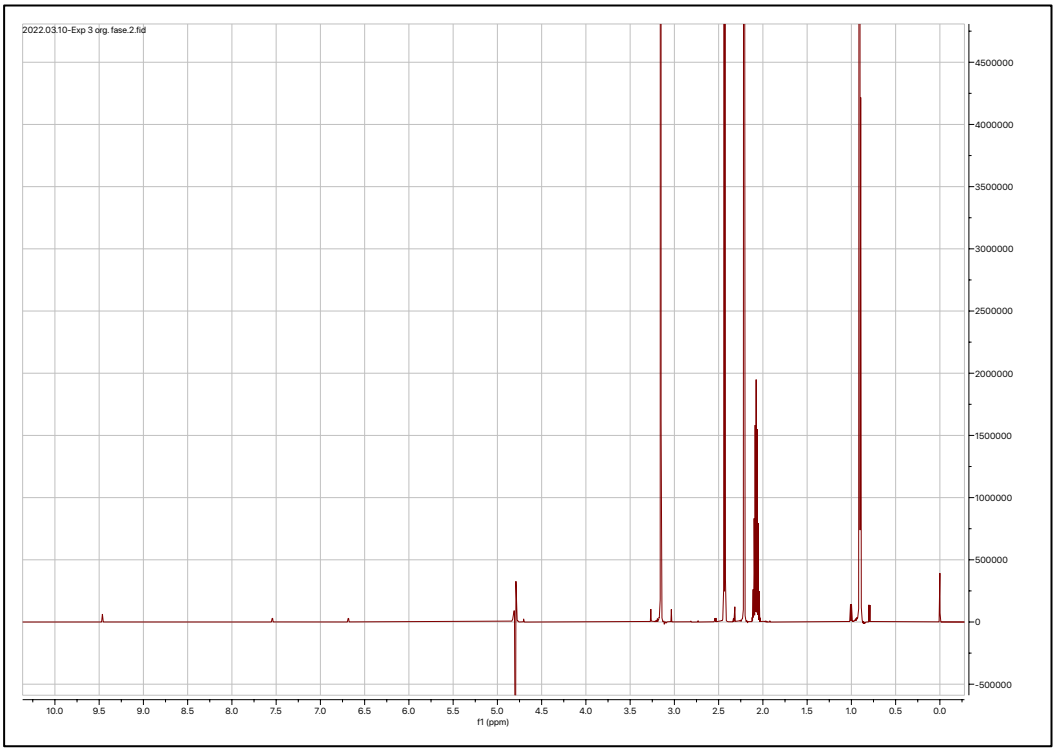


Figure C-7. The NMR spectrum of the organic product solution of experiment 3.

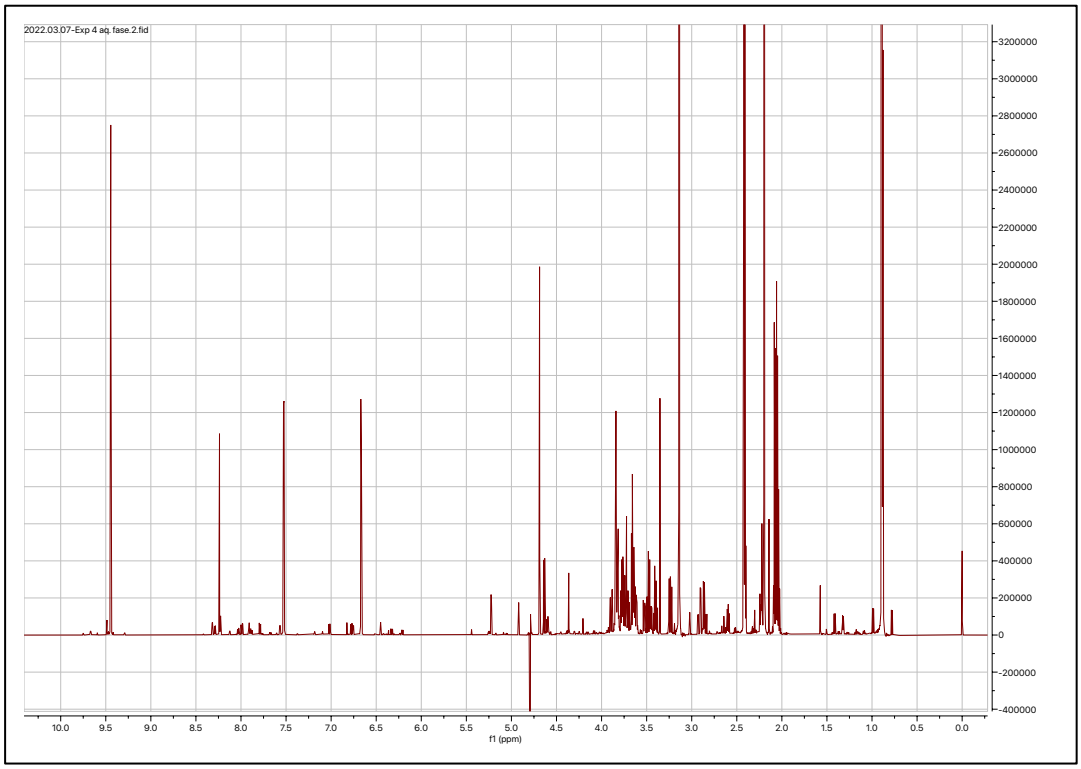


Figure C-8. The NMR spectrum of the aqueous product solution of experiment 4.

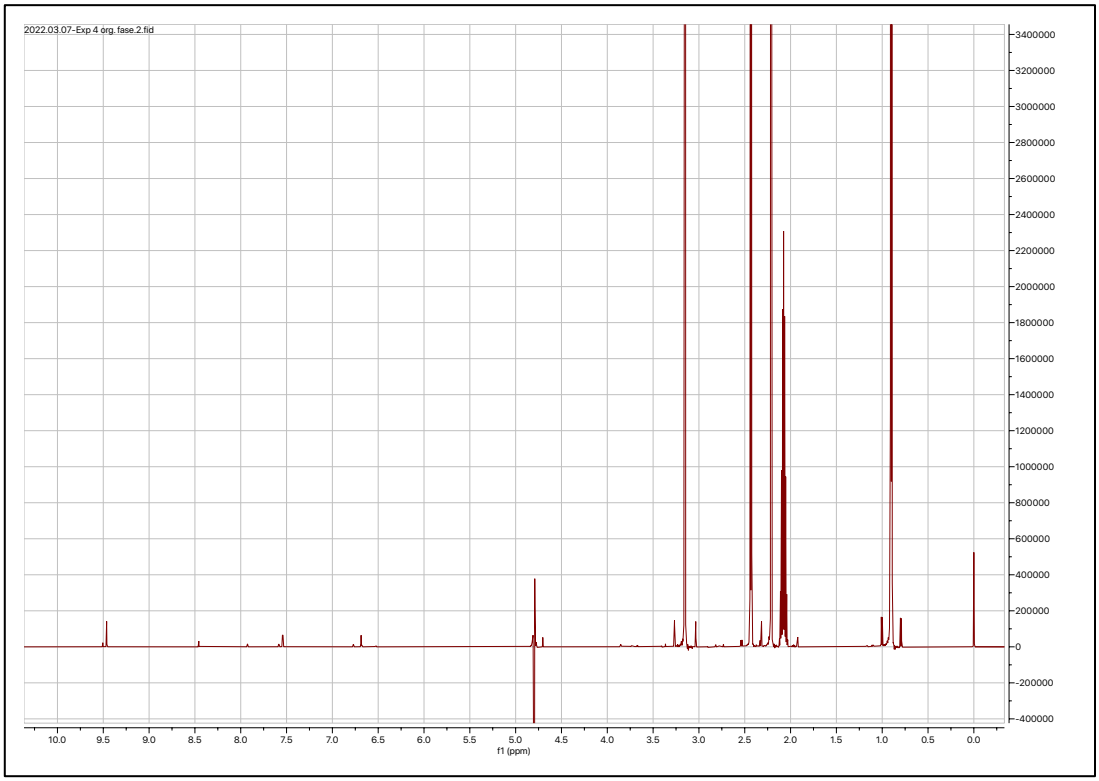


Figure C-9. The NMR spectrum of the organic product solution of experiment 4.

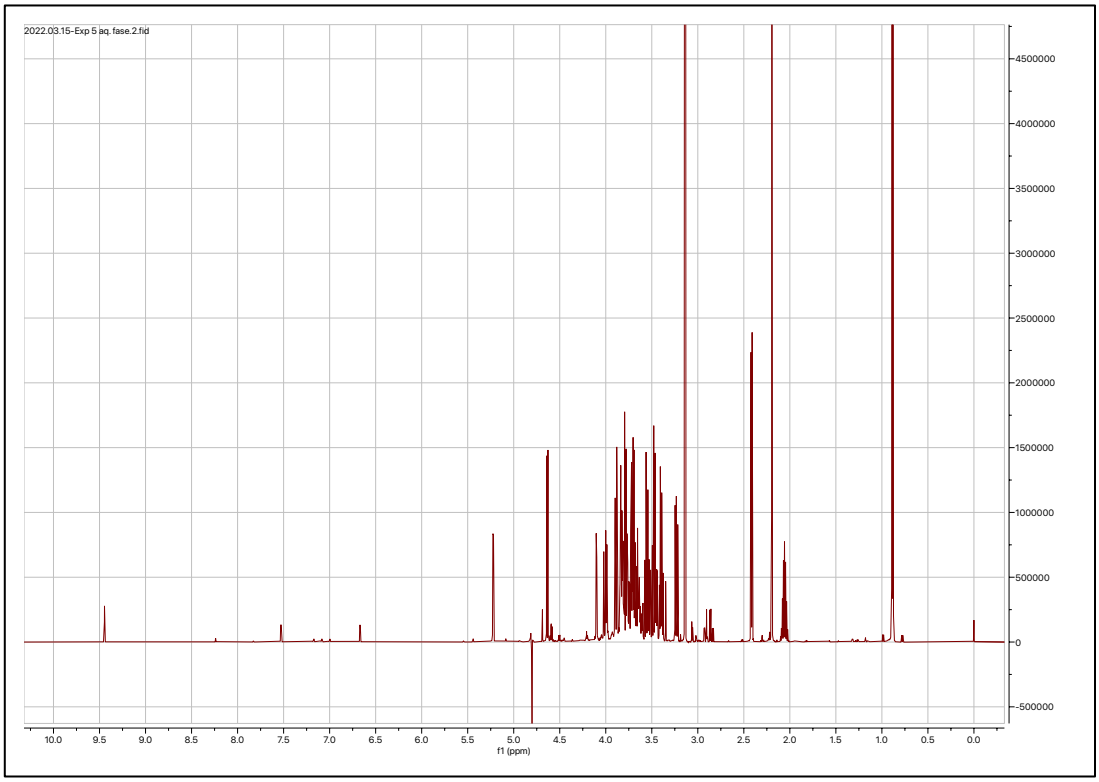


Figure C-10. The NMR spectrum of the aqueous product solution of experiment 5.

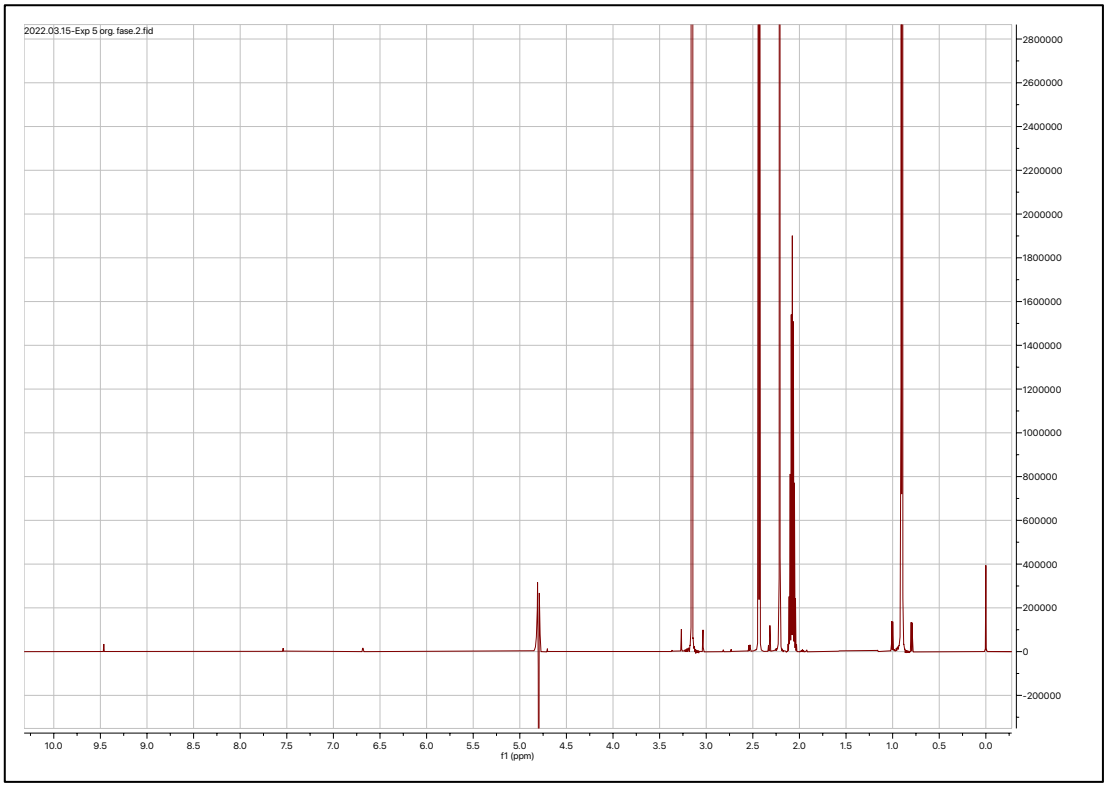


Figure C-11. The NMR spectrum of the organic product solution of experiment 5.

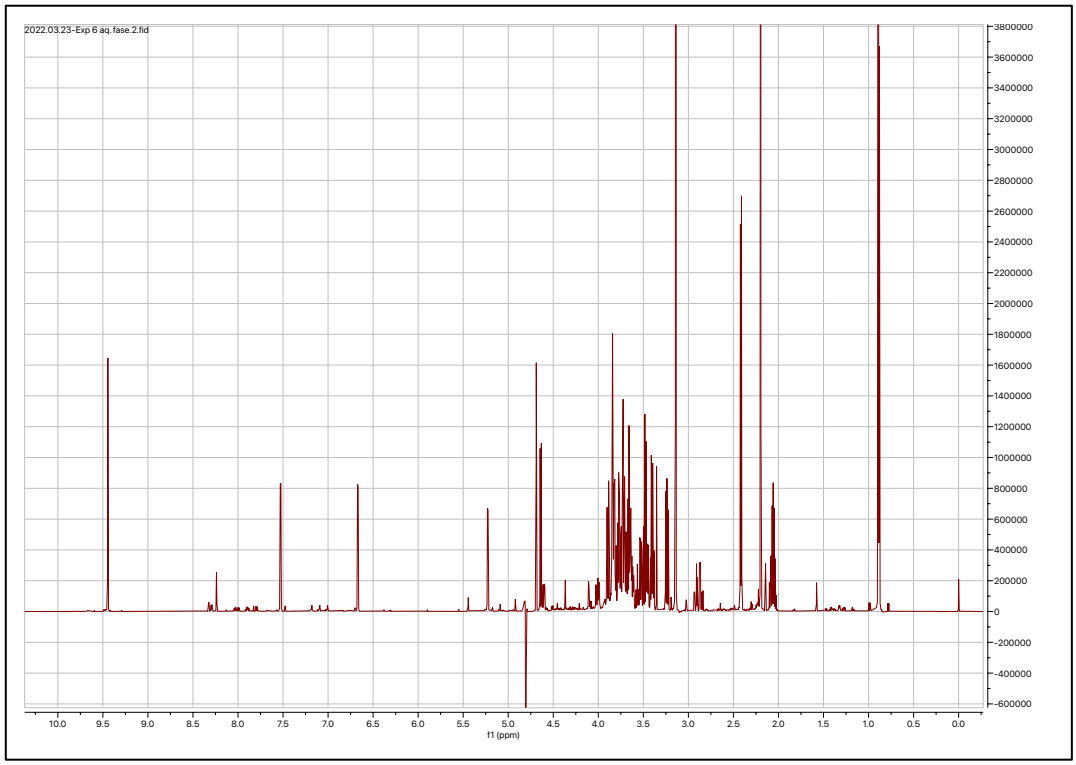


Figure C-12. The NMR spectrum of the aqueous product solution of experiment 6.

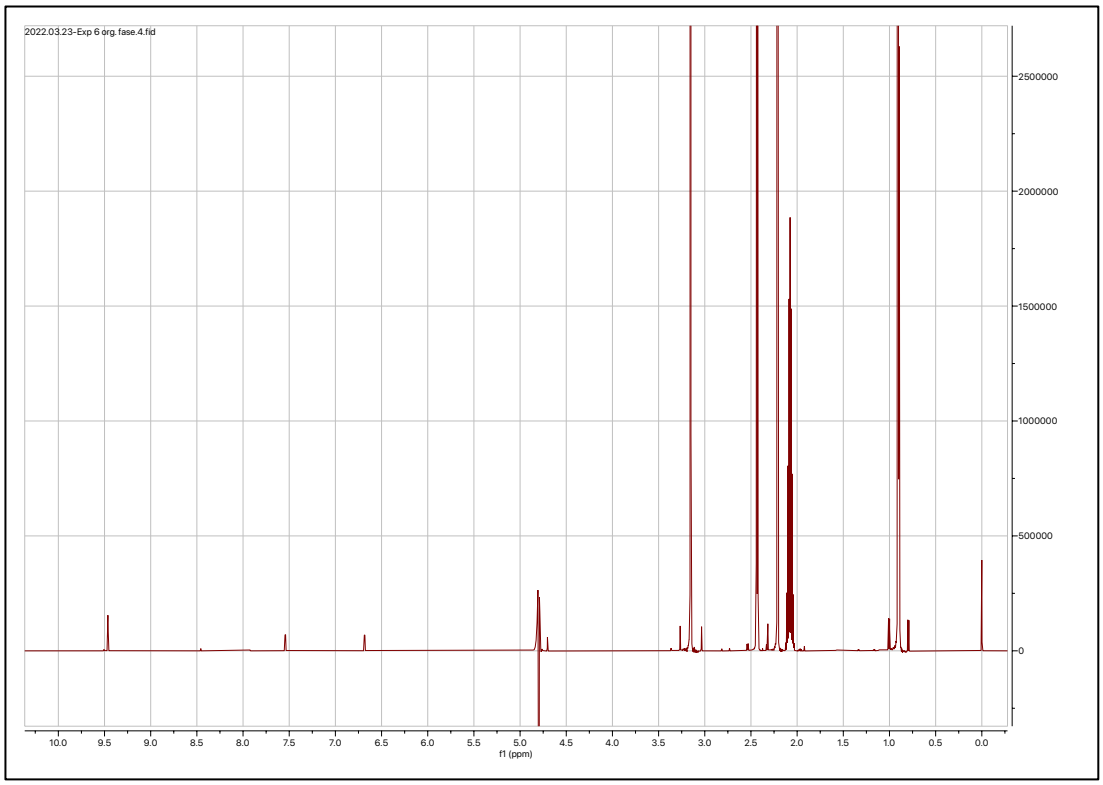


Figure C-13. The NMR spectrum of the organic product solution of experiment 6.

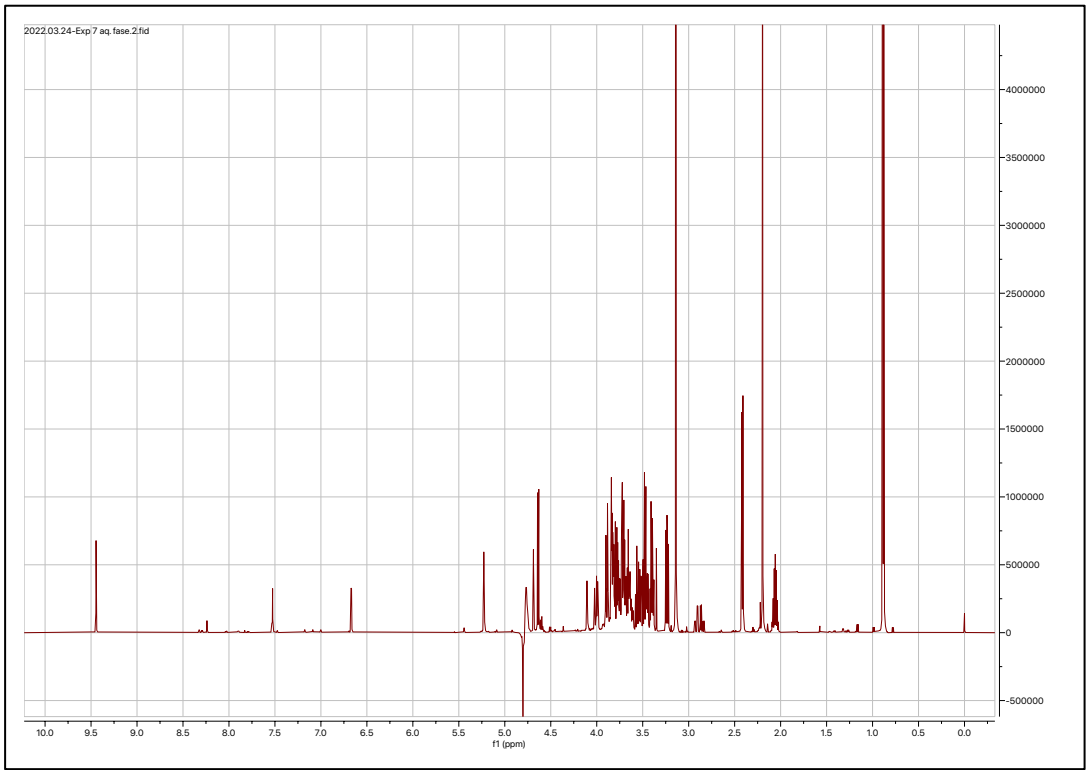


Figure C-14. The NMR spectrum of the aqueous product solution of experiment 7.

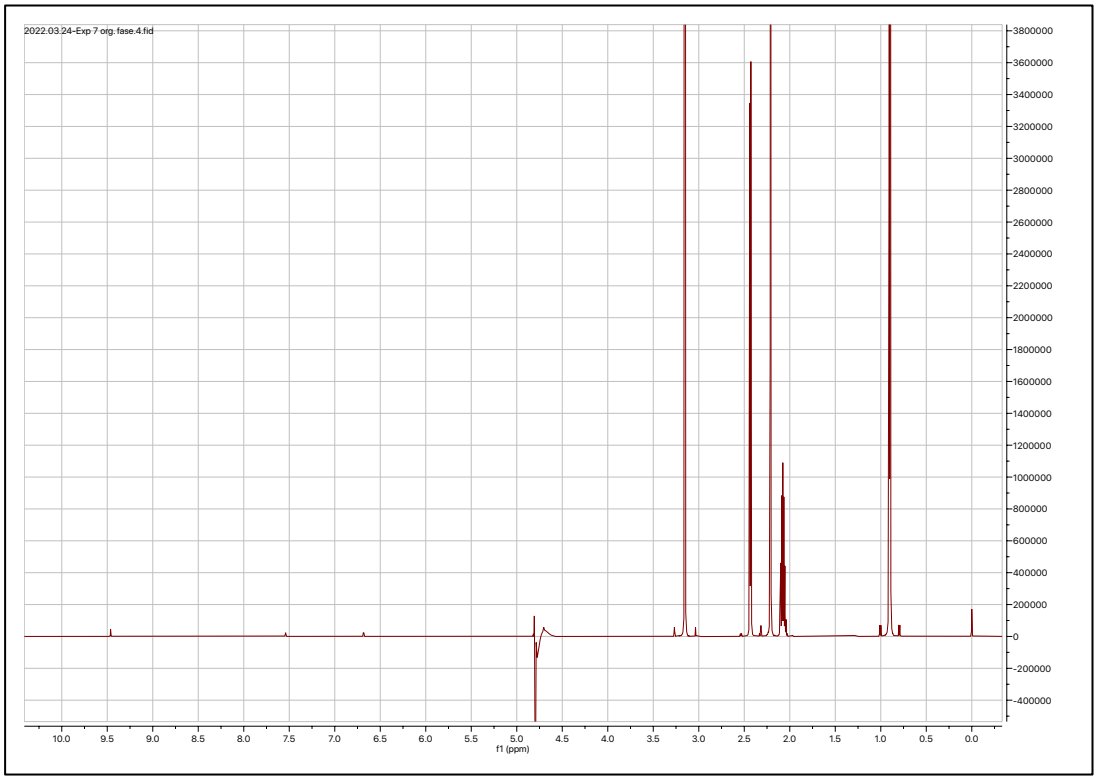


Figure C-15. The NMR spectrum of the organic product solution of experiment 7.

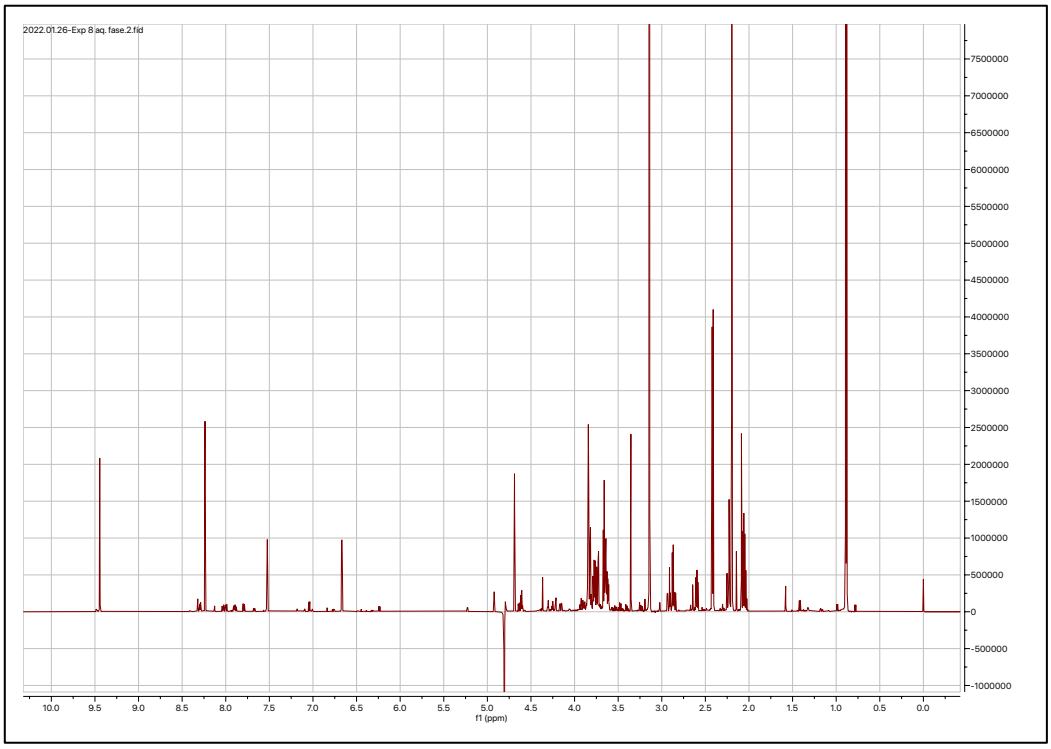


Figure C-16. The NMR spectrum of the aqueous product solution of experiment 8.

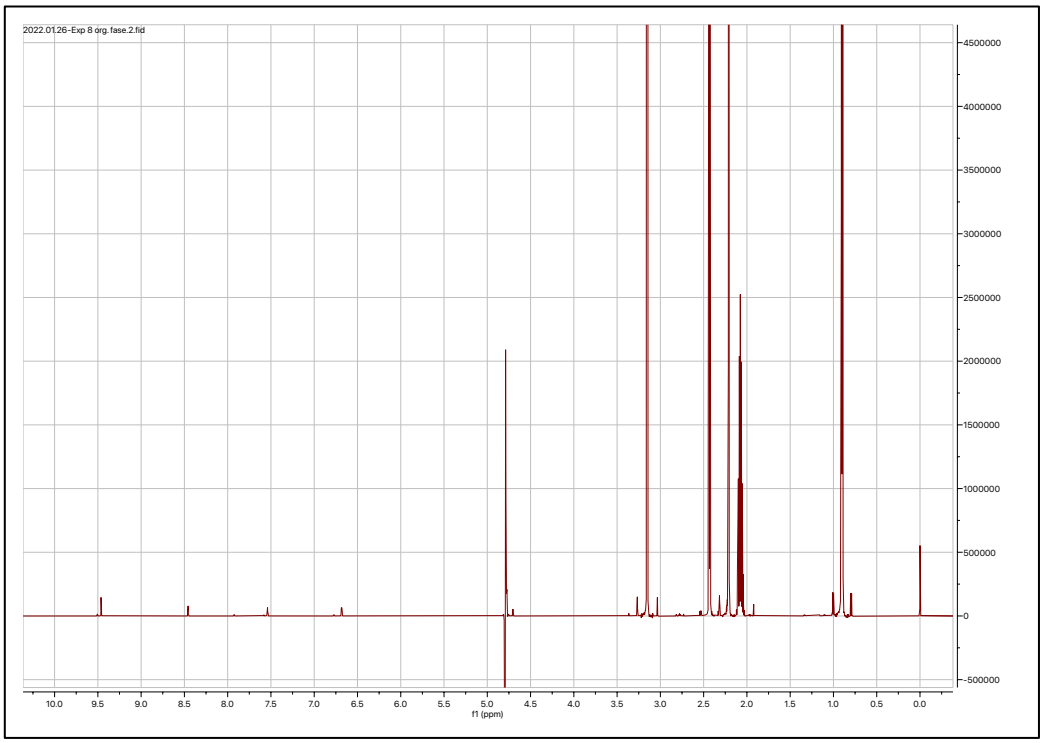


Figure C-17. The NMR spectrum of the organic product solution of experiment 8.

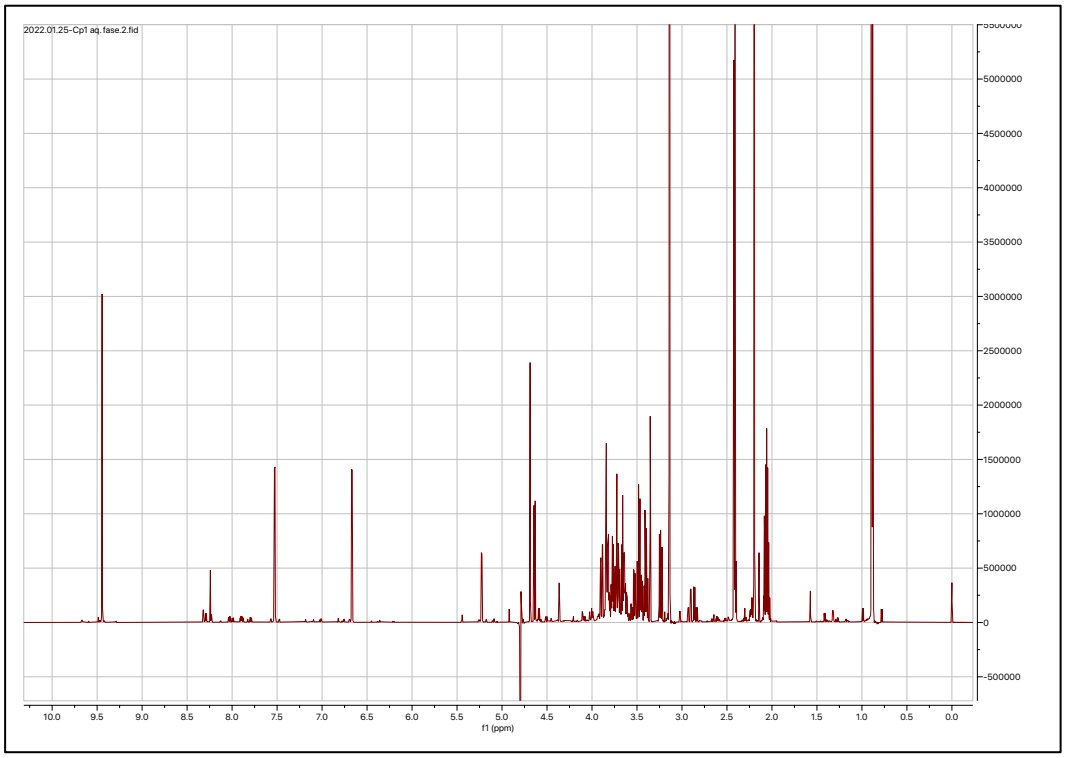


Figure C-18. The NMR spectrum of the aqueous product solution of centre point (CP) experiment 1.

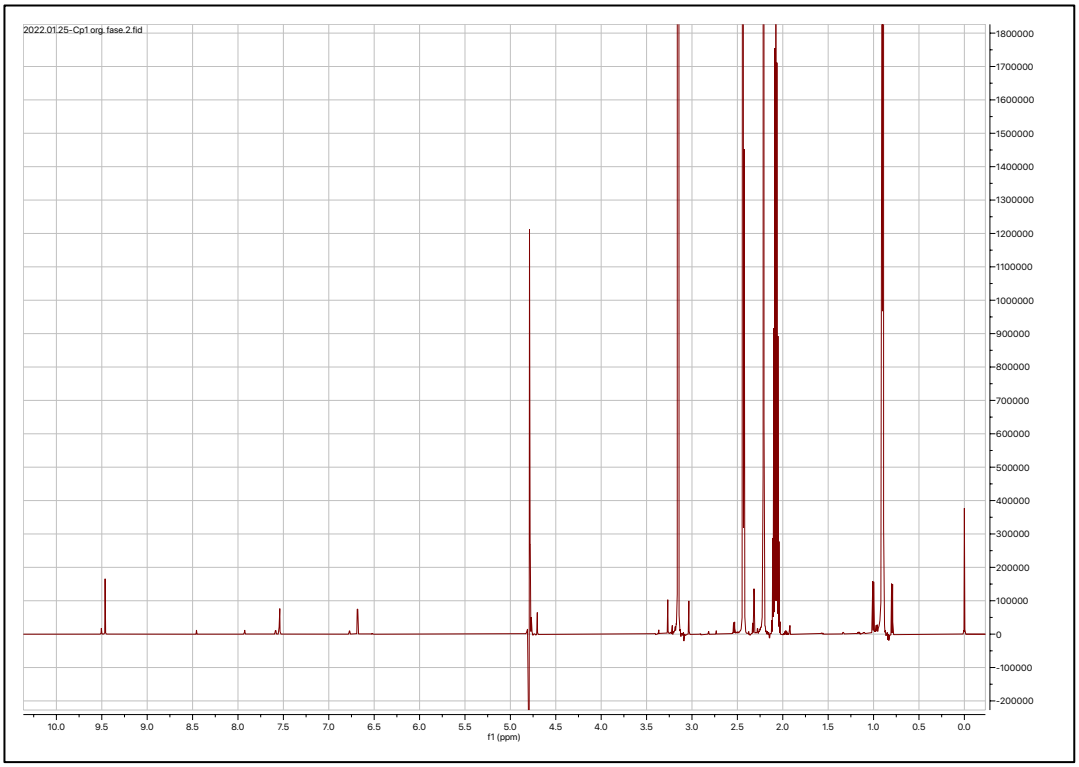


Figure C-19. The NMR spectrum of the organic product solution of centre point (CP) experiment 1.

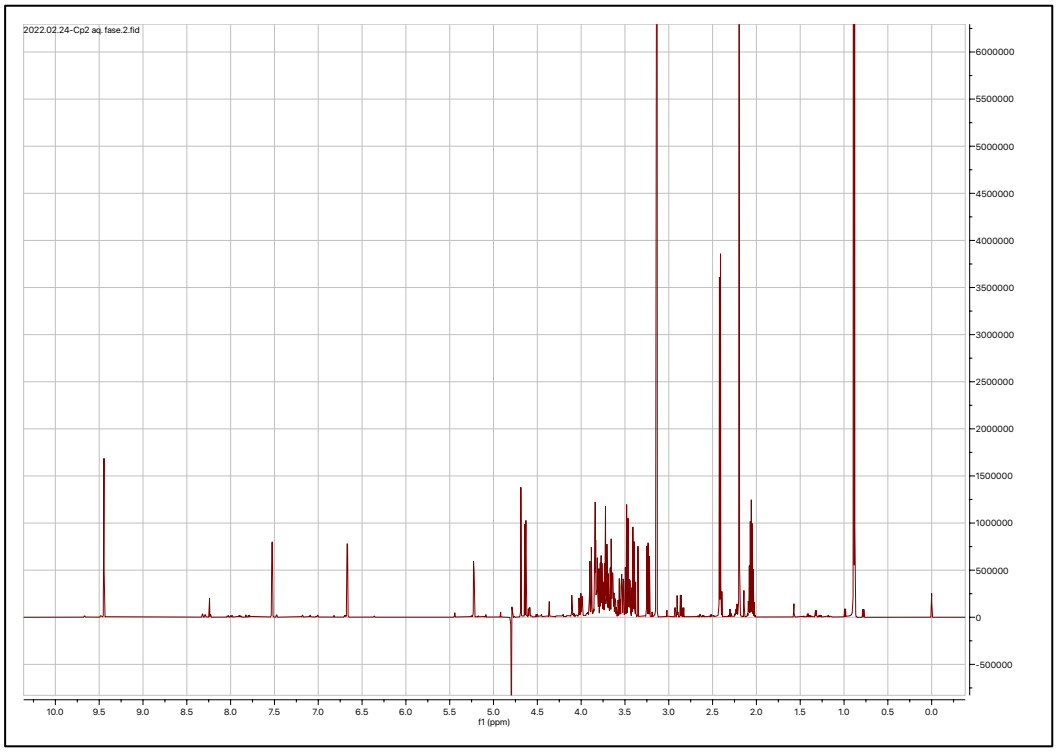


Figure C-20. The NMR spectrum of the aqueous product solution of centre point (CP) experiment 2.

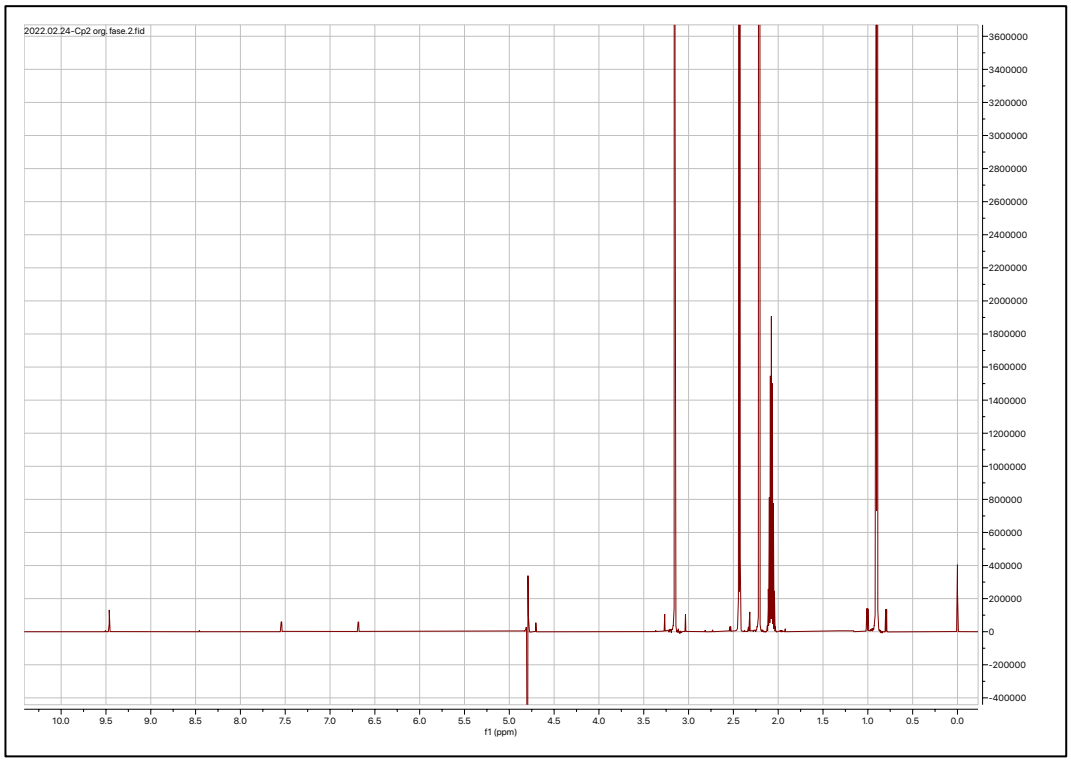


Figure C-21. The NMR spectrum of the organic product solution of centre point (CP) experiment 2.

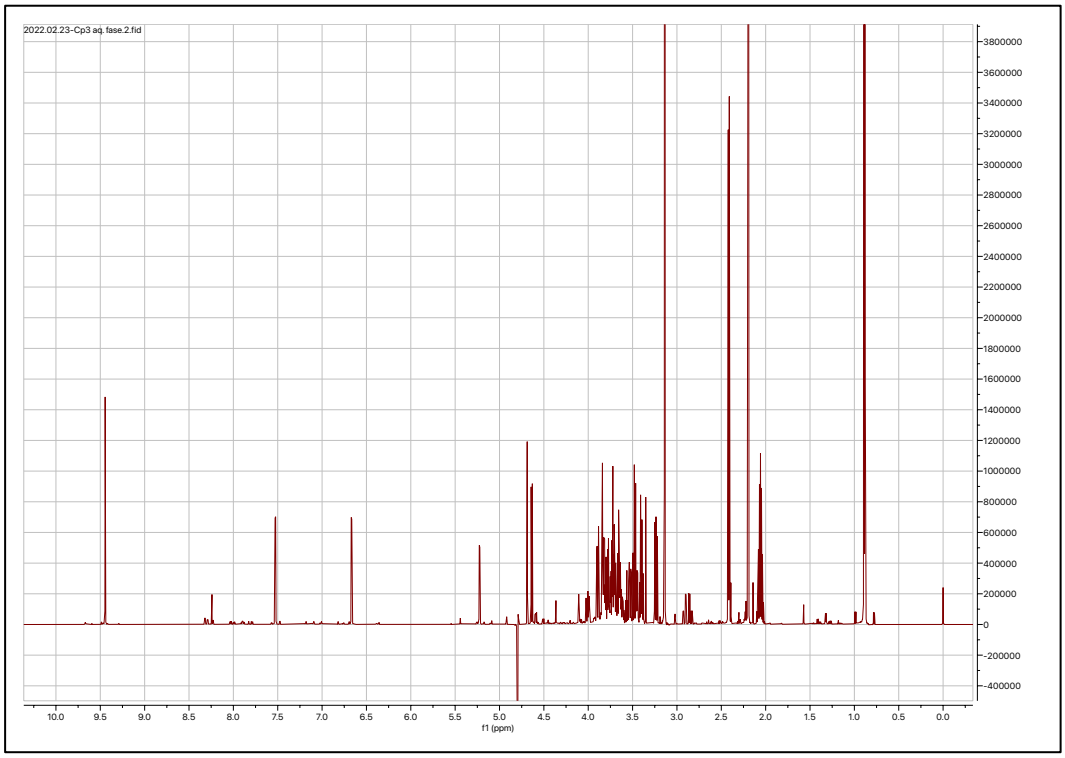


Figure C-22. The NMR spectrum of the aqueous product solution of centre point (CP) experiment 3.

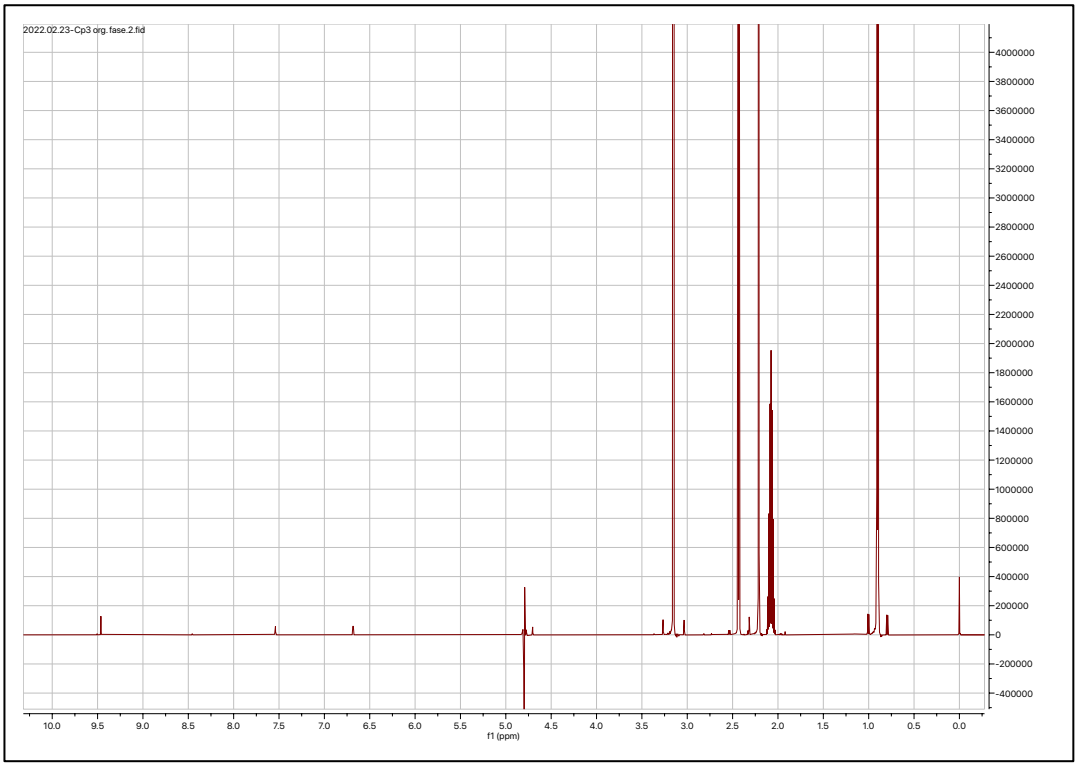


Figure C-23. The NMR spectrum of the organic product solution of centre point (CP) experiment 3.

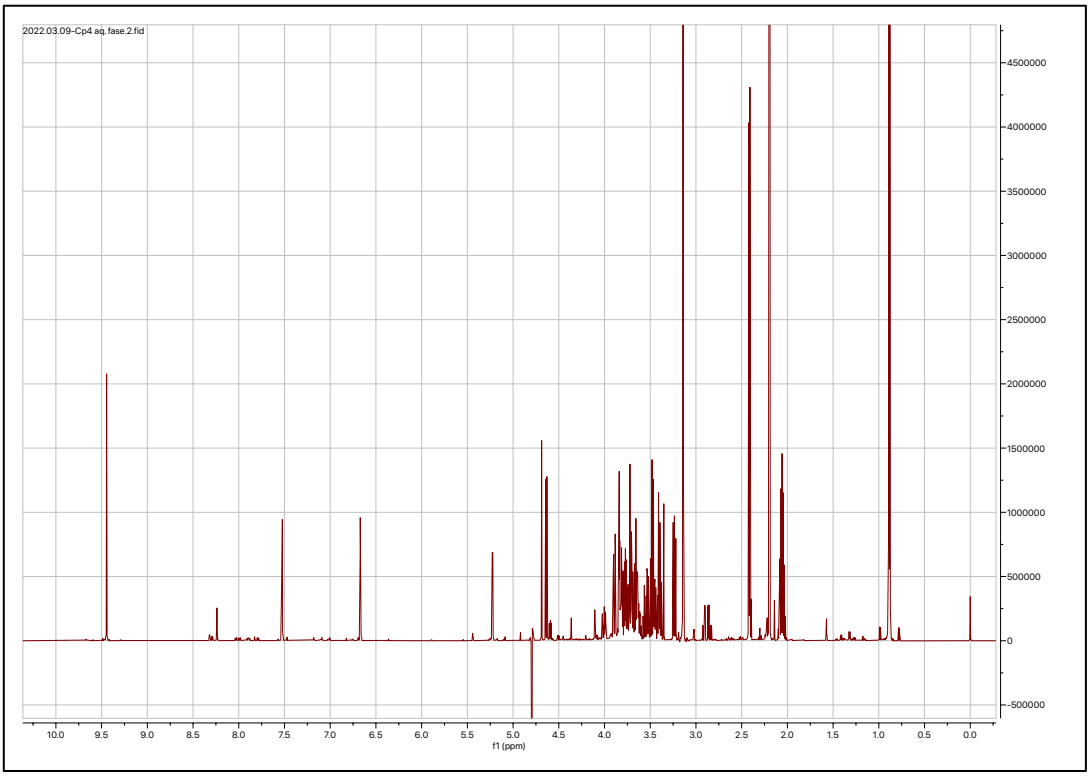


Figure C-24. The NMR spectrum of the aqueous product solution of centre point (CP) experiment 4.

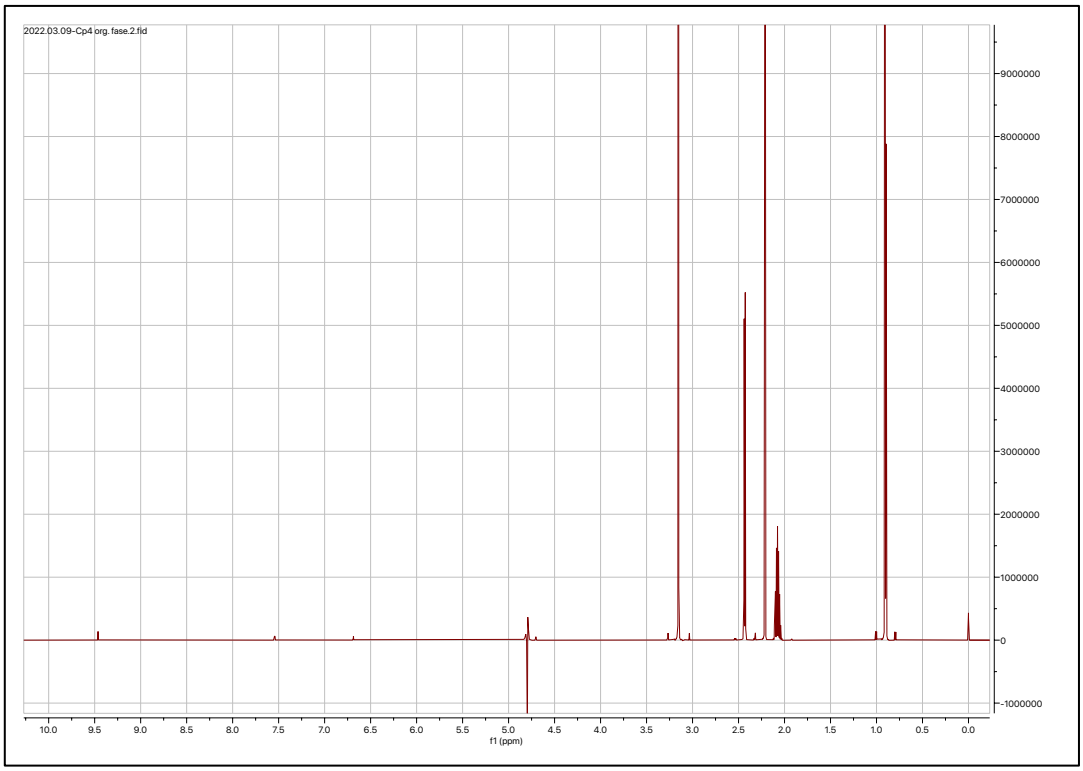


Figure C-25. The NMR spectrum of the organic product solution of centre point (CP) experiment 4.

Appendix D
Raw data from the optimizing full factorial design, and additional results

Table D-1. All the variables and constants for the optimizing full factorial design. Substrate wet is the weight of the raw cherries added to the reactor and substate dry is the calculated dry weight.

Exp #:	T (°C):	t (min):	Vaq(mL):	Vorg(mL):	Catalyst (g):	Substrate wet (g):	Substrate dry (g):
2.1	190	30	2.5	2.5	0.025	1.4131	0.2
2.2	230	30	2.5	2.5	0.025	1.4131	0.2
2.3	190	75	2.5	2.5	0.025	1.4131	0.2
2.4	230	75	2.5	2.5	0.025	1.4131	0.2

Table D-2. The results from the optimizing factorial design, including the concentrations of FUR and HMF in the aqueous and organic product solutions.

Exp #:	cFUR (aq) (mM):	cHMF (aq) (mM):	cFUR (org) (mM):	cHMF (org) (mM):	FUR yield, (m%):	HMF yield, (m%):
2.1	0.36	18.58	0.95	34.11	0.16	8.93
2.2	1.63	60.05	6.32	58.38	0.86	20.04
2.3	0.93	47.85	3.19	47.95	0.50	17.21
2.4	1.56	18.60	6.16	18.98	0.86	6.23

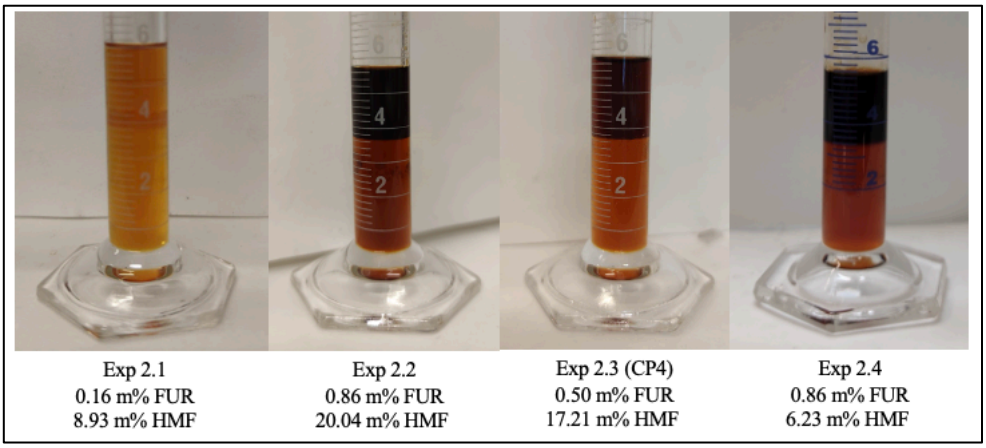


Figure D-1. Displays pictures of all the product solutions belonging to each experiment in the optimizing full factorial design before phase separation.

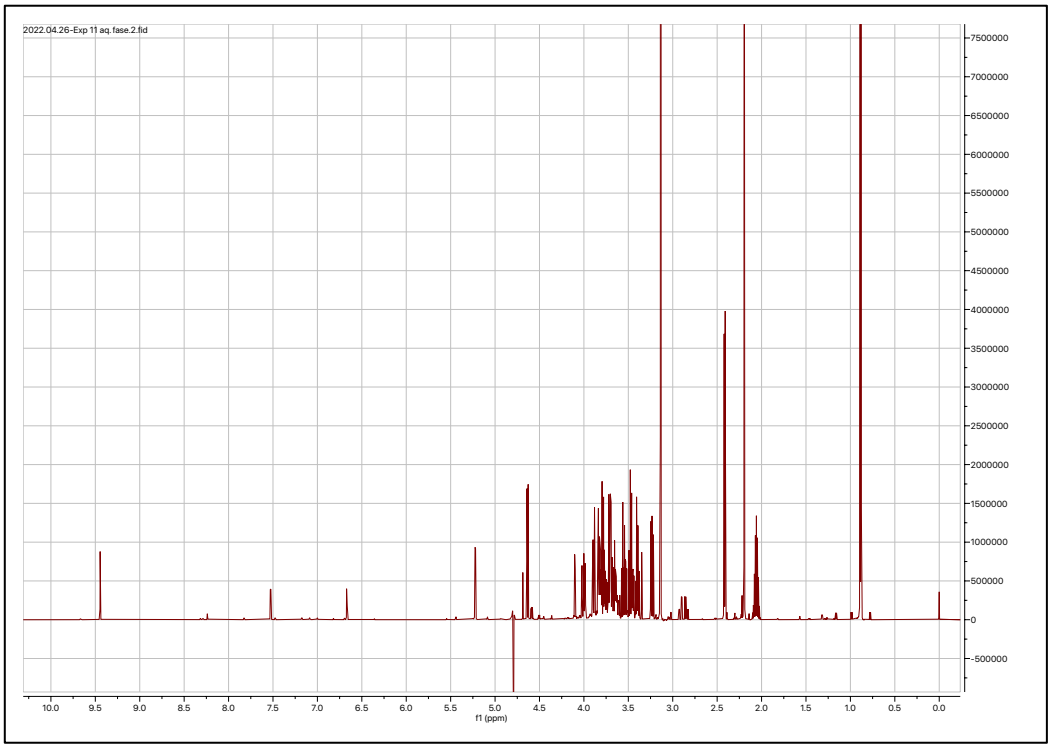


Figure D-2. The NMR spectrum of the aqueous product solution of experiment 2.1.

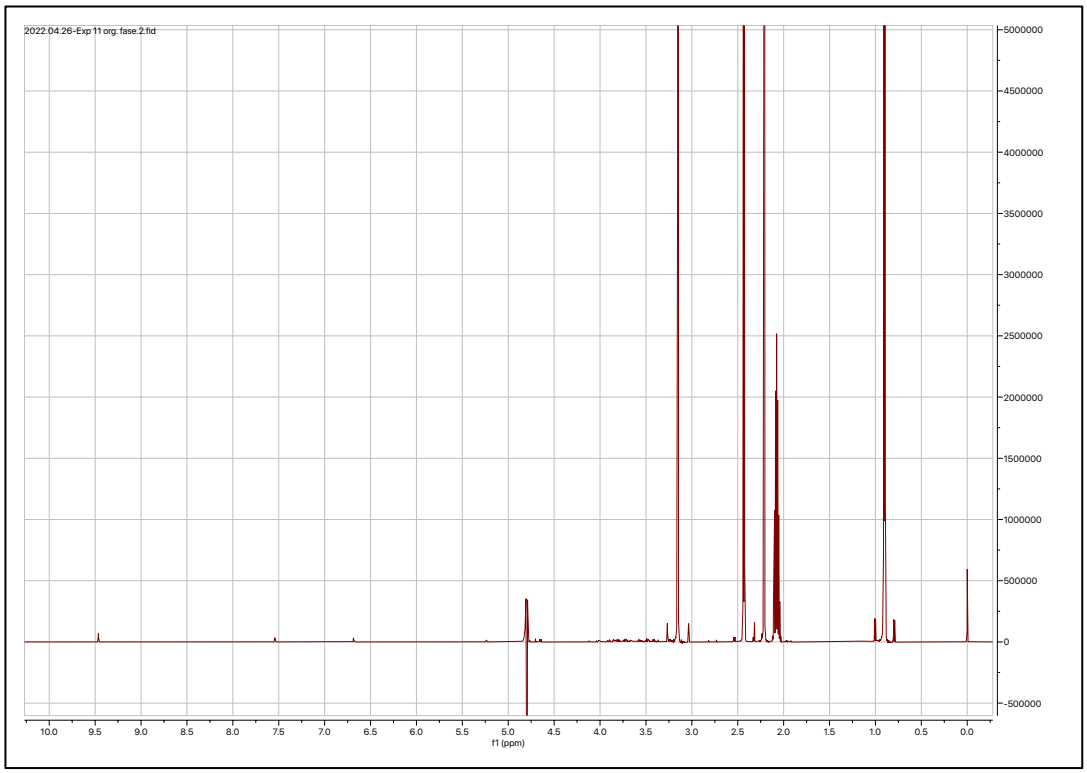


Figure D-3. The NMR spectrum of the organic product solution of experiment 2.1.

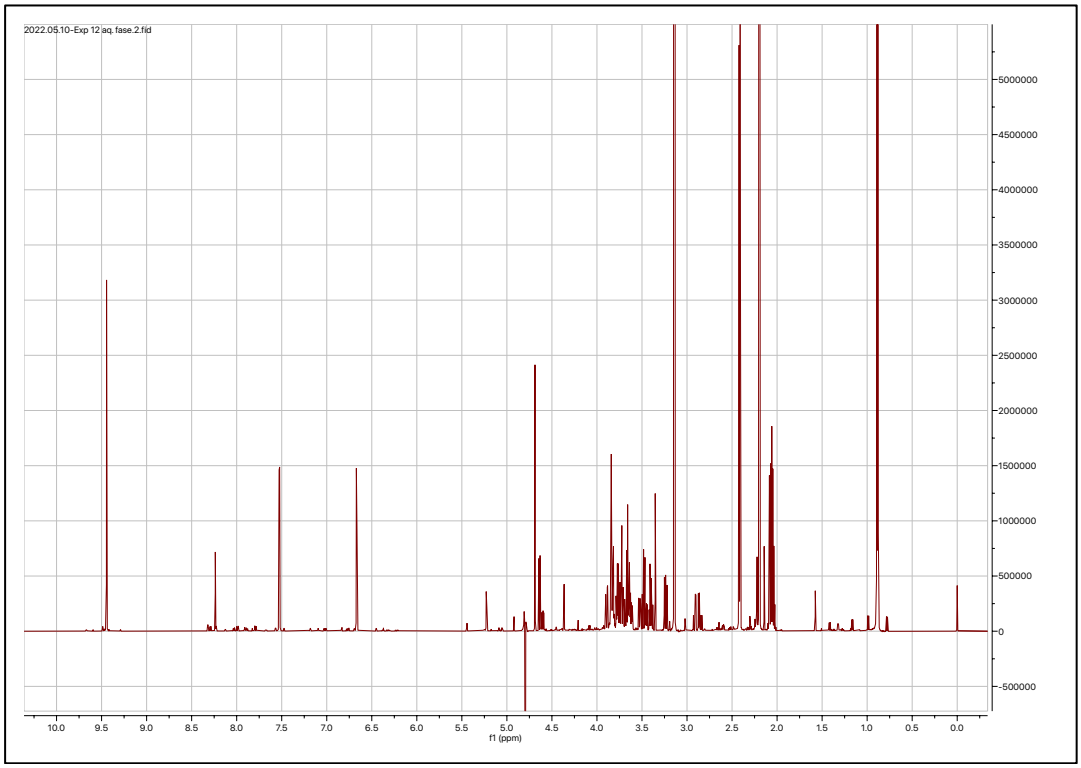


Figure D-4. The NMR spectrum of the aqueous product solution of experiment 2.2.

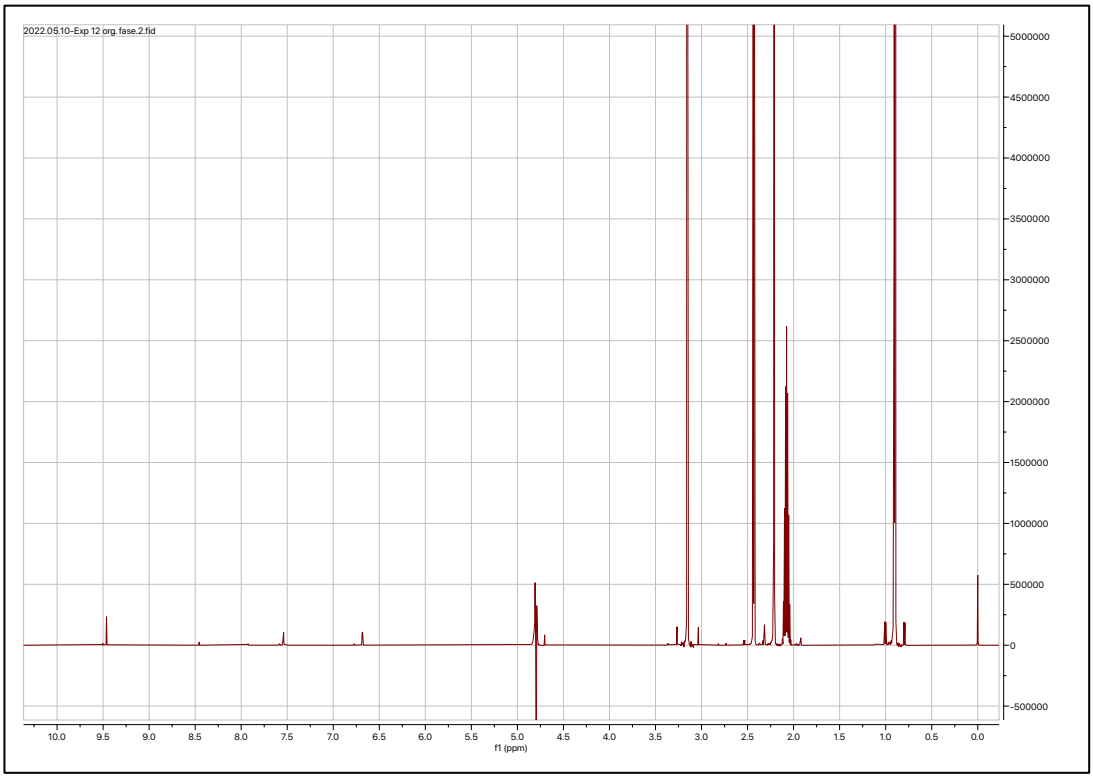


Figure D-5. The NMR spectrum of the organic product solution of experiment 2.2.

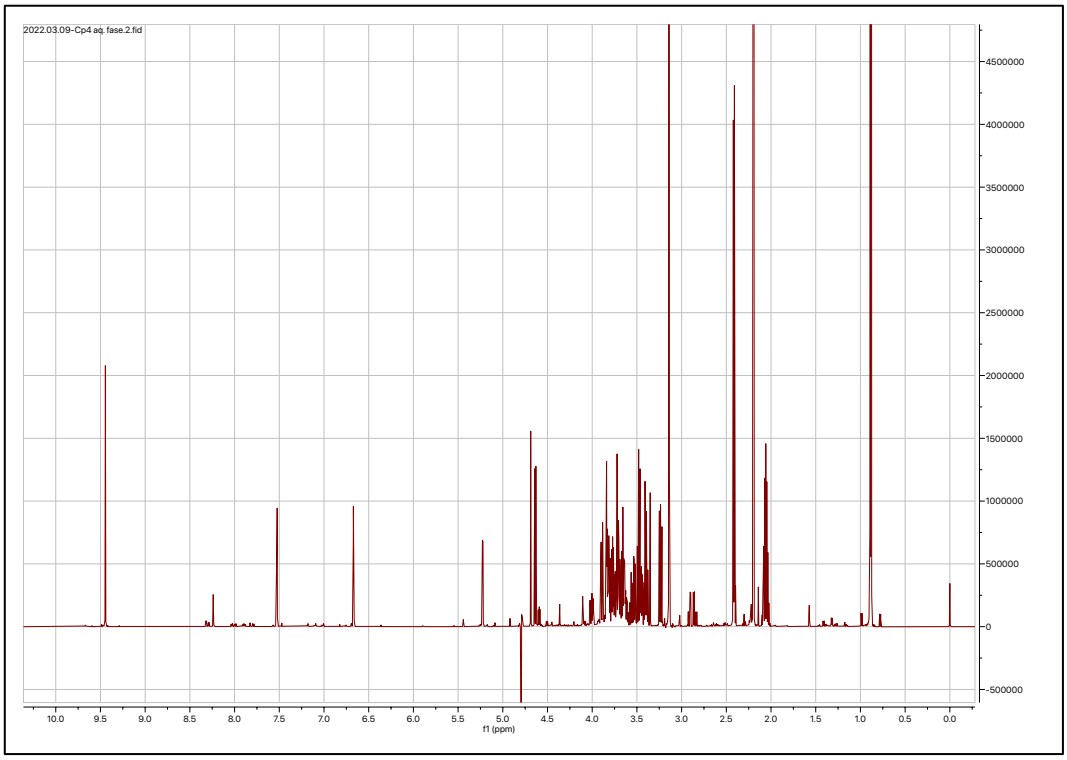


Figure D-6. The NMR spectrum of the aqueous product solution of experiment 2.3 (CP4).

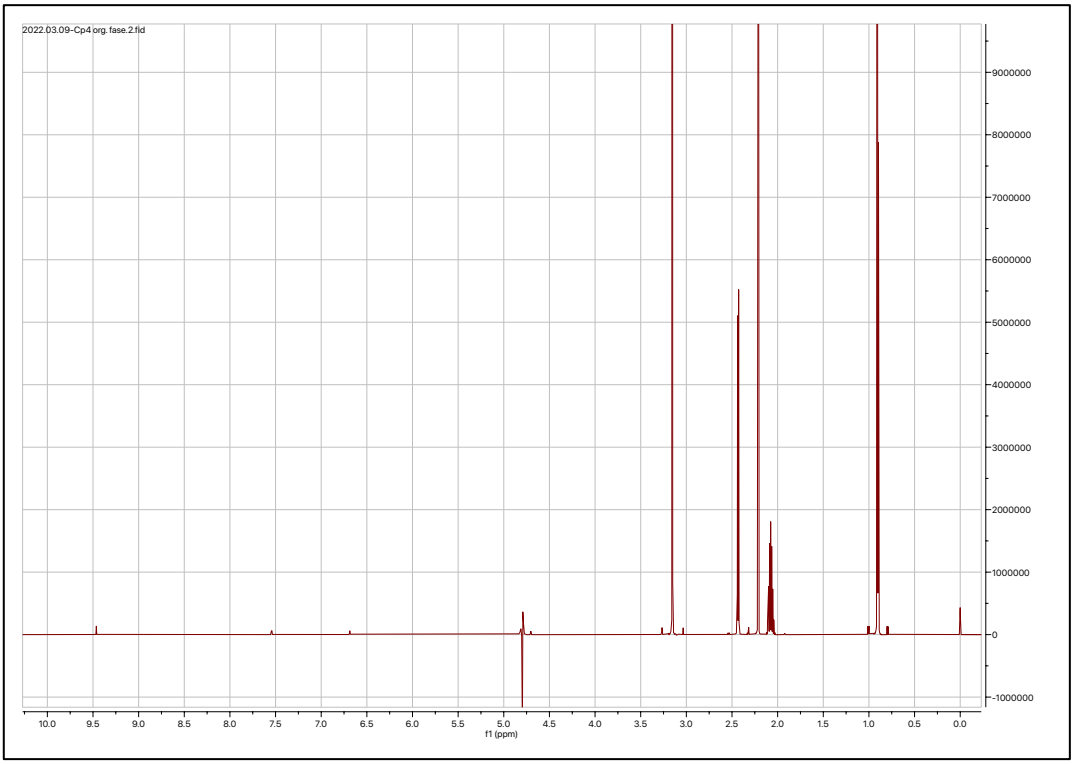


Figure D-7. The NMR spectrum of the organic product solution of experiment 2.3 (CP4).

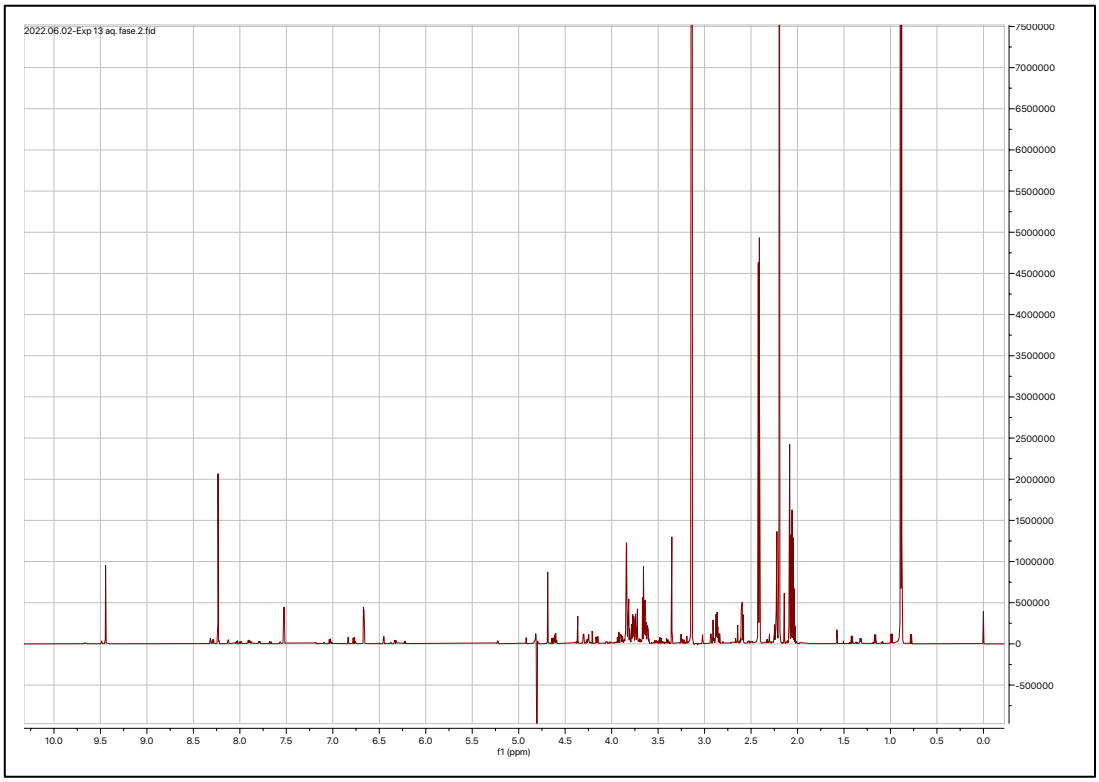


Figure D-8. The NMR spectrum of the aqueous product solution of experiment 2.4.

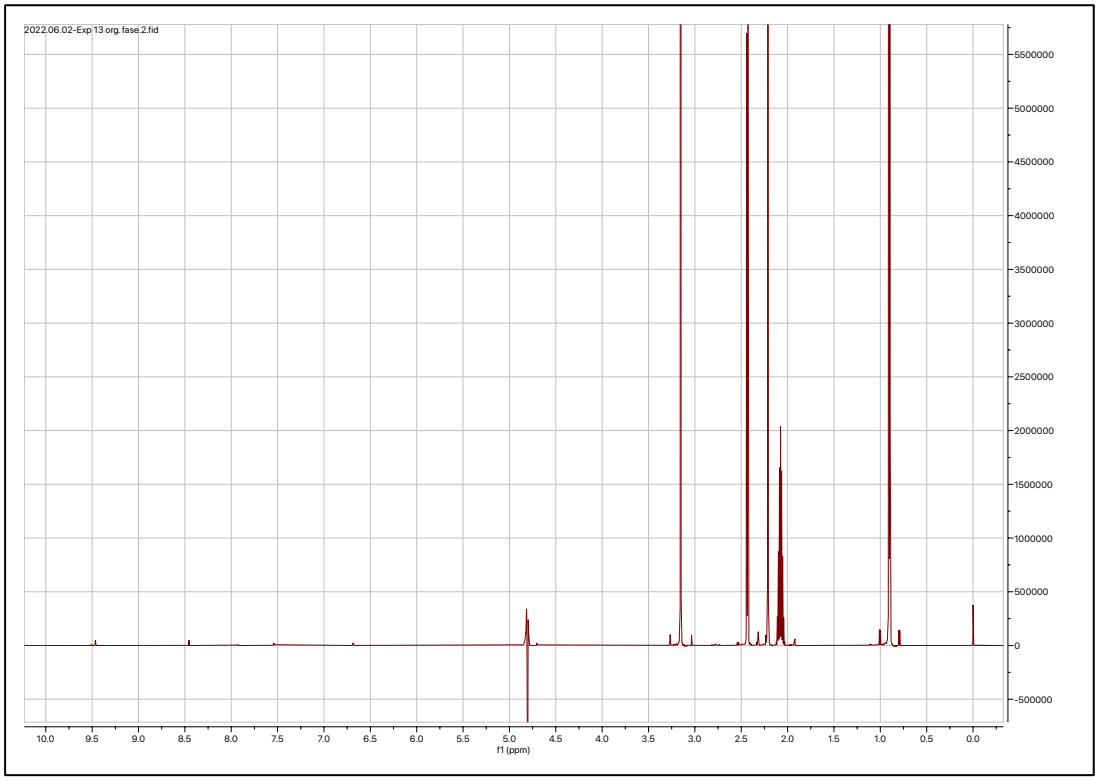


Figure D-9. The NMR spectrum of the organic product solution of experiment 2.4.

Appendix E
Some raw data and pictures of parts included in the reaction system

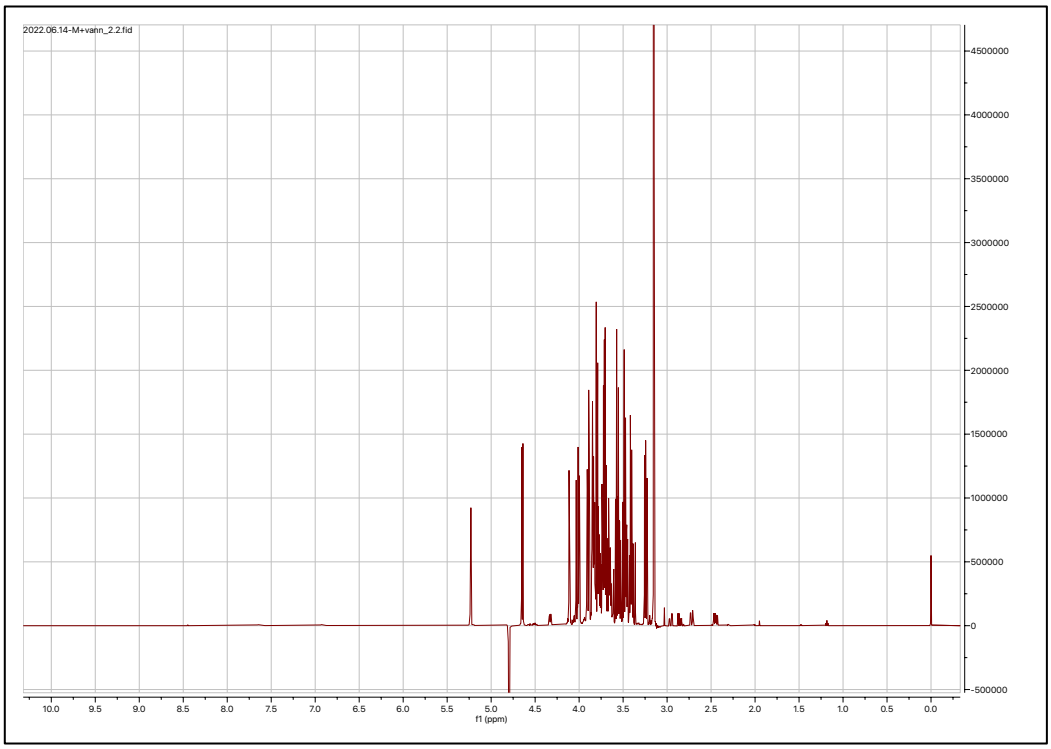


Figure E-1. NMR spectrum of cherries and water shaken together and filtered.

Figure E-2. Pictures of the reactor: Without lit, with lid and closed.

Figure E-3. Picture of the 10mL volumetric flask the organic phase was extracted with water in, after refrigeration.